Journal of Mathematical Biology (2026) 92:54
https://doi.org/10.1007/s00285-026-02376-4

Mathematical Biology

Derivations of animal movement models with accumulated memory

Tian Xu Wang[1,2] · Kyung-Han Choi[1,2] · Hao Wang[1,2]

Received: 26 March 2025 / Revised: 28 February 2026 / Accepted: 4 March 2026

Abstract

Animals continuously update their movement decisions using both real-time observations and historical information from experience, social interactions, or environmental cues, which we call accumulated memory (also called distributed delay memory). While memory is important for animals, how it influences movement strategies has received limited attention. We address this gap by integrating accumulated memory into three widely used models: advection-diffusion, Fickian-type diffusion, and Fokker–Planck type diffusion. These represent distinct strategies: (i) gradient-based movement, responding to environmental gradients; (ii) environment matching, symmetrically adjusting movement rates; and (iii) location-based movement, relying solely on local suitability. We derive each model from random walk models to compare how different memory-based movement strategies at the individual level give rise to distinct macroscopic population behaviors. Furthermore, we establish the local existence of solutions for a general model encompassing all three cases using fixed-point theory and provide a linear stability analysis. Numerical simulations show that the Fickian model always converges rapidly to a uniform state. Under memory-suppressed conditions, the advection–diffusion and Fokker–Planck models may exhibit aggregation, whereas under memory-enhanced conditions all models eventually reach uniformity, with the advection–diffusion and Fokker–Planck models sometimes displaying oscillatory wiggling pattern or periodic movement before stabilization.

Keywords Animal movement · Movement strategies · Accumulated memory · Formal derivation · Advection-Diffusion model · Fickian diffusion · Fokker–Planck type diffusion

Mathematics Subject Classification 35Q92 · 35K57 · 35A01 · 60J75

✉ Hao Wang
hao8@ualberta.ca

[1] Interdisciplinary Lab for Mathematical Ecology & Epidemiology, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

[2] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

Published online: 17 March 2026

1 Introduction

Animal movement is a fundamental ecological process that shapes populations, ecosystems, and evolutionary dynamics (Nathan et al. 2008). Memory plays a central role in animal movement by enabling individuals to act on past information derived from personal experience, social interactions, or environmental cues. At the individual level, animals recall resource patches, threats, or landmarks, for example, pigeons can recall and report past actions (Zentall et al. 2001), while chimpanzees use lexigram keyboards to indicate hidden food locations up to 16 hours after observing baiting events (Terrace 2005). Beyond personal memory, social animals benefit from information exchange within groups (Couzin et al. 2005). Ants, for instance, frequently adjust their behavior based on the actions of others, sharing and updating information in real time (Jackson and Ratnieks 2006). Such interactions allow individuals to access historical knowledge gathered by other members, even across spatially distant locations (Smith 1980). Animals also exploit information embedded in the physical environment, such as chemical cues left by conspecifics or persistent habitat features. These cues may carry information about past occupancy, the number and identity of individuals, the elapsed time since their presence. For example, ants can detect and interpret chemical footprints to reconstruct past movement paths and adjust their own trajectories accordingly (Wüst and Menzel 2017). Importantly, this form of memory use is not restricted to social species; even animals that are not typically social can exploit such environmental signals to guide movement decisions (Tóth et al. 2020). Taken together, these findings suggest that animals often have access to historical information, making it reasonable to assume that past population states contribute to present movement decisions.

In recent years, movement models have been extended to incorporate how animals process information and respond to their environment (Painter et al. 2024; Carrillo and Craig 2019; Potts and Lewis 2019). A widely used modeling framework for describing animal movement influenced by taxis toward or away from environmental factors is the advection–diffusion model:

$$u_t = D\Delta u + \nabla \cdot (u\nabla\gamma(v)), \tag{1}$$

where u represents the population density of animals, and v denotes a factor such as food availability, threat distribution, chemical concentration, environmental cues, or memory. Here, the diffusion coefficient D reflects unbiased random movement, while the advective potential $\gamma(v)$ introduces a directional bias in movement based on the factor v at spatial location x and time t. The specific form of $\gamma(v)$ determines how individuals respond to this factor. For example, (Winkler 2010) studied a chemotaxis model with a signal-dependent response function $\gamma(v) = \frac{\chi_0}{\alpha(1+\alpha v)}$, which provided a good fit to experimental data from (Dahlquist et al. 1972). Other common choices in the context of cell chemotaxis include $\gamma(v) = -(1+c)\ln(c+v)$ and $\gamma(v) = -v$ (Hillen and Painter 2009). Moreover, (Painter et al. 2024; Potts and Lewis 2019; Carrillo and Craig 2019; Giunta et al. 2022a; Salmaniw et al. 2025; Giunta et al. 2022b; Liu et al. 2025a; Ducrot et al. 2018) considered $\gamma(v)$ as spatially nonlocal information about population density.

(Wang and Salmaniw 2023) provided a comprehensive review of models where v is interpreted as various forms of memory or as the spatially nonlocal population information. If v represents fixed episodic memory of population density, i.e. $v(t) = u(t-\tau)$, the model becomes a delayed advection–diffusion model. This formulation assumes that movement decisions are made based on population information available exactly τ time units in the past. Such fixed-delay memory has motivated numerous studies, often under equation (1) with a linear response $\gamma(v) = v$ (Shi et al. 2020; Wang et al. 2023; Song et al. 2022, 2019; Shi et al. 2019; Zhang et al. 2023; Li et al. 2025; An et al. 2020; Li et al. 2022, 2023; Shi et al. 2021a; Song et al. 2021a; Wang et al. 2022a). Spatially nonlocal with fixed delay memory models have also been explored (Xue et al. 2024; Song et al. 2024; Wang et al. 2025). In this work, we restrict our attention to the local spatial memory case. More recently, (Shi and Shi 2024) and (Ji et al. 2024) investigated advection-diffusion models with accumulated population memory (also noted as distributed delay memory)

$$v = g *_t u = \int_{-\infty}^{t} g(t-\tau)u(\tau)\,\mathrm{d}\tau, \tag{2}$$

which assumes that all past population information contributes to current movement decisions, with their influence weighted by a memory kernel $g(\cdot)$, rather than relying on a fixed-time snapshot. This is consistent with empirical evidence that animals can act on all past information derived from personal experience, social interactions, or environmental cues. Mathematically, the fixed-delay case is recovered as a special case of the distributed delay when $g(t) = \delta(t-\tau)$, with δ the Dirac delta measure.

Based on earlier discussions, it is reasonable to assume that animals use past population information to inform their current movement decisions, especially more recent information. Typically, not all past experiences exert equal influence: recent information tends to be more impactful, while older information may diminish in relevance due to inaccessibility, uncertainty, or fading reliability. Thus, a memory kernel g that prioritizes recent data while gradually diminishing the impact of older information is biologically plausible. The Gamma distribution is often a suitable choice for such memory kernels (Shi and Shi 2024) (see further explanation in Section 6). In this paper, we also adopt v as the accumulated memory defined by (2), due to its greater generality. Figure 1 provides a visual explanation of this movement process, particularly on the discrete patch domain.

As a special case, if the accumulated memory decays over time at a fixed rate μ, this corresponds to using a kernel $g(t) = e^{-\mu t}$ in (2) (a weak kernel). In this case, if $\gamma(v) = Av$ for some constant A, the model (1) can be equivalently rewritten as the following coupled PDE–ODE system (Wang and Salmaniw 2023, (2.11)), (Potts and Lewis 2019, (1),(4)):

$$\begin{cases} u_t = D\Delta u + A\nabla\cdot(u\nabla v), \\ v_t = u - \mu v. \end{cases}$$

On the other hand, in heterogeneous environments, diffusion alone can generate non-uniform steady states without the involvement of additional advective dynamics (Kim et al. 2024). Several diffusion-based movement models have demonstrated such

effects, including aggregation and segregation phenomena (Funaki et al. 2012; Yoon and Kim 2017; Wang et al. 2022b). Among these, Fokker–Planck type diffusion and Fickian type diffusion are two commonly used formulations in animal movement models. In particular, Fokker–Planck type diffusion has been shown to provide a more effective survival strategy than constant diffusion in food-heterogeneous environments (Cho and Kim 2013; Kwon and Kim 2016). Classical movement models with memory have only been formulated within an advection–diffusion framework, where memory is embedded in the taxis terms. In this work, however, we broaden the perspective by allowing diffusion itself to be influenced by memory. Specifically, we consider three forms of movement models: advection–diffusion, Fokker–Planck type diffusion, and Fickian type diffusion. This broader framework enables us to investigate how memory-induced movement strategies shape population dynamics and generate spatial patterns that extend beyond the scope of classical formulations.

A general Fokker–Planck type diffusion model is given by

$$u_t = \Delta(\gamma(v)u). \tag{3}$$

Here, the motility function $\gamma(v)$ governs how individuals adjust their movement based on the value of factor v. (Liu et al. 2011) adopted a Fokker–Planck type diffusion to investigate the formation of periodic stripe patterns in engineered E. coli. In their system, cell motility is modulated by the concentration of acyl-homoserine lactone (AHL), using the function $\gamma(v) = \left(D_\rho + D_{\rho,0}\left(\frac{v}{K_h}\right)^m\right)/\left(1 + \left(\frac{v}{K_h}\right)^m\right)$, which captures the sharp decrease in motility at high AHL levels. Under this Fokker–Planck type diffusion framework, (Yoon and Kim 2017) studied a chemotaxis system where cell motility is given by the decreasing power-law function $\gamma(v) = cv^{-k}$, and proved that aggregation behavior can occur when $k > 1$. Further studies on pattern formation in density-suppressed motility models (i.e., $\gamma'(v) < 0$) can be found in (Ma et al. 2020; Wang and Xu 2021; Jin and Wang 2021; Smith-Roberge et al. 2019; Liu and Shi 2025b; Choi and Kim 2024), where different forms of $\gamma(v)$ are considered, such as $\gamma(v) = a + \frac{b}{(c+v)^k}$, $\gamma(v) = \frac{1}{c+v^k}$, and $\gamma(v) = e^{-v}$. The global existence of classical or weak solutions to (3) has also been established under suitable conditions on $\gamma(v)$; see (Ahn and Yoon 2019; Desvillettes et al. 2019; Fujie and Jiang 2020; Jiang et al. 2022; Lyu and Wang 2022; Tao and Winkler 2023) for details. When v represents accumulated memory (2), the model (3) implies that movement is determined solely by accumulated memory at the current position, without reference to its spatial gradient (see Section 3).

The Fickian type diffusion model can be written as

$$u_t = \nabla \cdot (\gamma(v)\nabla u), \tag{4}$$

which is commonly applied in physical diffusion influenced by geometric or boundary conditions (Fick 1855; Bringuier 2011; Okubo and Levin 2001). Both the Fickian and Fokker–Planck types can be derived from microscopic movement rules such as velocity-jump processes, as discussed by (Choi and Kim 2019). (Bringuier 2011) provided theoretical comparisons and physical examples: (i) Fickian diffusion arises in the movement of Na^+ and Cl^- ions in an open ocean; (ii) Fokker–Planck type

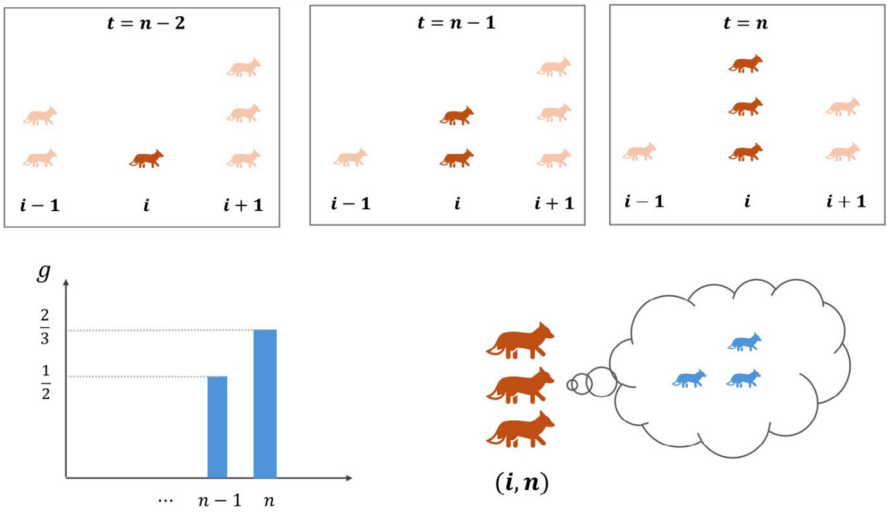

Fig. 1 An illustration of how foxes acquire and share the collective memory. The top panels depict the movement of six foxes across spatial locations $i-1$, i, and $i+1$ over time steps $t = n-2$, $n-1$, and n. The bottom-left graph shows memory retention, with foxes retaining memory only from the most recent two time steps, $t = n-1$ and n. If u_i^n denotes the population at location i and time n, then the memory of the foxes at (i, n) is computed as: $v_i^n = g(n-1)u_i^{n-1} + g(n)u_i^n = \frac{1}{2}u_i^{n-1} + \frac{2}{3}u_i^n$, which equals 3 in this example

diffusion occurs for H_3O^+ and HO^- in water, where the inhomogeneity is driven by a temperature gradient. (Wang et al. 2022b) also compared these two diffusion types in disease models over heterogeneous landscapes, assuming an increasing, bounded function γ. They found that Fokker–Planck type diffusion enhances spatial segregation between infected and susceptible populations, reducing overall infection size, which is not observed under Fickian diffusion.

When v is interpreted as accumulated memory (2), model (4) implies that animals compare the memory value v at their current location with that in surrounding areas and adjust their movement accordingly to match the movement rates of nearby individuals (see derivation in Section 3). Unlike the Fokker–Planck formulation, where movement depends solely on local memory v, Fickian diffusion reflects an environment-matching strategy that requires information from both the local and adjacent memory. In both the Fokker–Planck type diffusion (3) and the Fickian type diffusion (4), the monotonicity of $\gamma(v)$ induces biased movement rates based on the memory of past population v: if $\gamma(v)$ increases with v, areas of higher population density lead to faster dispersal; conversely, a decreasing $\gamma(v)$ implies slower dispersal in crowded areas.

For the Fickian type and Fokker–Planck type diffusion models, we assume that $\gamma(v)$ remains strictly positive. This strict positivity ensures that the model is uniformly elliptic, so classical analytical techniques are applicable. If γ is allowed to vanish at some value of v, the diffusion becomes degenerate, and different analytical methods are required. In contrast, for the advection–diffusion formulation, no restriction is imposed on the sign of γ itself.

All three of the aforementioned movement models, the advection–diffusion model (1), the Fokker–Planck type diffusion model (3), and the Fickian type diffusion model (4), can be unified within the general framework

$$u_t = \nabla \cdot \big(D(v)\nabla u + A(v)u\nabla v\big), \tag{5}$$

where again $u(x, t)$ represents the population density of animals, and $v(x, t)$ denotes accumulated memory (2). When $D(v) = D$ is constant and $A(v) = \gamma'(v)$, the model reduces to the advection–diffusion model (1); when $D(v) = \gamma(v)$ and $A(v) = \gamma'(v)$, it corresponds to a Fokker–Planck type diffusion (3); and when $D(v) = \gamma(v)$ and $A(v) = 0$, it reduces to the Fickian type diffusion model (4). This general framework was first derived by Patlak (1953) for particle random walk and later applied to describe bacteria movement guided by chemical gradients Keller and Segel (1971). It captures how both diffusion and directed movement can depend on environmental or internal factors, and different choices of $D(v)$ and $A(v)$ correspond to distinct movement strategies.

Since we are interested in how movement guided by memory redistributes the population within a bounded habitat, we impose a homogeneous Neumann boundary condition, which in our setting corresponds to a zero-flux boundary, so that the total mass is conserved. Although the movement operators include a time-convolved memory of past densities, this memory is nonlocal only in time; therefore, the Neumann condition still prevents any movement-driven gain or loss of population across the boundary. This ensures mass conservation with respect to movement, so that the diffusive and advective terms in (5) merely rearrange individuals within the domain without changing the total population. For the mathematical generality, the analysis of existence is carried out under a general homogeneous boundary operator. Stability analysis and all numerical simulations are conducted under homogeneous Neumann boundary conditions.

The objective of this paper is to investigate how different individual-level memory-based movement strategies give rise to three distinct movement models, (1), (3), and (4), and to examine the resulting differences in population-level movement patterns under homogeneous Neumann boundary conditions. The structure of the paper is as follows. In Section 2, we discuss the ecological implications of the three diffusion types. In Section 3, we formally derive these memory-dependent diffusion models using discretized methods and a discrete velocity-jump process, thereby explaining the underlying movement strategies. By tracing how continuous equations emerge from discrete movement behaviour, we aim to clarify how these models capture different movement strategies based on memory. The well-posedness of the generalized model with homogeneous boundary conditions (36) is established in Section 4, and the linear stability analysis is presented in Section 5. In Section 6, we compare the three diffusion types through numerical simulations under homogeneous Neumann boundary conditions, highlighting differences in movement patterns and population distributions.

2 Three memory-based movement strategies

Animals may adopt a variety of movement strategies when interacting with their environment. For instance, when the environment is approximately homogeneous or individuals have no particular preference, constant diffusion is commonly used to represent unbiased random roaming. However, in heterogeneous settings, animals often show preferences for certain environmental features, which requires movement models that account for spatial and temporal variability (Kim et al. 2024). We focus on three movement models, each reflecting a different movement strategy tied to how animals use memory. What follows is an introduction to these three representative movement strategies:

(i) Gradient-Based Movement: In this strategy, organisms use the gradient of accumulated memory at their current location to guide their movements. From a modeling perspective, this behavior can be described by advection–diffusion equations, which capture both unbiased random dispersal and active responses to environmental changes through directed movement along or against memory gradients. Gradient-based movement is common in nature (Wu et al. 2012; Smith et al. 2006), where animals exhibit clear preferences toward or away from specific environmental cues. For example, migratory birds such as geese rely on memory of landmarks or environmental cues (Zein et al. 2022), such as wind direction (Amélineau et al. 2014; Liechti 2006) or temperature gradients (Sapir et al. 2010), to navigate effectively during long migrations. Similarly, many territorial animals respond to population pressure by moving toward regions with either higher or lower density (Matthysen 2005). When animals can access not only real-time observations but also information about past population densities, they can rely on accumulated memory (2), which reflects the overall past crowding of the population. If animals move along this gradient, they tend to move toward locations that were previously highly crowded. Conversely, if they move opposite to this gradient, they prefer to remain in areas with lower population pressure.

(ii) Environment Matching: In this strategy, the movement rate depends on the accumulated memory at both the current location and adjacent locations. Animals share symmetric movement rates with nearby individuals based on these accumulated memories. Consequently, a change in memory at one location alters the movement rate for both that site and its neighbors simultaneously. This coordinated adjustment allows animals to avoid overcrowding and promote population stability. Assuming total mass is conserved, such matching behavior is expected to lead, over long times, to a uniform population distribution (see Lemma 5.2 and Section 6). From a modeling perspective, this process is well captured by Fickian diffusion equations. Unlike gradient-based movement, where spatial heterogeneity of accumulated memory induces directional movement bias and leads to heterogeneous population distributions, environment matching promotes symmetric movement between neighboring locations based on accumulated memory, thereby reducing spatial heterogeneity and fostering a more uniform population distribution.

(iii) Location-Based Movement: In this strategy, animals decide whether to move or stay based solely on the suitability of their current location. When population pressure is a determining factor, animals rely on accumulated memory to assess not only the current population pressure at a given site but also its past levels. When animals are attracted to higher population densities, locations with greater accumulated memory correspond to lower movement rates. Conversely, when animals prefer lower population pressure, locations with higher population pressure in the past correspond to higher movement rates. Importantly, this movement is unbiased, with animals having an equal probability of moving to any other neighboring site. However, because accumulated memory is spatially heterogeneous, movement rates vary across the landscape. This movement strategy can be captured by Fokker–Planck type diffusion equations. In contrast to gradient-based movement or Fickian diffusion, this approach does not rely on comparing the current site with surrounding locations.

These three movement strategies play essential roles in natural systems. In this work, we focus on accumulated population memory; however, the same movement strategies also apply when animals respond to other social or environmental cues, such as species competition, mating pressure, predation risk, territorial marking, resource distribution, or temperature. To better understand how individual movement strategies influence macroscopic population distribution patterns, we now turn to the derivation of the corresponding memory-based movement models.

3 Formal derivations of explicit memory models

In this section, we derive three classes of memory-dependent movement models, employing two classical approaches: discretization of space and time, and discrete velocity-jump processes. The main assumption is that individuals at a given location can access historical population information about that location. Such memory may arise from personal experience, communication within a social group, or external cues such as scent marks left by other individuals or environmental signals (Zentall et al. 2001; Couzin et al. 2005; Wüst and Menzel 2017).

Throughout these models, we refer to *accumulated memory* as a spatiotemporal map

$$v(x,t) = (g^*_{tu})(x,t)u = \int_{-\infty}^{t} g(t-s)\,u(x,s)\,\mathrm{d}s, \tag{6}$$

where at each point $(x,t) \in \Omega \times \mathbb{R}$, it represents the weighted accumulation of population information from all previous times up to t at location x. Here, $g \in L^1([0,\infty))$ is a kernel function representing the importance of historical population information. For example, one can consider that more recent information is weighted more heavily, while older information gradually becomes less influential. The memory-dependent mobility function $\gamma \in C^3(\mathbb{R})$ describes how memory modulates the local movement rate and determines how strongly memory influences movement. For example, if γ is constant, accumulated memory does not affect movement. If $\gamma(v) = e^v$, the effect of memory grows rapidly with v, so individuals respond strongly in regions with high

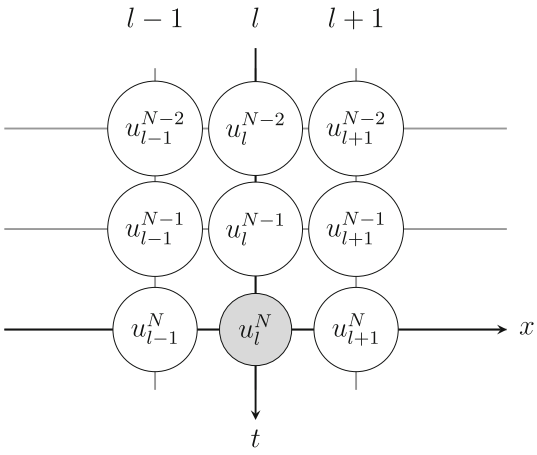

Fig. 2 Grid representation of the discrete method

accumulated memory. If $\gamma(v) = \frac{av}{c+v}$ with $a, c > 0$, the influence of memory saturates for large v, meaning additional memory has a diminishing effect on movement. A linear function, $\gamma(v) = av$, produces a proportional change in movement with memory, with the slope a controlling sensitivity (see more discussion in Section 1). Such memory-dependent movement may result from mechanisms like within-species competition, which promotes movement (positive density-dependence), or social cohesion and group foraging, which reduce movement (negative density-dependence) (Matthysen 2005). For the derivations of the Fickian-type and Fokker–Planck-type diffusion models, we assume $\gamma > 0$; whereas in the advection–diffusion formulation, γ is allowed to change sign.

The discrete method begins with a lattice walk or space-jump process, and specifies how transition probabilities depend on accumulated memory. For a discrete-time walk, these movement equations are derived in the diffusion limit by letting the space step size h and the time step δt go to zero such that the ratio $h^2/\delta t$ remains constant. An alternative stochastic process that may more accurately represent cell motion is the velocity-jump process (Othmer et al. 1988). In this case, the velocity and movement speed, rather than the spatial position, change according to accumulated memory.

Throughout this section, the domain $\Omega \subset \mathbb{R}^n$ is bounded with a smooth boundary, and $u(x, t)$ is a population density function satisfying $u \in C_b^{2,1}(\bar{\Omega} \times (-\infty, t])$.

3.1 One-Dimensional discrete method

We now derive the equation using a discretization approach. Let the one-dimensional spatial domain be $\Omega = [0, L]$, divided into M uniform subintervals of size $h = L/M$, with grid points $x_l = lh$ for $l = 0, 1, \ldots, M$. The temporal domain is $(-\infty, t]$, discretized with time step $\tau > 0$, giving time points $t_k = t - (N - k)\tau$ for integers $k \leq N$. Figure 2 provides a grid representation of the discrete method. Let $u(x, t)$ be a smooth population-density function. The population density at position x_l and time t_k (i.e., $u(x_l, t_k)$) is denoted by u_l^k for brevity.

The continuous accumulated memory function is defined in (6) and the corresponding discretized accumulated memory at position x_l up to time t_N is

$$v_l^N = \tau \sum_{k=-\infty}^{N} g^{N-k} u_l^k, \tag{7}$$

where $g^k = g(k\tau)$ quantifies the influence of past population information at time t_k. As $u \in C_b^{2,1}(\bar{\Omega} \times (-\infty, t])$, for small τ we have

$$v(x_l, t_N) = v_l^N + O(\tau^2).$$

3.1.1 Advection–Diffusion model

We now consider the case where the relative spatial gradient of the accumulated memory of population density influences the probability of local movement. For species that do not prefer crowding, if the gradient of the accumulated memory is negative, i.e., the accumulated memory at position x_{l+1} is lower than at x_l, then individuals at x_l are more likely to move toward x_{l+1}, and vice versa. Hence, we refer to it as gradient-based movement.

This accumulated memory reflects the historical population distribution, which can vary spatially due to heterogeneity in population density. At x_l, we denote it by v_l^N and introduce $\gamma(v_l^N)$, a memory-dependent mobility function describing how memory modulates the local movement rate. The difference between positions x_{l+1} and x_l is then

$$\hat{\gamma}_l^N = \gamma\left(v_{l+1}^N\right) - \gamma\left(v_l^N\right).$$

At time t_N, the probabilities for an individual at position x_l to move left or right are determined by a combination of unbiased random movement and biased movement:

$$\begin{aligned} P_{\text{left}}(x_l) &= p + \hat{\gamma}_{l-1}^N, \\ P_{\text{right}}(x_l) &= p - \hat{\gamma}_l^N, \end{aligned} \tag{8}$$

where $0 < p \leq \frac{1}{2}$ represents the probability of unbiased random movement, and $\hat{\gamma}_i^N$ captures the influence of spatial gradients of accumulated memory on directed movement.

Using the mean value theorem for $\gamma \in C^3(\mathbb{R})$, there exists ξ_i^N between v_i^N and v_{i+1}^N such that

$$\hat{\gamma}_i^N = \gamma(v_{i+1}^N) - \gamma(v_i^N) = \gamma'(\xi_i^N)\,(v_{i+1}^N - v_i^N).$$

From the discrete memory definition (7),

$$v_{i+1}^N - v_i^N = \tau \sum_{k=-\infty}^{N} g^{N-k}\left(u_{i+1}^k - u_i^k\right).$$

Applying a Taylor expansion to u at u_i^k, for small h,

$$u_{i+1}^k - u_i^k = h\, u_x(x_i, t_k) + O(h^2),$$

where $u_x(x_i, t_k)$ denotes the value of the derivative function u_x at (x_i, t_k). As $g \in L^1([0, \infty))$ and $u \in C_b^{2,1}(\bar{\Omega} \times (-\infty, t])$, we have

$$\hat{\gamma}_i^N = O(h).$$

Therefore, for sufficiently small h, we always have

$$0 < P_{\text{left}}(x_l),\ P_{\text{right}}(x_l),\ P_{\text{left}}(x_l) + P_{\text{right}}(x_l) \le 1.$$

The function γ may be either decreasing or increasing. When γ is increasing, a positive gradient in accumulated memory v implies a positive gradient in $\gamma(v)$; when γ is decreasing, the gradients have opposite signs. In the case of an increasing γ, the movement probability (8) implies that animals tend to move away from regions with higher accumulated memory of the population. For example, if the accumulated memory at x_{l+1} exceeds that at x_l, then $\hat{\gamma}_l^N > 0$, it follows that

$$P_{\text{left}}(x_l) > p > P_{\text{right}}(x_l) > 0,$$

which indicates a decreasing tendency of movement toward the right. Conversely, when γ is decreasing, it implies a tendency to move toward regions with higher accumulated memory of the population.

The change in population at position x_l over a short time interval from t_N to $t_{N+1} = t_N + \tau$ is determined by individuals leaving x_l and individuals arriving from x_{l-1} and x_{l+1}:

$$u_l^{N+1} - u_l^N = -P_{\text{left}}(x_l)u_l^N - P_{\text{right}}(x_l)u_l^N + P_{\text{right}}(x_{l-1})u_{l-1}^N + P_{\text{left}}(x_{l+1})u_{l+1}^N. \tag{9}$$

Substituting (8) into (9), we obtain

$$\begin{aligned} u_l^{N+1} - u_l^N &= -\big(p + \hat{\gamma}_{l-1}^N\big)u_l^N - \big(p - \hat{\gamma}_l^N\big)u_l^N + \big(p - \hat{\gamma}_{l-1}^N\big)u_{l-1}^N + \big(p + \hat{\gamma}_l^N\big)u_{l+1}^N \\ &= p\big(u_{l+1}^N + u_{l-1}^N - 2u_l^N\big) \\ &\quad + \big(u_{l+1}^N\big(\gamma(v_{l-1}^N) - \gamma(v_l^N)\big) - u_{l-1}^N\big(\gamma(v_l^N) - \gamma(v_{l-1}^N)\big)\big) \\ &\quad + u_l^N\big(\gamma(v_{l-1}^N) + \gamma(v_{l+1}^N) - 2\gamma(v_l^N)\big). \end{aligned} \tag{10}$$

Note that $v(x_l, t_N) = v_l^N + O(\tau^2)$ when τ is small. It follows from $g \in L^1([0, \infty))$ and $u \in C_b^{2,1}(\bar{\Omega} \times (-\infty, t])$ that

$$v_x(x_l, t_N) = \tau \sum_{k=-\infty}^{N} g^{N-k} u_x(x_l, t_k) + O(\tau^2),$$
$$v_{xx}(x_l, t_N) = \tau \sum_{k=-\infty}^{N} g^{N-k} u_{xx}(x_l, t_k) + O(\tau^2). \tag{11}$$

Here, $u_x(x_l, t_N)$ and $u_{xx}(x_l, t_N)$ denote derivatives of the smooth function $u(x, t)$ evaluated at (x_l, t_N).

Linearizing $u^N_{l\pm1}$ around u^N_l gives

$$u^N_{l\pm1} = u^N_l \pm h\, u_x(x_l, t_N) + \frac{h^2}{2} u_{xx}(x_l, t_N) + O(h^3). \tag{12}$$

Substituting (12) into (10) and keeping terms up to $O(h^2)$, and noting that $\gamma \in C^3(\mathbb{R})$, together with (11), for τ, h are small, we obtain:

$$\begin{aligned}
&u^N_{l+1}\big(\gamma(v^N_{l-1}) - \gamma(v^N_l)\big) - u^N_{l-1}\big(\gamma(v^N_l) - \gamma(v^N_{l-1})\big) + u^N_l\big(2\gamma(v^N_{l-1}) - 2\gamma(v^N_l)\big)\\
=&2h^2\Big[u_x(x_l, t_N)\,\gamma'(v^N_l)\,v_x(x_l, t_N) + u^N_l\big(\gamma''(v^N_l)\,v^2_x(x_l, t_N) + \gamma'(v^N_l)\,v_{xx}(x_l, t_N)\big)\Big]\\
&+O(h^3) + O(\tau^2)\\
=\;& 2h^2\big(u\,(\gamma(v))_x\big)_x\Big|_{(x_l, t_N)} + O(h^3) + O(\tau^2).
\end{aligned}$$

Therefore, the discrete population change at x_l can be expressed as

$$u^{N+1}_l - u^N_l = ph^2 u_{xx}(x_l, t_N) + 2h^2\big(u\,(\gamma(v))_x\big)_x\Big|_{(x_l, t_N)} + O(h^3) + O(\tau^2). \tag{13}$$

Note that

$$u_t(x_l, t_N) = \lim_{\tau\to0} \frac{u^{N+1}_l - u^N_l}{\tau}.$$

Assume that the limit $\lim\limits_{h\to0,\tau\to0} \frac{h^2}{\tau}$ exists. Define the diffusion and taxis coefficients by

$$D = p \lim_{h\to0,\tau\to0} \frac{h^2}{\tau}, \qquad \alpha = 2 \lim_{h\to0,\tau\to0} \frac{h^2}{\tau}.$$

Since (13) holds for any (x_l, t_N), dividing both sides of (13) by τ and taking the limit as $h, \tau \to 0$, we obtain that, for any $x \in \Omega$ and $t > 0$, the population density $u(x, t)$ satisfies

$$u_t = D\, u_{xx} + \alpha\,\big(u\,\gamma(v)_x\big)_x. \tag{14}$$

To further simplify the expression, we introduce the rescaled function

$$\tilde{\gamma}(v) = \frac{2qD}{p}\,\gamma(v).$$

For notational convenience, we drop the tilde and obtain

$$u_t = D\,u_{xx} + \Big(u\,\gamma(v)_x\Big)_x. \tag{15}$$

This is an advection-diffusion type model. The higher-dimensional version is

$$u_t = D\Delta u + \nabla\cdot(u\,\nabla\gamma(v))\,. \tag{16}$$

When $\gamma(v)$ is linear, the model (16) reduces to the simplest advection–diffusion form, $u_t = D\Delta u + a\nabla\cdot\big(u\nabla v\big)$.

In summary, this movement strategy combines random, unbiased movement with biased movement guided by local gradients of certain environmental factors. The accumulated memory represents this factor, and individuals adjust their movement by comparing the gradient of memory between their current location and the arrival site: they move along the gradient if they prefer crowding (e.g., for foraging efficiency, mating opportunities, or shelter) or away from it if they avoid crowded areas (e.g., to reduce intraspecific competition).

3.1.2 Fickian type diffusion

Consider the case where animals share a symmetric movement rate between adjacent sites, so that the probability of an individual moving from x_i to x_{i+1} is equal to the probability of moving from x_{i+1} to x_i. In other words, animals match their movement rates to those of their neighbors: if individuals at x_i have a higher probability of moving right, then those at x_{i+1} will have the same higher probability of moving left.

This common probability depends on environmental factors that influence movement between two locations. Since this movement strategy involves matching movement rates between a site and its neighbor according to local environmental conditions, we refer to it as environment-matching movement. In this work, we consider accumulated memory as the environmental factor.

In general, the movement probability can be modeled as a linear combination of the accumulated memory at the two adjacent positions,

$$a\,\gamma(v_i^N) + b\,\gamma(v_{i+1}^N), \qquad a, b \geq 0,\ a+b=1.$$

where the parameters a and b determine the relative influence of the accumulated memory at the departure site and the arrival site, respectively.

Accordingly, the probabilities for an individual at position x_l to move left or right are

$$P_{\text{left}}(x_l) \;=\; a\,\gamma\Big(v_{l-1}^N\Big) + b\,\gamma\Big(v_l^N\Big), \qquad P_{\text{right}}(x_l) \;=\; a\,\gamma\Big(v_l^N\Big) + b\,\gamma\Big(v_{l+1}^N\Big). \tag{17}$$

To ensure valid probabilities, we require

$$0 < \gamma(v) \le \frac{1}{2(a+b)} = \frac{1}{2},$$

so that

$$0 < P_{\text{left}},\ P_{\text{right}},\ P_{\text{left}} + P_{\text{right}} \le 1.$$

According to the probabilities in (17), when γ is increasing, higher accumulated memory corresponds to a higher movement probability for animals. Conversely, when γ is decreasing, higher accumulated memory leads to a lower probability of movement.

Substituting the movement probabilities (17) into (9), we obtain the population change at location x_l from t_N to t_{N+1}:

$$\begin{aligned} u_l^{N+1} - u_l^N &= -\big(a\gamma(v_{l-1}^N) + b\gamma(v_l^N) + a\gamma(v_l^N) + b\gamma(v_{l+1}^N)\big)u_l^N \\ &\quad + \big(a\gamma(v_{l-1}^N) + b\gamma(v_l^N)\big)u_{l-1}^N + \big(a\gamma(v_l^N) + b\gamma(v_{l+1}^N)\big)u_{l+1}^N \\ &= a\Big(\gamma(v_l^N)\big(u_{l+1}^N - u_l^N\big) - \gamma(v_{l-1}^N)\big(u_l^N - u_{l-1}^N\big)\Big) \\ &\quad + b\Big(\gamma(v_{l+1}^N)\big(u_{l+1}^N - u_l^N\big) - \gamma(v_l^N)\big(u_l^N - u_{l-1}^N\big)\Big). \end{aligned}$$

Applying the Taylor expansions (12), using the convergence result (11) for the first derivative of v, and the C^3-regularity of γ, we obtain, for small τ and h, that

$$u_l^{N+1} - u_l^N = (a+b)h^2\Big(\big(\gamma(v)\,u_x\big)_x\Big)\Big|_{(x_l,t_N)} + O(h^3) + O(\tau^2). \tag{18}$$

Assume $\lim\limits_{h\to 0,\tau\to 0} \frac{h^2}{\tau}$ exists. Divide both sides of (18) by τ, take the limit as $h, \tau \to 0$ and define

$$\alpha = (a+b)\lim_{h\to 0,\tau\to 0}\frac{h^2}{\tau} = \lim_{h\to 0,\tau\to 0}\frac{h^2}{\tau}.$$

Since (18) holds for any (x_l, t_N), it follows that for all $x \in \Omega$ and $t > 0$, the population density $u(x,t)$ satisfies

$$u_t = \alpha\big(\gamma(v)\,u_x\big)_x,$$

with continuous accumulated memory v (2).

For convenience, we introduce the rescaled function $\tilde{\gamma}(v) = \alpha\gamma(v)$, it follows that $0 < \tilde{\gamma}(v) \le \frac{\alpha}{2}$. Dropping the tilde notation for simplicity, the resulting continuum equation is

$$u_t = \big(\gamma(v)\,u_x\big)_x.$$

This is a one-dimensional Fickian type diffusion model. A natural extension to higher dimensions is

$$u_t = \nabla\cdot(\gamma(v)\,\nabla u)\,. \tag{19}$$

In summary, this movement strategy represents individuals adjusting their movement rates based on environmental factors (accumulated memory) at both the departure and arrival sites. Movement probabilities are symmetrically matched between adjacent

sites according to the local accumulated memory at both locations. This matching of movement rates between adjacent sites reduces population spatial heterogeneity.

3.1.3 Fokker–Planck type diffusion

Consider the case where, at any fixed site, the probabilities of moving left and right are equal. This common probability is assumed to depend only on environmental factors at only the current location, and we therefore refer to it as location-based movement, in contrast to the other two cases, which require comparison with adjacent sites. Specifically, we assume it depends on the accumulated memory v_l^N at the current site.

For an individual at position x_l, this gives

$$P_{\text{left}}(x_l) = P_{\text{right}}(x_l) = \gamma(v_l^N). \tag{20}$$

Here, $\gamma(v_l^N)$ represents how the accumulated memory specifically influences movement. If γ is increasing, higher accumulated memory leads to a greater likelihood of leaving the current position, corresponding to memory-enhanced movement. Conversely, if γ is decreasing, higher accumulated memory reduces the movement rate, a behavior often referred to as memory-suppressed movement.

To ensure valid probabilities, we require

$$0 < \gamma(v) \leq \tfrac{1}{2},$$

so that

$$0 < P_{\text{left}},\ P_{\text{right}},\ P_{\text{left}} + P_{\text{right}} \leq 1.$$

Substituting (20) into (9), the population change at x_l from t_N to t_{N+1} is

$$u_l^{N+1} - u_l^N = -2\gamma(v_l^N)\, u_l^N + \gamma(v_{l-1}^N)\, u_{l-1}^N + \gamma(v_{l+1}^N)\, u_{l+1}^N.$$

Applying the Taylor expansions (12), using the convergence result (11), and the C^3-regularity of γ, we obtain

$$\begin{aligned} u_l^{N+1} - u_l^N &= h^2\Big(\gamma(v_l^N)u_{xx}(x_l,t_N) + \gamma''(v_l^N)\left(v_x(x_l,t_N)\right)^2 u_l^N \\ &\quad + \gamma'(v_l^N)\, v_{xx}(x_l,t_N)u_l^N + 2\gamma'(v_l^N)v_x(x_l,t_N)u_x(x_l,t_N)\Big) + O(h^3) + O(\tau^2) \\ &= h^2\Big((\gamma(v)u)_{xx}\Big)\Big|_{(x_l,t_N)} + O(h^3) + O(\tau^2). \end{aligned} \tag{21}$$

Assume $\lim_{h\to 0,\tau\to 0} \frac{h^2}{\tau} = \alpha$. Dividing both sides of (21) by τ and taking the limit $h, \tau \to 0$, since (21) holds for any (x_l, t_N), we obtain, for all $x \in \Omega$ and $t > 0$, $u(x,t)$ satisfies

$$u_t = \alpha\,(\gamma(v)\,u)_{xx},$$

where v denotes the accumulated memory.

Introducing the rescaled function $0 < \tilde{\gamma}(v) = \alpha\gamma(v) \leq \frac{\alpha}{2}$ and dropping the tilde, it becomes

$$u_t = (\gamma(v)\, u)_{xx}.$$

This is a one-dimensional Fokker–Planck type diffusion equation. The higher-dimensional analogue is

$$u_t = \Delta(\gamma(v)\, u)\,. \tag{22}$$

In this movement strategy, the movement probability depends solely on the accumulated memory at the local position, without adjustment relative to neighboring sites, which distinguishes it from the Fickian type diffusion case. Although this strategy assumes that individuals have no directional bias when choosing to move, they assign equal probability to all adjacent locations based on local accumulated memory. However, spatial heterogeneity in this memory field leads to uneven movement rates, so that an effective directional bias arises implicitly from spatial variation in accumulated memory. This illustrates how microscopically unbiased movement can lead to heterogeneous population distributions at the macroscopic level.

For both the Fickian and Fokker–Planck type diffusion models, the function $\gamma \in C^3(\mathbb{R})$ must satisfy $0 < \gamma(v) \leq \frac{1}{2} \lim\limits_{h\to 0,\ \tau\to 0} \frac{h^2}{\tau}$. An increasing γ corresponds to memory-enhanced movement, whereas a decreasing γ corresponds to memory-suppressed movement. In contrast, for the advection–diffusion-type model, no such restriction is imposed; γ may take negative values. In this case, a decreasing γ indicates that animals tend to move toward crowded areas, whereas an increasing γ indicates that they tend to move away from crowded areas.

3.2 Discrete velocity jump process

The velocity jump process provides a microscopic explanation of classical thermodynamics and serves as a classical method for studying the motion of chemical particles, bacteria, or individual animals. For example, (Erban and Othmer 2004) studied bacterial movement in which the turning frequency depends on an intracellular state evolving according to a system of ordinary differential equations. Similarly, (Eftimie et al. 2007) used the velocity jump process to model animal communication mechanisms based on attraction, repulsion, and alignment.

Consider a particle moving along the x-axis at a constant speed c. At random times, the particle reverses direction according to a Poisson process with constant intensity λ. Let $u^+(x,t)$ and $u^-(x,t)$ denote the densities of particles at position x and time t moving to the right $(+)$ and left $(-)$, respectively. In one space dimension, this process is described by the Carleman model (first used in Carleman (1957)):

$$\begin{aligned} \frac{\partial u^+}{\partial t} + c\frac{\partial u^+}{\partial x} &= -\lambda u^+ + \lambda u^-, \\ \frac{\partial u^-}{\partial t} - c\frac{\partial u^-}{\partial x} &= \lambda u^+ - \lambda u^-. \end{aligned} \tag{23}$$

The total population is given by $u = u^+ + u^-$. Formally, this system leads to the diffusion equation $u_t = Du_{xx}$ in the limit $\lim_{c\to\infty,\lambda\to\infty} \frac{c^2}{2\lambda} = D$. This process was studied by (Taylor 1922), (Goldstein 1951), and (Kac 1974). Later, (Hillen and Othmer 2000) and (Othmer and Hillen 2002) formally derived diffusion equations with constant turning frequency from the continuum kinetic equation using a parabolic scaling limit, and provided a rigorous justification of this limit. Model (23) represents the case that individuals switch direction randomly without bias, analogous to heat diffusion. However, in biological contexts, animals often adjust their movement in response to environmental factors (Laca 2008). For nonconstant movement speeds and turning frequencies, movement models with heterogeneity have been extensively analyzed, yielding more diffusion operators (Choi and Kim 2019; Lim et al. 2021).

One of the main goals of diffusion modeling is to understand the connection between macroscopic collective phenomena and microscopic individual behaviors shaped by biological traits. In what follows, we use the discrete velocity jump process to formally derive three classes of diffusion equations and highlight their differences. In particular, we focus on heterogeneity induced by accumulated memory, which represents past population densities. A higher accumulated memory indicates regions where the population was relatively high in the past. This accumulated memory (2) is modulated by $\gamma(v)$, which represents how sensitively animals respond to accumulated memory.

In the following analysis, we assume $c(v) \in C^1(\mathbb{R})$ and $\lambda(v) \in C^0(\mathbb{R})$ to be bounded and bounded away from zero. These positivity and smoothness conditions ensure the existence of a well-defined diffusion limit of the jump process (Lim et al. 2021).

3.2.1 Advection–Diffusion Type model

We consider the turning frequency to be non-constant, consisting of two components: an unbiased random turning frequency and a biased turning frequency that depends on the gradient of the accumulated memory. Let the spatial dimension be $n \in \mathbb{N}$. The total population is defined as

$$u := \sum_{k=1}^{n} u^k,$$

where u^k is the total population moving in the kth direction. Each directional population is decomposed as

$$u^k := u^{k+} + u^{k-},$$

with u^{k+} and u^{k-} denoting the population moving along or opposite to the unit vector $\vec{e}_k$, for $k = 1, 2, \ldots, n$.

We define the directional and total fluxes as

$$j^k := \frac{c}{\varepsilon}\left(u^{k+} - u^{k-}\right), \qquad \vec{j} := \sum_{k=1}^{n} j^k \vec{e}_k.$$

Here, $\vec{e}_k$ denotes the unit vector in the kth direction. The gradient of accumulated memory leads to distinct turning frequencies for forward- and backward-moving individuals u^{k+} and u^{k-}, denoted by λ^{k+} and λ^{k-}:

$$\lambda^{k\pm} = \lambda_1 \pm \lambda_2 \partial_{x^k}(\gamma(v)), \tag{24}$$

where $\lambda_1 > 0$ represents the unbiased random turning rate, and $\lambda_2 > 0$ measures the strength of the bias in the turning frequency induced by the gradient $\partial_{x^k}(\gamma(v)) = \frac{\partial \gamma(v)}{\partial x^k}$. Here ∂_{x^k} denotes the derivative in the k-th coordinate. Here, $v = g *_t u$ denotes the accumulated memory, which reflects historical population density. A higher accumulated memory corresponds to locations that were more crowded in the past. The function $\gamma(v)$, which may be linear or nonlinear, modulates how accumulated memory influences the biased turning frequency.

When γ is increasing, the gradient of $\gamma(v)$ has the same sign as the gradient of the accumulated memory v, corresponding to animals attracted to crowded regions. For example, if $\partial_{x^k} v > 0$, then $\partial_{x^k}\gamma(v) > 0$, which gives

$$\lambda^{k+} > \lambda_1 > \lambda^{k-} > 0.$$

This implies that forward-moving individuals, who are moving toward more crowded regions, turn more frequently, while backward-moving individuals turn less.

In contrast, if γ is decreasing, the gradients of $\gamma(v)$ and v have opposite signs, corresponding to animals that avoid crowded regions. For example, if $\partial_{x^k} v > 0$, then $\partial_{x^k}\gamma(v) < 0$, implying

$$0 < \lambda^{k+} < \lambda_1 < \lambda^{k-}.$$

In this case, forward-moving individuals entering regions with higher accumulated memory have a lower turning frequency, while backward-moving individuals have a higher turning frequency.

Considering the biased turning frequency $\lambda^{k\pm}$ (24) and assuming constant movement speed c, the kinetic equations (23) under parabolic scaling are, for $k = 1, \ldots, n$:

$$\partial_t u^{k+} + \frac{c}{\varepsilon}\partial_{x^k} u^{k+} = -\frac{1}{\varepsilon^2}\big(\lambda_1 + \lambda_2 \partial_{x^k}\gamma(v)\big)u^{k+} + \frac{1}{2n\varepsilon^2}\sum_{l=1\pm}^{n\pm}\big(\lambda_1 + \lambda_2 \partial_{x^l}\gamma(v)\big)u^l, \tag{25}$$

$$\partial_t u^{k-} - \frac{c}{\varepsilon}\partial_{x^k} u^{k-} = -\frac{1}{\varepsilon^2}\big(\lambda_1 - \lambda_2 \partial_{x^k}\gamma(v)\big)u^{k-} + \frac{1}{2n\varepsilon^2}\sum_{l=1\pm}^{n\pm}\big(\lambda_1 + \lambda_2 \partial_{x^l}\gamma(v)\big)u^l, \tag{26}$$

where ε is the diffusion-scale singular limit parameter.

Adding (25) and (26) yields the exact conservation law for each direction:

$$\partial_t u^k + \partial_{x^k} j^k = -\frac{\lambda_1}{\varepsilon^2}u^k - \frac{\lambda_2}{c\varepsilon} j^k \partial_{x^k}\gamma(v) + \frac{\lambda_1}{n\varepsilon^2}u + \frac{\lambda_2}{cn\varepsilon}\sum_{l=1}^{n} j^l \partial_{x^l}\gamma(v),$$

and summing over k gives the global conservation law

$$\partial_t u + \nabla \cdot \vec{j} = 0. \tag{27}$$

Subtracting (26) from (25) gives

$$\frac{\varepsilon}{c}\frac{\partial j^k}{\partial t} + \frac{c}{\varepsilon}\partial_{x^k} u^k = -\frac{\lambda_1}{\varepsilon c} j^k - \frac{\lambda_2}{\varepsilon^2} u^k \partial_{x^k}\gamma(v),$$

or equivalently,

$$j^k + \frac{\varepsilon^2}{\lambda_1}\frac{\partial j^k}{\partial t} + \frac{c^2}{\lambda_1}\partial_{x^k} u^k = -\frac{\lambda_2 c}{\lambda_1 \varepsilon} u^k \partial_{x^k}\gamma(v).$$

We choose $\lambda_2 = \varepsilon\tilde{\lambda}_2$ so that $\frac{\lambda_2 c}{\lambda_1 \varepsilon} = \frac{\tilde{\lambda}_2 c}{\lambda_1} = O(1)$ as $\varepsilon \to 0$. This scaling represents the only nontrivial regime in which the small per-turn bias due to the memory gradient builds up over diffusive time scales, generating a finite macroscopic drift comparable to diffusion. Under this scaling, the leading-order flux is

$$j^k = -\frac{\lambda_2 c}{\lambda_1 \varepsilon} u^k \partial_{x^k}\gamma(v) - \frac{c^2}{\lambda_1}\partial_{x^k} u^k + O(\varepsilon^2),$$

and substituting into (27) yields

$$\frac{\partial u}{\partial t} = \frac{c^2}{\lambda_1}\Delta u + \frac{\lambda_2 c}{\varepsilon \lambda_1}\nabla \cdot (u \nabla \gamma(v)) + O(\varepsilon^2).$$

Taking the limit $\varepsilon \to 0$ and defining $K = \lim_{\varepsilon\to 0} \frac{\lambda_2 c}{\varepsilon \lambda_1}$ and $D = \lim_{\varepsilon \to 0} \frac{c^2}{\lambda_1}$ with $D, K > 0$, we formally obtain the advection–diffusion equation

$$\frac{\partial u}{\partial t} = D\Delta u + K\nabla \cdot (u \nabla \gamma(v)).$$

For simplicity, define $\tilde{\gamma}(v) = K\gamma(v)$ and drop the tilde:

$$\frac{\partial u}{\partial t} = D\Delta u + \nabla \cdot (u \nabla \gamma(v)).$$

Here, the first term represents unbiased random movement, while the second term represents directed movement driven by the gradient of accumulated memory.

Remark 3.1 This derivation relies on the assumption $\frac{\lambda_2 c}{\lambda_1 \varepsilon} = O(1)$, which ensures that both diffusion and memory-driven advection contribute at leading order. If $\frac{\lambda_2 c}{\lambda_1 \varepsilon} = O(\varepsilon)$, the drift term vanishes as $\varepsilon \to 0$, yielding pure diffusion: $j^k = -\frac{c^2}{\lambda_1}\partial_{x^k} u^k +$

$O(\varepsilon)$, which leads to $u_t = D\Delta u$. If $\dfrac{\lambda_2 c}{\lambda_1 \varepsilon} = O\left(\frac{1}{\varepsilon}\right)$, the drift dominates, producing a singular limit in which the leading-order algebraic equation enforces constraints $u^k \partial_{x^k} \gamma = 0 + O(\varepsilon)$, which is either unphysical or requires a different rescaling.

3.2.2 Fickian- and fokker–planck-type diffusion models

Inspired by (Lim et al. 2021), we consider the case where both the speed $c(v)$ and the turning frequency $\lambda(v)$ are positive functions of the accumulated memory v, rather than constants. We assume $c(v) \in C^1(\mathbb{R})$. Moreover, assuming that the discrete velocities are symmetric and that u^{k+} and u^{k-} share the same turning frequency $\lambda(v)$, the discrete-velocity kinetic system (generalising (23)) in n directions takes the form

$$\frac{\partial u^{k+}}{\partial t} + \frac{1}{\varepsilon}\frac{\partial}{\partial x^k}\big(c(v)u^{k+}\big) = \frac{1}{2n\varepsilon^2}\lambda(v)\sum_{\ell=1\pm}^{n\pm}\big(u^{\ell} - u^{k+}\big), \tag{28}$$

$$\frac{\partial u^{k-}}{\partial t} - \frac{1}{\varepsilon}\frac{\partial}{\partial x^k}\big(c(v)u^{k-}\big) = \frac{1}{2n\varepsilon^2}\lambda(v)\sum_{\ell=1\pm}^{n\pm}\big(u^{\ell} - u^{k-}\big), \tag{29}$$

for $k = 1, \ldots, n$.

Define the directional and total populations, as well as the directional and total fluxes by

$$u^k := u^{k+} + u^{k-}, \qquad u := \sum_{k=1}^{n} u^k, \qquad j^k := \frac{c(v)}{\varepsilon}\left(u^{k+} - u^{k-}\right), \qquad \vec{j} := \sum_{k=1}^{n} j^k \vec{e}_k.$$

Adding (28) and (29) gives

$$\partial_t u^k + \partial_{x^k} j^k = \frac{1}{n\varepsilon^2}\lambda(v)u - \frac{1}{\varepsilon^2}\lambda(v)\left(u^{k+} + u^{k-}\right), \tag{30}$$

and summing over n directions yields the global conservation law

$$\partial_t u + \nabla\cdot\vec{j} = 0. \tag{31}$$

Subtracting (29) from (28), we obtain

$$\frac{\varepsilon}{c(v)}\frac{\partial j^k}{\partial t} + \frac{1}{\varepsilon}\partial_{x^k}\big(c(v)u^k\big) = -\frac{\lambda(v)}{c(v)\varepsilon}j^k + \frac{\varepsilon}{(c(v))^2}\frac{\partial c(v)}{\partial t}j^k. \tag{32}$$

We expand the directional populations and fluxes in powers of ε as

$$u^{k\pm} = u_0^{k\pm} + \varepsilon u_1^{k\pm} + O(\varepsilon^2), \qquad j^k = j_0^k + \varepsilon j_1^k + O(\varepsilon^2).$$

At leading order, (32) gives

$$\partial_{x^k}\big(c(v)u_0^k\big) = -\frac{\lambda(v)}{c(v)}\, j_0^k,$$

so that

$$j^k = -\frac{c(v)}{\lambda(v)}\, \partial_{x^k}\big(c(v)u_0^k\big) + O(\varepsilon).$$

It follows that

$$\vec{j} = -\frac{c(v)}{\lambda(v)}\, \nabla\big(c(v)u_0\big) + O(\varepsilon). \tag{33}$$

Inserting (33) into (31) yields

$$\partial_t u = \nabla\cdot\Big(\frac{c(v)}{\lambda(v)}\, \nabla\big(c(v)u_0\big)\Big) + O(\varepsilon).$$

Thus, in the limit $\varepsilon \to 0$, the macroscopic equation is

$$\partial_t u = \nabla\cdot\Big(\frac{c(v)}{\lambda(v)}\, \nabla\big(c(v)u\big)\Big). \tag{34}$$

The movement behavior of individuals is determined by both the speed $c(v)$ and the turning frequency $\lambda(v)$. The speed $c(v)$ describes how fast individuals move. If $c(v)$ increases with accumulated memory v, then individuals in high-memory regions move faster and leave those areas more quickly, while those in low-memory regions move more slowly and remain longer. This smooths out population differences in spatial heterogeneity. Conversely, if $c(v)$ decreases with v, individuals move more slowly in high-memory regions and spend more time there, while leaving low-memory regions more quickly. This effect strengthens spatial aggregation of the population. The turning frequency $\lambda(v)$ describes how often individuals change their movement direction and reflects the randomness of movement. A lower $\lambda(v)$ means individuals are more likely to continue moving in the same direction, while a higher $\lambda(v)$ makes their motion more random. The combination of speed and turning frequency leads to different overall movement patterns. For example, if speed is high but turning frequency is low, individuals move quickly and in a persistent direction, resulting in strong directed movement. If speed is low but turning frequency is high, individuals move slowly and turn often, so they mostly remain in the same area but wander around randomly.

- **Fokker–Planck type diffusion.** If $\lambda(v) = c(v) = \gamma(v) > 0$, then (34) reduces to

$$u_t = \Delta\big(\gamma(v)u\big),$$

a Fokker–Planck type diffusion equation. Here the movement speed $c(v)$ and turning frequency $\lambda(v)$ are identical. If $\gamma(v)$ is decreasing, individuals in regions with higher accumulated memory move more slowly and turn less frequently, so they remain longer in crowded areas. This corresponds to memory-suppressed

movement. If $\gamma(v)$ is increasing, individuals in high-memory regions move faster and turn more frequently, which corresponds to memory-enhanced movement.

- **Fickian type diffusion.** If $c(v) = 1$ and $\lambda(v) = 1/\gamma(v) > 0$, then (34) becomes

$$u_t = \nabla \cdot \big(\gamma(v)\,\nabla u\big).$$

In the Fickian type diffusion case, individuals move at a constant speed, and only the turning frequency depends on the accumulated memory through $1/\gamma(v)$. When $\gamma(v)$ increases, individuals in high-memory regions turn less frequently, so they persist in their direction and spread out more efficiently. Conversely, when $\gamma(v)$ decreases in high-memory regions, individuals turn more often, move more randomly, and "wiggle" around, which slows their spread.

4 Well-posedness

Most memory-related movement studies are formulated under the diffusion–advection framework

$$u_t = D\Delta u + \chi \nabla \cdot (u \nabla v), \tag{35}$$

where v represents memory, with constant diffusion D and taxis coefficient χ. When v corresponds to a fixed delayed memory, i.e., $v(x,t) = u(x,t-\tau)$, (Shi et al. 2020) proved the existence of solutions. Further work on fixed-delay memory models can be found in (An et al. 2020; Li et al. 2022, 2023; Shi et al. 2019, 2021a; Song et al. 2021a, 2022, 2019; Wang et al. 2022a). However, for accumulated memory where

$$v(x,t) = g *_t u = \int_{-\infty}^{t} g(t-s)u(x,s)\,\mathrm{d}s,$$

and $g \in L^1([0,\infty))$, the existence of solutions to (35) has not yet been established. Nevertheless, (Shi 2025; Ji et al. 2024) assumed the existence of solutions to study spatially nonconstant steady states. For fixed temporal delay with nonlocal spatial perception (Xue et al. 2024), i.e.,

$$v(x,t) = g *_x u(x,t-\tau) = \int_{\Omega} g(x-y)u(y,t-\tau)\,\mathrm{d}y,$$

existence of solutions has been proved, with additional studies in (Song et al. 2024; Wang et al. 2025). Many studies focus on purely spatial nonlocal perception (not memory),

$$v = g *_x u = \int_{\Omega} g(x-y)u(y,t)\,\mathrm{d}y,$$

as in (Giunta et al. 2022a; Salmaniw et al. 2025; Painter et al. 2024; Giunta et al. 2022b; Liu et al. 2025a; Ducrot et al. 2018). Shi et al. (Shi et al. 2021b) and Song et

al. (Song et al. 2021b) studied bifurcations from constant equilibrium for models with both nonlocal spatial and accumulated memory, i.e.,

$$v = \int_{-\infty}^{t} \int_{\Omega} G(x, y, t-s) g(t-s) u(y, s) \, \mathrm{d}y \, \mathrm{d}s.$$

Note that all the above-mentioned models fall under the (35) framework. For the other two types of diffusion models, Fokker–Planck diffusion and Fickian type diffusion, to the best of our knowledge, no studies have investigated memory-influenced movement.

To bridge these gaps, we prove local existence for the general model (5), which covers advection–diffusion (1), Fokker–Planck (3), and Fickian type diffusion models (4), where both the diffusion and taxis coefficients can be nonlinear and depend on memory. To further enhance generality and identify the minimal requirements on the temporal kernel, we consider a general measure μ rather than a specific kernel function g. Let $\Omega \in \mathbb{R}^n$ be a bounded domain with smooth boundary. Then we rewrite the general model (5) into

$$\begin{cases} u_t = \nabla \cdot (D(\mu *_t u) \nabla u + A(\mu *_t u) u \nabla (\mu *_t u)), & x \in \Omega, \ t > 0, \\ \mathcal{B}u = 0, & x \in \partial\Omega, \ t > 0, \\ u(x, t) = \phi(x, t), & x \in \bar{\Omega}, \ t \le 0, \end{cases} \tag{36}$$

where $\mathcal{B}u = \beta_1(x)\partial_{\bar{\nu}} u + \beta_2(x)u$, with $\beta_1, \beta_2 \in C^2(\partial\Omega)$ and $\beta_1^2 + \beta_2^2 > 0$.

For the analysis in this chapter, we always assume the following conditions on D, A, and μ:

(A1) $D, A \in C^3([0, \infty))$ and $\inf_{s \ge 0} D(s) > 0$.
(A2) $\mu \in \mathcal{M}$, the space of finite Borel measures on $[0, \infty)$ and $\mu(\{0\}) = 0$.
(A3) The nonnegative initial condition ϕ satisfies

$$\phi \in C^{2+\theta, 1+\theta/2}(\bar{\Omega} \times (-\infty, 0]), \qquad \theta \in (0, 1),$$

and

$$\mathcal{B}\phi = 0 \quad \text{on } \partial\Omega \times (-\infty, 0].$$

Assumption (A2) accommodates both $L^1([0, \infty))$ kernels and discrete delays. If μ is absolutely continuous with respect to Lebesgue measure, for example, when μ has a density $g \in L^1([0, \infty))$ (such as the Gamma kernel (50)), then

$$(\mu *_t u)(x, t) = \int_{-\infty}^{t} g(t-r) \, u(x, r) \, \mathrm{d}r = \int_{0}^{\infty} g(r) \, u(x, t-r) \, \mathrm{d}r.$$

If instead μ is a Dirac measure, say $\mu = \delta_\tau$ with $\tau > 0$, then

$$(\mu *_t u)(x, t) = \int_{0}^{\infty} u(x, t-r) \, \mathrm{d}\mu(r) = u(x, t-\tau).$$

For $T > 0$, we denote $\Omega_T := \Omega \times (-\infty, T]$, and $C^{\alpha,\beta}(\Omega_T) := C^{\alpha}(\Omega) \cap C^{\beta}([0, T])$. Denote total variation of μ as $\|\mu\|_{\mathcal{M}}$, i.e. $\|\mu\|_{\mathcal{M}} = \mu([0, \infty))$.

Lemma 4.1 *Let $\mu \in \mathcal{M}$ and let $v \in C^{\theta,\frac{\theta}{2}}(\Omega_T)$ for some $\theta \in [0, 1)$. Suppose the support of v is contained in $\Omega \times I$, where $I = (-\infty, T]$ or $I = [0, T]$. Then*

$$\|\mu *_t v\|_{C^{\theta,\frac{\theta}{2}}(\Omega\times I)} \leq \mu(T - I)\, \|v\|_{C^{\theta,\frac{\theta}{2}}(\Omega\times I)},$$

where

$$T - I := \{T - s : s \in I\}.$$

Proof The case $\theta = 0$ is immediate. Suppose $0 < \theta < 1$. For any $x, y \in \Omega$ and $t, s \in [0, T]$ with $t \geq s$, we consider two cases. First, if $I = (-\infty, T]$, then

$$\begin{aligned}
\big|(\mu *_t v)(x, t) - (\mu *_s v)(y, s)\big| &= \left| \int_0^{\infty} v(x, t - r)\, \mathrm{d}\mu(r) - \int_0^{\infty} v(y, s - r)\, \mathrm{d}\mu(r) \right| \\
&\leq [v]_{C^{\theta,\frac{\theta}{2}}(\Omega_T)} \int_0^{\infty} \mathrm{d}\mu(r) \Big(|x - y|^{\theta} + |t - s|^{\theta/2} \Big) \\
&= \|\mu\|_{\mathcal{M}}\, [v]_{C^{\theta,\frac{\theta}{2}}(\Omega_T)} \Big(|x - y|^{\theta} + |t - s|^{\theta/2} \Big).
\end{aligned}$$

If the support $I = [0, T]$, then $v = 0$ for all $t < 0$, and the integral reduces to an integration over $[0, T]$. By a similar argument, one obtains

$$\big|(\mu *_t v)(x, t) - (\mu *_s v)(y, s)\big| \leq \mu([0, T])\, [v]_{C^{\theta,\frac{\theta}{2}}(\Omega\times[0,T])} \Big(|x - y|^{\theta} + |t - s|^{\theta/2} \Big).$$

Hence, we conclude that

$$[\mu *_t v]_{C^{\theta,\frac{\theta}{2}}(\Omega\times I)} \leq \mu(T - I)\, [v]_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}.$$

Therefore, combining this with the $C^0(\Omega \times I)$ norm of $\mu *_t v$, we obtain

$$\|\mu *_t v\|_{C^{\theta,\frac{\theta}{2}}(\Omega\times I)} \leq \mu(T - I)\, \|v\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}.$$

□

Remark 4.2 When $I = [0, T]$, $T - I = [0, T]$; when $I = (-\infty, T]$, $T - I = [0, \infty)$. Because μ is a finite Borel measure (see **(A2)**), in either case we have

$$\mu(T - I) \leq \mu([0, \infty)) < \infty.$$

We are ready to prove the local existence of (36).

Theorem 4.3 (Local existence) *Let* $0 < \theta < 1$, *and suppose that* $(A1) - (A3)$ *hold. Then there exists* $T_{\max} > 0$ *and unique nonnegative* $u \in C^{2+\theta,1+\frac{\theta}{2}}(\bar{\Omega}_{T_{\max}})$ *that solves the system* (36). *In addition, we have*

$$\text{either } T_{\max} = \infty \text{ or } \|u\|_{C^{2+\theta,1+\frac{\theta}{2}}(\bar{\Omega}_T)} \to \infty \text{ as } T \to T_{\max} \text{ for some } \gamma \in (0,1). \tag{37}$$

Proof Fix $R \in \left(\|\phi\|_{C^{2+\theta,1+\frac{\theta}{2}}(\Omega_0)}, \infty \right)$ and $\theta \in (0,1)$. For any $T > 0$, define

$$K_T = \left\{ u : \|u\|_{C^{2+\theta,1+\frac{\theta}{2}}(\bar{\Omega}_T)} \le R, u \ge 0 \right\}.$$

For $v \in K_T$, consider the solution u to the following equation

$$\begin{cases} u_t = \nabla \cdot (D(\mu *_t v)\nabla u + A(\mu *_t v)u\nabla(\mu *_t v)), & x \in \Omega, t > 0, \\ \mathcal{B}u = 0, & x \in \partial\Omega, t > 0, \\ u(x,t) = \phi(x,t), & x \in \bar{\Omega}, t \le 0. \end{cases} \tag{38}$$

By (Proposition 7.3.3 (Lunardi 2012)), there exists a unique classical solution defined on $\bar{\Omega}_{T_1}$ for some $T_1 > 0$ independent of $v \in K_T$.

By Lemma 4.1 and assumption (A2), we have $\mu *_t v \in C^{2+\theta,1+\frac{\theta}{2}}(\Omega_T)$. Moreover, by assumption (A3), $D(\mu *_t v)$, $A(\mu *_t v)$, $\nabla(\mu *_t v) \in C^{1+\theta,1+\frac{\theta}{2}}(\Omega_T)$. By the parabolic Schauder estimate (Theorem 4.31 of (Lieberman 1996)), $\exists\ T_0 \in (0, T_1]$ such that $u \in K_{T_0}$. Define a map $S : K_{T_0} \to K_{T_0}$ by $S(v) = u$. S can also be regarded as a map $K_T \to K_T$ for all $T \le T_0$.

We claim that S is a contraction in K_T if T is sufficiently small. Let $T \in (0, T_0]$ and take $v^{[1]}, v^{[2]} \in K_T$. Denote $u^{[i]} = S\left(v^{[i]}\right)$, $w^{[i]} = \nabla\left(\mu *_t v^{[i]}\right)$, $i = 1, 2$ and $\tilde{u} = u^{[1]} - u^{[2]}$, $\tilde{v} = v^{[1]} - v^{[2]}$, $\tilde{w} = w^{[1]} - w^{[2]}$. By calculation, $\tilde{u}$ satisfies the equation

$$\begin{cases} \tilde{u}_t = \nabla \cdot (D(\mu *_t v^{[1]})\nabla\tilde{u}) + \nabla \cdot \left(A(\mu *_t v^{[1]})w^{[1]}\tilde{u}\right) + F(x,t), & x \in \Omega, t > 0, \\ \mathcal{B}u = 0, & x \in \partial\Omega, t > 0, \\ \tilde{u}(x,t) = 0, & x \in \bar{\Omega}, t \le 0, \end{cases}$$

where

$$\begin{aligned} F(x,t) = {} & \nabla \cdot \left((D(\mu *_t v^{[1]}) - D(\mu *_t v^{[2]}))\nabla u^{[2]} \right) + \nabla \cdot \left(A(\mu *_t v^{[2]})u^{[2]}\tilde{w} \right) \\ & + \nabla \cdot \left((A(\mu *_t v^{[1]}) - A(\mu *_t v^{[2]}))u^{[2]}w^{[1]} \right). \end{aligned}$$

Since $v^{[1]}, v^{[2]} \in K_T$, $0 \le \mu *_t v^{[1]}, \mu *_t v^{[2]} \le \|\mu\|_{\mathcal{M}} R$. Let $\|D\|_{C^3([0,\|\mu\|_{\mathcal{M}}R])}$, $\|A\|_{C^3([0,\|\mu\|_{\mathcal{M}}R])} \le M$. By the Schauder estimate (Theorem 4.31 of (Lieberman 1996)), there exists C_1 dependent on T, M, $\left\| D(\mu *_t v^{[1]}) \right\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}$ and

$\left\|A(\mu *_t v^{[1]})w^{[1]}\right\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}$ such that

$$\|\tilde{u}\|_{C^{2+\theta,1+\frac{\theta}{2}}(\Omega_T)} \le C_1 \|F\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}. \tag{39}$$

By the C^3 regularity of D, we have

$$\begin{aligned}
&\left\|\nabla\cdot\left((D(\mu *_t v^{[1]}) - D(\mu *_t v^{[2]}))\nabla u^{[2]}\right)\right\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}\\
&\le \|D(\mu *_t v^{[1]}) - D(\mu *_t v^{[2]})\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)} \|u^{[2]}\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)}\\
&\le R\|D(\mu *_t v^{[1]}) - D(\mu *_t v^{[2]})\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)}\\
&\le R\int_{\mu *_t v^{[2]}}^{\mu *_t v^{[1]}} \left\|D'(s)\right\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)} \mathrm{d}s\\
&= R\int_0^1 \left\|D'(v^s)(\mu *_t \tilde{v})\right\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)} \mathrm{d}s\\
&\le R\int_0^1 \left\|D'(v^s)\right\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)} \mathrm{d}s \|\mu *_t \tilde{v}\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)},
\end{aligned}$$

where $v^s = s\mu *_t \tilde{v} + \mu *_t v^{[2]}$. For $t \in [0, T]$ and $x \in \Omega$, we have

$$\mu *_t v(x,t) = \int_0^T v(x, t-r)\,\mathrm{d}\mu(r) + \int_T^\infty \phi(x, t-r)\,\mathrm{d}\mu(r).$$

As $v^{[1]}$ and $v^{[2]}$ share the same initial condition, it follows that

$$\|\mu *_t \tilde{v}\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)} = \left\|\int_0^T \tilde{v}(x, t-r)\,\mathrm{d}\mu(r)\right\|_{C^{\theta,\frac{\theta}{2}}(\Omega\times[0,T])}.$$

By Lemma 4.1, with $\operatorname{supp} v = \Omega \times [0, T]$, we obtain

$$\|\mu *_t \tilde{v}\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)} \le \mu([0,T])\,\|\tilde{v}\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega\times[0,T])}.$$

Moreover, for the integral term, we have

$$\begin{aligned}
&\int_0^1 \left\|D'(v^s)\right\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)} \mathrm{d}s\\
&= \int_0^1 \left\|D'(v^s)\right\|_{C([0,T])} \mathrm{d}s + \int_0^1 \left\|D''(v^s)\nabla v^s\right\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)} \mathrm{d}s\\
&\le M + \int_0^1 \left\|D''(v^s)\right\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)} \mathrm{d}s \left(\|\mu *_t \nabla\tilde{v}\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)} + \|\mu *_t \nabla v^{[2]}\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}\right).
\end{aligned}$$

Since $v^{[2]}$ and $\tilde{v}$ belong to $C^{2+\theta,1+\frac{\theta}{2}}(\Omega_T)$, by Lemma 4.1, both $\|\mu *_t \nabla\tilde{v}\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}$ and $\|\mu *_t \nabla v^{[2]}\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}$ are bounded by $\|\mu\|_{\mathcal{M}}R$. Using (A1) regularity of D, we deduce that $D''(v^s) \in C^{\theta,\frac{\theta}{2}}(\Omega_T)$ for $s \in [0,1]$. Therefore, there exists a constant $c_1 > 0$ such that

$$\int_0^1 \left\| D'(v^s) \right\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)} \mathrm{d}s \le c_1.$$

Combining the above, we conclude that

$$\begin{aligned}&\|\nabla\cdot\left((D(\mu *_t v^{[1]}) - D(\mu *_t v^{[2]}))\nabla u^{[2]}\right)\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}\\ &\le c_1 R\mu([0,T])\,\|\tilde{v}\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega\times[0,T])}.\end{aligned}$$

By a similar argument, for $A \in C^3[0,T]$, we can show

$$\begin{aligned}&\|\nabla\cdot\left((A(\mu *_t v^{[1]}) - A(\mu *_t v^{[2]}))u^{[2]}w^{[1]}\right)\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}\\ &\le c_2 R\mu([0,T])\,\|\tilde{v}\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega\times[0,T])},\end{aligned}$$

for some constant c_2. By Lemma 4.1 and the condition (A1), we have

$$\begin{aligned}\left\|\nabla\cdot\left(A(\mu *_t v^{[2]})u^{[2]}\tilde{w}\right)\right\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)} &\le \left\|A(\mu *_t v^{[2]})u^{[2]}\tilde{w}\right\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)}\\ &\le \left\|A(\mu *_t v^{[2]})\right\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)}\left\|u^{[2]}\right\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)}\\ &\quad \|\tilde{w}\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)}\\ &\le c_3 R\mu([0,T])\,\|\tilde{v}\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega\times[0,T])},\end{aligned}$$

where $c_3 = M(1 + R\,\|\mu\|_{\mathcal{M}})$. Similar estimates hold for $\left\|D(\mu *_t v^{[1]})\right\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}$, $\left\|A(\mu *_t v^{[1]})w^{[1]}\right\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}$, and $\left\|\nabla\cdot\left(A(\mu *_t v^{[1]})w^{[1]}\right)\right\|_{C^{\theta,\frac{\theta}{2}}(\Omega_T)}$. Therefore, C_1 depends only on T, M, $\|\mu\|_{\mathcal{M}}$, and R.

Now we continue with (7) to get

$$\|\tilde{u}\|_{C^{2+\theta,1+\frac{\theta}{2}}(\Omega_T)} \le C_1(c_1+c_2+c_3)R\mu([0,T])\,\|\tilde{v}\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega\times[0,T])}.$$

Since $\mu \in \mathcal{M}$, by choosing $T > 0$ sufficiently small, $\mu([0,T])$ can be arbitrarily small. Hence the constant multiplying $\|\tilde{v}\|_{C^{1+\theta,\frac{\theta}{2}}(\Omega_T)}$ is < 1, so S is a contraction on K_T. Therefore, S is a contraction when T is small enough. By Banach fixed point theorem, there exists a unique $u \in K_T$, such that $S(u) = u$. That is, u is the unique solution in $[0,T]$. We deduce from (38) and the nonnegative initial data u_0 with $u_0 \not\equiv 0$ that the fixed point u is nonnegative by using the maximum principle. By the standard argument, the solution u can be extended up to some $T_{\max} > 0$. Since the above choice of T depends only on $\|u_0\|_{C^{2+\theta}(\bar{\Omega})}$, the conclusion (37) follows. □

Remark 4.4 Under Assumption (A1), the advection–diffusion and Fokker–Planck models require $\gamma \in C^4(\mathbb{R})$, whereas the Fickian-type diffusion model requires $\gamma \in C^3(\mathbb{R})$ (see relationship between $D(v)$ and $\gamma(v)$ in Section 1 after (5)).

Typical examples of $\gamma \in C^k([0,\infty))$ include power-law decay $\gamma(v) = a + \frac{1}{(c+v)^b}$ ($a \geq 0$, $b, c > 0$), Holling type II or III responses $\gamma(v) = \frac{v^a}{c+v^a}$ ($c > 0$, $a \geq 1$), exponential decay $\gamma(v) = e^{-v}$, and unbounded growth $\gamma(v) = v$ or $\gamma(v) = e^v$.

Remark 4.5 We have proved existence under the general homogeneous boundary condition $\mathcal{B}u = 0$. In the modeling perspective, the homogeneous Dirichlet boundary condition represents a hostile boundary where animals die on the boundary. In this case, the total population, however, goes to 0 as $t \to \infty$ unless it has population dynamics. By including a suitable reaction term, such as $f(x,u) \in C^{1,3}\left(\mathbb{R}; \bar{\Omega} \times \mathbb{R}\right)$, which satisfies the condition that there exists a constant $\bar{u}$ such that $f(x,u) < 0$ for any $x \in \Omega$ when $u > \bar{u}$, and $f(x,u) \geq 0$ when $0 \leq u < \bar{u}$, local existence can also be established using a similar argument. Moreover, global existence is expected to hold if f is bounded.

Remark 4.6 To accurately observe the movement behavior of animals, it is preferable to exclude population dynamics (birth and death) and ensure that the total number of animals within the domain remains conserved, i.e.,

$$\int_\Omega u(x,t)\,dx = \int_\Omega \phi(x,0)\,dx.$$

This corresponds to the homogeneous Neumann boundary condition:

$$\nabla u \cdot \vec{\nu} = 0 \quad \text{on } \partial\Omega \times (0,T],$$

where $\vec{\nu}$ is the outward normal vector to the boundary $\partial\Omega$. This boundary condition ensures that no flux of u crosses the boundary. For the following simulation analysis in Section 6, we will use homogeneous Neumann boundary condition.

Remark 4.7 As we have seen from the above proof, it is possible to observe finite time blow-up from the equation (36). For example, if we take $\mu(s) = \frac{1}{\tau}e^{-s/\tau}$, then system (36) with homogeneous Neumann boundary condition becomes

$$\begin{cases} u_t = \nabla \cdot (D(v)\nabla u + A(v)u\nabla v), & x \in \Omega, t > 0, \\ \tau v_t = u - v, & x \in \Omega, t > 0, \\ \partial_{\vec{\nu}} u = 0, & x \in \partial\Omega, t > 0, \\ u(x,t) = \phi(x,t),\ v(x,t) = \dfrac{1}{\tau}\displaystyle\int_{-\infty}^{t} e^{-(t-s)/\tau}\phi(x,s)\,ds, & x \in \bar{\Omega}, t \leq 0. \end{cases} \tag{40}$$

If we incorporate diffusion $\varepsilon\Delta v$ in the second equation, it reads

$$\begin{cases} u_t = \nabla\cdot(D(v)\nabla u + A(v)u\nabla v), & x\in\Omega, t>0,\\ \tau v_t = \varepsilon\Delta v + u - v, & x\in\Omega, t>0,\\ \partial_{\vec{\nu}} u = \partial_{\vec{\nu}} v = 0, & x\in\partial\Omega, t>0,\\ u(x,t)=\phi(x,t),\ v(x,t)=\dfrac{1}{\tau}\displaystyle\int_{-\infty}^{t} e^{-(t-s)/\tau}\phi(x,s)\,ds, & x\in\bar{\Omega}, t\le 0. \end{cases} \tag{41}$$

The system (40) can be considered a limit system of (41) as $\varepsilon\to 0$. Previous studies have explored finite-time blow-up phenomena in special cases of (41) (Winkler 2010). While the solution of (41) exists globally, its height depends on ε, so we may expect the solution or its derivative of (40) blows up when $\varepsilon\to 0$ (Wang and Xu 2021; Choi and Kim 2024).

5 Linear stability of constant steady states

In this section, we study the linear stability of constant steady-state solutions. Since we are interested in accumulated memory and its resulting spatial patterns, we focus on integrable kernels $g\in L^1([0,\infty))$ and not consider the case $g=\delta_\tau$, which corresponds to a fixed delay memory $v(x,t)=u(x,t-\tau)$.

We begin with the special case (40) of the model (36), with the weak kernel $g(t)=\tau^{-1}e^{-t/\tau}$ for $\tau>0$ (see Section 6). We then extend the analysis to more general kernels $g\in L^1([0,\infty))$ and derive a sufficient condition for the stability of constant steady states.

We rewrite the equation in (36) as

$$u_t = \nabla\cdot(D(v)\nabla u + A(v)u\nabla v) = \nabla\cdot\left(D(v)\left(\nabla u - \chi(v)\frac{u}{v}\nabla v\right)\right), \tag{42}$$

where $v=g*_t u$ and

$$\chi(v) = -\frac{A(v)v}{D(v)}.$$

The equation (42) has the standard logarithmic chemotaxis form. The function $\chi(v)$ is called the chemosensitivity, and the stability of the spatially homogeneous state is often governed by the threshold $\chi>1$ in the logarithmic chemotaxis setting (see, e.g., (Yoon and Kim 2017; Choi and Kim 2024)); see also Figure 3.

As in the logarithmic chemotaxis framework, we track the quantity

$$\chi(s) := -\frac{A(s)s}{D(s)} \tag{43}$$

as the key parameter for linear stability.

Note that if $(\bar{u},\bar{v})$ is a constant steady state of (40), then necessarily $\bar{u}=\bar{v}$.

Lemma 5.1 *Suppose that $D, A \in C^1([0,\infty))$ and $D(s) > 0$ for all $s \geq 0$. The constant steady state $(\bar{u}, \bar{v})$ of (40) is linearly stable if $\chi(\bar{u}) < \frac{1}{\tau}$ and unstable if $\chi(\bar{u}) > \frac{1}{\tau}$, provided the initial condition is not spatially uniform. If the initial condition is constant, the solution remains constant for all time.*

Proof We linearize (40) around the constant steady state $(\bar{u}, \bar{v})$ and obtain the linearized system:

$$\partial_t \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} D(\bar{v})\Delta & A(\bar{v})\bar{u}\Delta \\ 1 & -\tau^{-1} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} =: M(\bar{u}, \bar{v}) \begin{pmatrix} u \\ v \end{pmatrix}.$$

Let $(\mu, \mathbf{x})$ be an eigenpair of $-\Delta$ under the homogeneous boundary condition $\partial_{\vec{\nu}} u = 0$, with $\mu \geq 0$. Consider the matrix

$$B(\bar{u}, \bar{v}) := \begin{pmatrix} -D(\bar{v})\mu & -A(\bar{v})\bar{u}\mu \\ 1 & -\tau^{-1} \end{pmatrix}.$$

Let $(\lambda, \mathbf{c})$ be an eigenpair of B. Then,

$$M(\bar{u}, \bar{v})(\mathbf{x}e^{\lambda t}\mathbf{c}) = B(\bar{u}, \bar{v})(\mathbf{x}e^{\lambda t}\mathbf{c}) = \lambda \mathbf{x}e^{\lambda t}\mathbf{c} = \partial_t(\mathbf{x}e^{\lambda t}\mathbf{c}).$$

Hence, λ serves as an eigenvalue of the linearized system, with corresponding eigenfunction $\mathbf{x}e^{\lambda t}\mathbf{c}$. Thus, the sign of λ determines the linear stability of (40). We have

$$\mathrm{tr}(B) = -D(\bar{v})\mu - \tau^{-1} < 0, \quad \det(B) = \mu\left(\frac{D(\bar{v})}{\tau} - A(\bar{v})\bar{u}\right).$$

Note that $\mu = 0$ corresponds to the constant mode, which always yields $\lambda_1 = -\tau^{-1} < 0$ and $\lambda_2 = 0$. For $\mu > 0$, linear stability (instability) of $(\bar{u}, \bar{v})$ occurs if and only if $\det(B) > 0$ ($\det(B) < 0$), which in turn holds if and only if $\chi(\bar{u}) < \frac{1}{\tau}$ ($\chi(\bar{u}) > \frac{1}{\tau}$), since $\bar{v} = \bar{u}$. □

Throughout the simulations in Section 6.1, we adopt

$$\gamma(v) = a + \frac{b}{(v+c)^d}, \quad a, b, c, d > 0, \tag{44}$$

which is a strictly decreasing and positive function. Its derivative is given by

$$\gamma'(v) = \frac{-bd}{(v+c)^{d+1}} < 0.$$

The corresponding χ for each movement model is

1. (Advection–Diffusion)

$$\chi(v) = -\frac{\gamma'(v)v}{D} = \frac{bdv}{D(v+c)^{d+1}},$$

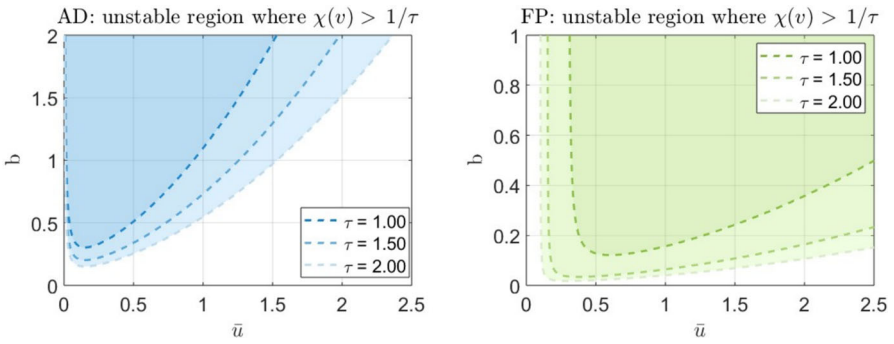

Fig. 3 Stability region in parameter space for the Advection–Diffusion (AD) and Fokker–Planck (FP) models. Shaded regions indicate instability of the constant steady state in the $(\bar{u}, b)$ plane for fixed γ with parameters $a = 0.05$, $c = 0.3$, $d = 2$, and $D = 1$

2. (Fickian type) $\chi(v) = 0$, and
3. (Fokker–Planck type)

$$\chi(v) = -\frac{\gamma'(v)v}{\gamma(v)} = \frac{bdv}{(v+c)(a(v+c)^d + b)}.$$

Because spatially homogeneous steady states are always linearly stable in the Fickian case, we focus on the remaining two models. We identify the instability regions as b and $\bar{u}$ vary, while keeping all other parameters fixed. Here, b controls the overall magnitude of γ, and $\bar{u}$ (the domain average of the initial condition) influences linear stability. See Figure 3. The shaded regions represent instability of the constant steady state, with the extent of the instability depending on the choice of τ. Outside the shaded areas, the constant equilibrium is stable. When b is very small, the constant equilibrium is always stable, as the influence of memory on movement is too weak.

The above lemma provides a necessary and sufficient condition for the stability of constant steady states of the general model (36) in the weak-kernel case. To address more general cases (including the strong-kernel case), we introduce the following lemma, which gives a sufficient condition for the stability of constant equilibria of (36). We take the Neumann boundary operator $\mathcal{B}u = \partial_{\vec{\nu}} u$.

Lemma 5.2 *Suppose* $D, A \in C^1([0,\infty))$, $\inf D > 0$, *and* $g \in L^1([0,\infty))$ *with* $g \geq 0$. *Then any nonnegative constant steady state* $\bar{u}$ *of general model* (36) *with* $\mathcal{B}u = \partial_{\vec{\nu}} u$ *is linearly stable if either* $A(\bar{u}) \geq 0$, *or* $A(\bar{u}) < 0$ *and* $\chi(\bar{u}) < \frac{1}{\|g\|_{L^1([0,\infty))}}$.

Proof Linearizing general model (36) around $\bar{u}$, and denoting the perturbation again by u, yields

$$\begin{cases} u_t = D(\bar{u})\Delta u + A(\bar{u})\bar{u}\,\Delta(g *_t u), & x \in \Omega,\ t > 0, \\ \partial_{\vec{\nu}} u = 0, & x \in \partial\Omega,\ t > 0, \\ u(x,t) = \psi(x), & x \in \bar{\Omega},\ t \leq 0. \end{cases} \tag{45}$$

Applying the Laplace transform in time, with $\mathcal{L}u(x,s) = \int_0^\infty e^{-st}u(x,t)\,\mathrm{d}t$, the memory term becomes

$$\begin{aligned}\mathcal{L}(g *_t u(x,\cdot))(s) &= \int_0^\infty e^{-st}\int_0^t g(t-\tau)u(x,\tau)\mathrm{d}\tau\mathrm{d}t + \psi(x)\int_0^\infty e^{-st}\int_{-\infty}^0 g(t-\tau)\mathrm{d}\tau\mathrm{d}t\\ &= \int_0^\infty e^{-st}\int_\tau^\infty g(t-\tau)u(x,\tau)\mathrm{d}t\mathrm{d}\tau + \psi(x)\int_0^\infty e^{-st}\int_t^\infty g(\tau)\mathrm{d}\tau\mathrm{d}t\\ &= \int_0^\infty e^{-s\tau}u(x,\tau)\mathrm{d}\tau\int_0^\infty e^{-st}g(t)\mathrm{d}t + \psi(x)\int_0^\infty g(\tau)\frac{1-e^{-s\tau}}{s}\mathrm{d}\tau\\ &= \mathcal{L}u(s)G(s) + \frac{\psi(x)(G(0)-G(s))}{s},\end{aligned} \tag{46}$$

where $G(s) = \mathcal{L}g(s)$ and $G(0) = \|g\|_{L^1([0,\infty))}$. The Laplace transform of (45) then becomes

$$\begin{cases} -\psi(x) + s\mathcal{L}u = \Big(D(\bar{u}) + A(\bar{u})\bar{u}\,G(s)\Big)\Delta(\mathcal{L}u) + \Delta\psi(x)\dfrac{G(0)-G(s)}{s}, & x\in\Omega\\ \partial_{\vec{\nu}}\mathcal{L}u = 0, & x\in\partial\Omega, \end{cases} \tag{47}$$

Let $\{u_n\}_{n=0}^\infty$ be Laplacian eigenfunctions satisfying

$$-\Delta u_n = \lambda_n u_n, \qquad \partial_{\vec{\nu}}u_n = 0, \qquad 0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \cdots,\ \lambda_n \to \infty.$$

Consider $u(x,t) = \sum_{n\ge0}\xi_n(t)u_n(x)$. Then $\psi(x) = \sum_{n\ge0}\xi_n(0)u_n(x)$. Substituting this expansion into (47) gives, for any $n \ge 1$,

$$\begin{aligned}(\mathcal{L}\xi_n)(s) &= \frac{\xi_n(0)\left(1 - \lambda_n A(\bar{u})\bar{u}\frac{G(0)-G(s)}{s}\right)}{s + \lambda_n\Big(D(\bar{u}) + A(\bar{u})\bar{u}\,G(s)\Big)}\\ &= \frac{\xi_n(0)}{s + \lambda_n\Big(D(\bar{u}) + A(\bar{u})\bar{u}\,G(s)\Big)} - \frac{\xi_n(0)\lambda_n A(\bar{u})\bar{u}(G(0)-G(s))}{s\Big(s + \lambda_n\Big(D(\bar{u}) + A(\bar{u})\bar{u}\,G(s)\Big)\Big)}.\end{aligned} \tag{48}$$

Note that $\xi_n(t)$ is real-valued for all $n \ge 0$ due to the self-adjoint property of $-\Delta$ under Neumann boundary condition. Assume

$$\inf_{s\ge0}\Big(D(\bar{u}) + A(\bar{u})\bar{u}\,G(s)\Big) =: a > 0. \tag{49}$$

Since $G(s) \le G(0)$ for all $s \ge 0$, it follows that for any $s \ge 0$,

$$|(\mathcal{L}\xi_n)(s)| \le \left|\frac{\xi_n(0)}{s + \lambda_n\Big(D(\bar{u}) + A(\bar{u})\bar{u}\,G(s)\Big)}\right| \le \frac{|\xi_n(0)|}{s + a\lambda_n},$$

which upon inversion yields

$$|\xi_n(t)| \le |\xi_n(0)|e^{-a\lambda_n t}.$$

Thus, for all modes with $\lambda_n > 0, \xi_n(t) \to 0$ as $t \to \infty$. For $\lambda_0 = 0$, the corresponding eigenfunctions u_n are constant, so this mode remains unchanged in time. Hence, the constant equilibrium $\bar{u}$ is stable whenever condition (49) holds.

If $A(\bar{u}) \geq 0$, then since $D(\bar{u}) > 0$, (49) always holds, so the constant steady state is stable. If $A(\bar{u}) < 0$, note that $\sup_{s\geq 0} G(s) = \|g\|_{L^1([0,\infty))}$, so condition (49) is equivalent to

$$\chi(\bar{u}) = -\frac{A(\bar{u})\bar{u}}{D(\bar{u})} < \frac{1}{\|g\|_{L^1([0,\infty))}},$$

where χ is defined in (43). This proves the lemma. □

By Lemma 5.2, we can further confirm that, for the Fickian-type diffusion model (4), the constant steady state is always linearly stable, since it corresponds to $D(v) = \gamma(v) > 0$ and $A \equiv 0$ in (36).

6 Numerical simulation

We next perform numerical simulations to further illustrate how accumulated-memory-driven movement strategies give rise to distinct movement patterns.

We begin by specifying the temporal kernel. Since we focus on accumulated memory, we take $g(t)$ to be a function in $L^1([0,\infty))$. The temporal weighting function $g(t)$ characterizes the dependence of memory on past information. Here, we consider the Gamma distribution. Specifically, the Gamma distribution function of order k (where $k \in \mathbb{N} \cup \{0\}$) is defined as

$$g_k(t) = \frac{t^k e^{-t/\tau}}{\tau^{k+1}\Gamma(k+1)}. \tag{50}$$

Two special cases of the Gamma kernel, commonly analyzed, are referred to as the weak kernel and the strong kernel (Shi et al. 2021b; Wang and Salmaniw 2023). When $k = 0$, the kernel is called the weak kernel. Function $g_0(t) = \tau^{-1}e^{-t/\tau}$ is strictly decreasing with t, reflecting the fading memory over time, where older information becomes less important. In contrast, when $k > 0$, the kernel is called the strong kernel. This kernel increases over $[0, k\tau)$ and then decreases over $(k\tau, \infty)$, representing information acquisition followed by memory decay, as illustrated in Figure 4. The parameters k and τ jointly determine the average delay and the weighting of past information, which are critical for quantifying the impact of memory on movement dynamics. The mean and variance of $g_k(t)$ are given by

$$\mathbb{E}(g_k(t)) = (k+1)\tau, \quad \text{and} \quad \mathbb{V}\text{ar}(g_k(t)) = (k+1)\tau^2.$$

As the parameters k or τ increase, the temporal kernel $g_k(t)$ emphasizes information from earlier time points, as shown in Figure 4. For the subsequent analysis, we set $\tau = 1$ and examine both the weak kernel ($k = 0$) and the strong kernel ($k > 0$).

The three movement models, given in (1), (4), and (3), with boundary and initial conditions same as in general equation (36), become:

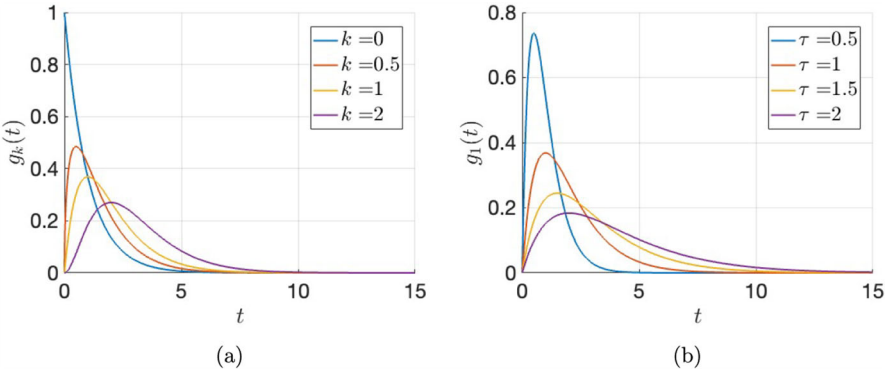

Fig. 4 Temporal kernel functions $g_k(t)$ over the time interval [0, 15]: (a) varying k with fixed $\tau = 1$; (b) varying τ with fixed $k = 1$

Advection–diffusion model:

$$\begin{cases} u_t = \nabla \cdot (D\,\nabla u + u\nabla\gamma(g *_t u)), & x \in \Omega,\ t > 0, \\ \partial_{\vec{\nu}} u = 0, & x \in \partial\Omega,\ t > 0, \\ u(x,t) = \phi(x,t), & x \in \bar{\Omega},\ t \le 0. \end{cases} \tag{51}$$

Fickian type diffusion model:

$$\begin{cases} u_t = \nabla \cdot (\gamma(g *_t u)\nabla u), & x \in \Omega,\ t > 0, \\ \partial_{\vec{\nu}} u = 0, & x \in \partial\Omega,\ t > 0, \\ u(x,t) = \phi(x,t), & x \in \bar{\Omega},\ t \le 0. \end{cases} \tag{52}$$

Fokker–Planck type diffusion model:

$$\begin{cases} u_t = \Delta\,(\gamma(g *_t u)u), & x \in \Omega,\ t > 0, \\ \partial_{\vec{\nu}} u = 0, & x \in \partial\Omega,\ t > 0, \\ u(x,t) = \phi(x,t), & x \in \bar{\Omega},\ t \le 0. \end{cases} \tag{53}$$

Here, the temporal kernel g is given in (50). γ is required to be positive in (52) and (53).

Areas with higher accumulated memory indicate locations that are not only currently crowded but also have been heavily populated in the past. Some animals may prefer such areas to enhance group hunting, maintain social cohesion, or increase mating opportunities. In this case, animals either reduce their movement rate within high-memory regions or actively move toward them, a phenomenon referred to as memory-suppressed movement. Conversely, some animals may avoid areas of high population pressure to reduce intraspecific or interspecific competition. In this case, animals in high-memory regions increase their movement rate in order to leave more

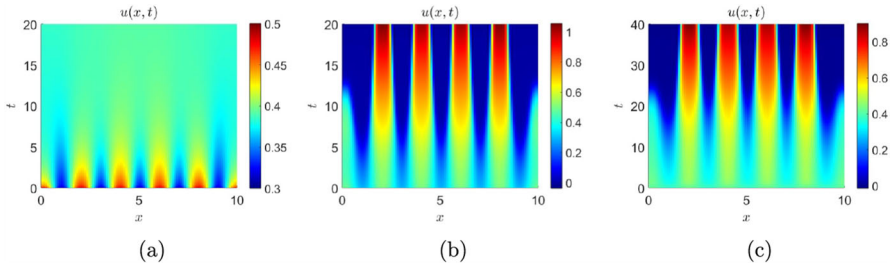

Fig. 5 Spatiotemporal distribution of population density u for the advection–diffusion models: (a) $(\bar{u}, b) = (0.4, 0.3)$ with the weak kernel; (b) $(\bar{u}, b) = (0.4, 0.5)$ with the weak kernel; (c) $(\bar{u}, b) = (0.4, 0.5)$ with the strong kernel ($k = 1$). In cases (b) and (c), the peaks of the initial function grow over time and an aggregation movement pattern is observed, whereas in case (a) the total population becomes uniform. For the strong kernel case (c), it undergoes a delay to form aggregates. Initial condition (54)

quickly, which is termed memory-enhanced movement. Both memory-suppressed and memory-enhanced movement cases are considered in our simulations.

(1) **Memory-suppressed movement:** In the three movement models (51)–(53), this corresponds to the case where γ is a decreasing function. Compared to the general model (36), this implies $D(v) = D, \gamma(v), \gamma(v) > 0$ for all three models (51)–(53), while $A(v) = \gamma'(v) < 0$ for (51) and (53), and $A(v) = 0$ for (52).
(2) **Memory-enhanced movement:** In the three movement models (51)–(53), this corresponds to the case where γ is an increasing function. Compared to the general model (36), this implies $D(v) = D, \gamma(v), \gamma(v) > 0$ for all three models (51)–(53), while $A(v) = \gamma'(v) > 0$ for (51) and (53), and $A(v) = 0$ for (52).

We use the finite difference scheme to simulate the equations. We take time-independent initial data. To simulate the memory map $v = g *_t u$ for each time step, we use the zero-order hold discretization. Details are in Appendix 8.

6.1 The case of memory-suppressed movement

As we can see from Lemma 5.1, it is possible that $\chi(\bar{u}) > \frac{1}{\tau}$ depending on the choice of parameters, in which case we expect a non-constant steady state. Since memory-suppressed movement implies that animals prefer regions of higher population pressure, either by actively moving toward high-density areas or by reducing their movement rate within them, we expect the population to form aggregates. In this case, the kernel function g influences only the speed of aggregate formation.

For the weak memory kernel $g(t) = e^{-t}$, which is the case of $k = 0, \tau = 1$, we can apply Lemma 5.1 to guide the choice of parameters for each equation (See Figure 3). Specifically, we take a positive decreasing function

$$\gamma(v) = 0.05 + \frac{b}{(v + 0.3)^2},$$

which corresponds to $a = 0.05$, $c = 0.3$, and $d = 2$ in (44).

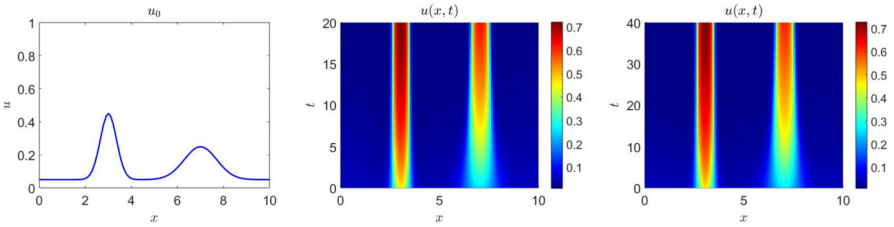

Fig. 6 Simulation results with an asymmetric initial population u_0 on the left. The figure in the middle is the case with the weak kernel. The one on the right is the case with the strong kernel for $k = 1$. Formation of aggregates with the strong kernel occurs more slowly than that of the weak kernel

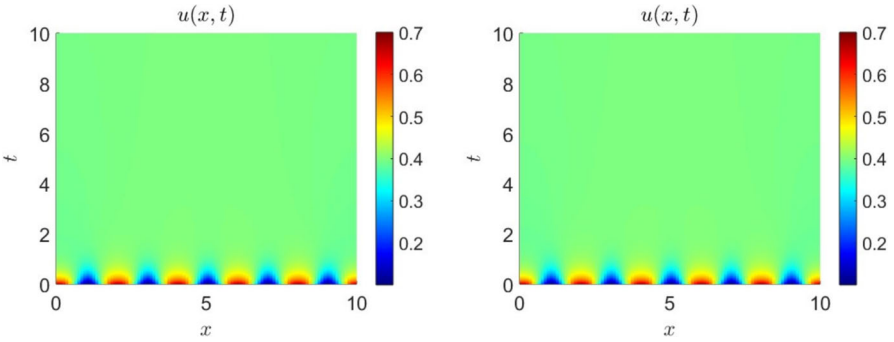

Fig. 7 Spatiotemporal solution of Fickian type diffusion model. The left figure is for the weak kernel $k = 0$, while the right one is for the strong kernel $k = 1$. We took the same initial condition and γ as the advection-diffusion model. It shows uniformization in the end

Advection–Diffusion model: We fix $D = 1$. For the parameters, we chose pairs $(\bar{u}, b) = (0.4, 0.3)$ for stable state, $(\bar{u}, b) = (0.4, 0.5)$ for unstable state. Let an initial function be given by

$$\phi(x,t) = 0.1\cos(\pi x) + 0.4, \tag{54}$$

which has an average value $\bar{u} = \frac{1}{|\Omega|}\int_\Omega u(x,0)\mathrm{d}x = 0.4$. The results for the advection–diffusion model with the initial condition (54) are shown in Figure 5. When $(\bar{u}, b) = (0.4, 0.3)$ (see Figure 5a), the constant equilibrium is stable, and the animals reach a uniform distribution across space. In contrast, when $(\bar{u}, b) = (0.4, 0.5)$ (Figures 5b, 5c), the constant equilibrium becomes unstable and a nonhomogeneous pattern emerges. For both the strong and weak kernels, the animals exhibit aggregation behavior, but in the strong kernel case there is a significant delay in the formation of aggregates compared to the weak kernel. When the initial condition is asymmetric, the nonhomogeneous pattern still emerges but displays noticeable asymmetry, and the delay in aggregation under the strong kernel still remains evident (see Figure 6).

Fickian type diffusion model: We use the same parameters and γ as in the advection–diffusion model. Even when γ is chosen to be a decreasing function, the population distribution always tends to become uniform quickly for both strong and weak kernels (Figure 7). No collective behavior is observed, which is consistent with our analysis in Lemma 5.2.

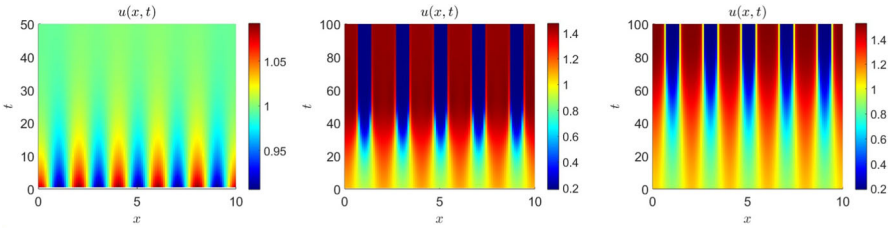

Fig. 8 patiotemporal distribution of population density for the Fokker–Planck type diffusion model: (a) $(\bar{u}, b) = (1, 0.1)$ with the weak kernel; (b) $(\bar{u}, b) = (1, 0.2)$ with the weak kernel; (c) $(\bar{u}, b) = (1, 0.2)$ with the strong kernel ($k = 1$). In case (a), the population remains uniformly distributed. In cases (b) and (c), aggregation emerges, with the strong kernel producing a delayed onset of pattern formation

Fokker-Planck type diffusion model: According to Figure 3, we choose pairs of the parameters $(\bar{u}, b) = (1, 0.1)$ for stable constant state, $(\bar{u}, b) = (1, 0.2)$ for unstable one. Let an initial function be given by

$$\phi(x, t) = 0.1 \cos(\pi x) + 1,$$

which has an average value $\bar{u} = 1$.

As illustrated in Figure 8, aggregation behavior is also observed in the Fokker—Planck type diffusion model when the parameters lie in the instability regime. Specifically, when $(\bar{u}, b) = (1, 0.1)$ (Figure 8a), the constant equilibrium is stable, and the animals gradually reach a uniform spatial distribution. In contrast, when $(\bar{u}, b) = (1, 0.2)$ (Figure 8b, 8c), the constant equilibrium becomes unstable, leading to heterogeneous movement patterns. Aggregation was observed over time, with the resulting peaks showing strong dependence on the initial data. This implies that reliance on accumulated memory makes movement more sensitive to the initial or past population states.

6.2 The case of memory-enhanced movement

Throughout, we take a bounded increasing function

$$\gamma(v) = a + b\left(\frac{v}{v + c}\right)^d, \qquad a, b, c, d > 0,$$

and fix $D = 1$. Then $\gamma'(v) > 0$ for $v > 0$.

We first consider the weak memory kernel $g(t) = e^{-t}$ ($\tau = 1$). By Lemma 5.1, any constant steady state $\bar{u}$ is linearly stable since

$$\chi(\bar{u}) = -\frac{A(\bar{u})\,\bar{u}}{D(\bar{u})} \in \left\{-\frac{\bar{u}\gamma'(\bar{u})}{D},\ 0,\ -\frac{\bar{u}\gamma'(\bar{u})}{\gamma(\bar{u})}\right\} \leq 0 < 1.$$

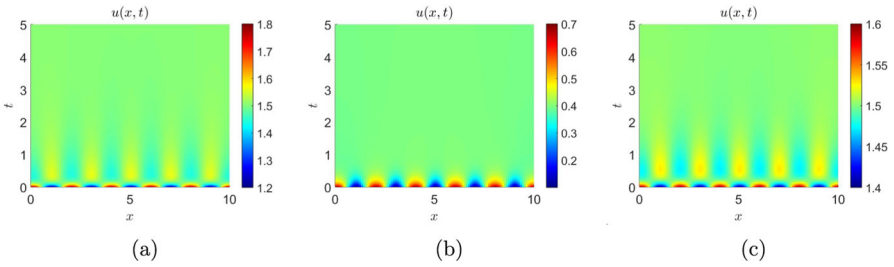

Fig. 9 Spatiotemporal distribution of population density with a weak memory kernel ($k = 0$) for memory-enhanced movement: (a) advection–diffusion model; (b) Fickian type diffusion model; (c) Fokker–Planck type diffusion model. We took $(a, b, c, d) = (0.01, 0.98, 0.4, 3)$ for each model

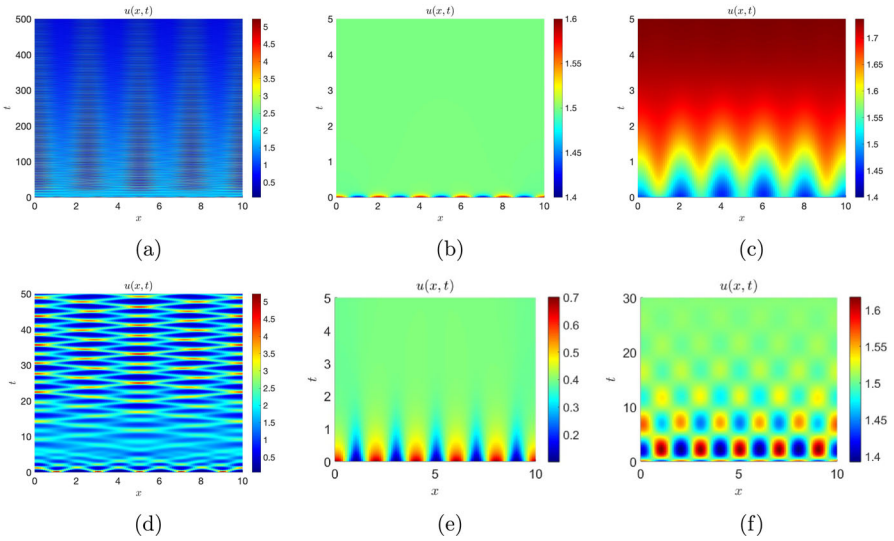

Fig. 10 Spatiotemporal distribution of population density with a strong memory kernel ($k > 0$) for memory-enhanced movement: (a,d) advection–diffusion model showing periodic nonconstant solutions at short times (d, $t \leq 50$) and convergence to uniform steady state at long times (a, $t \leq 500$); (b,e) Fickian type diffusion for $k = 1$ and $k = 3$; (c,f) Fokker–Planck type diffusion for $k = 1$ and $k = 3$. Parameters: $(a, b, c, d) = (0.01, 3.5, 0.3, 3)$ for (a)-(d) and $(a, b, c, d) = (0.01, 0.98, 1.5, 3)$ for (e)-(f)

One can also confirm from Lemma 5.2 that, since $A(v) \in \{\gamma'(v), 0\} \geq 0$ in all three models, the constant steady state is always linearly stable for both the strong-kernel and weak-kernel cases.

We use the initial condition

$$\phi(x, t) = 0.3\cos(\pi x) + 1.5,$$

with spatial mean $\bar{u} = 1.5$,

Simulation results show that in both the weak- and strong-kernel cases, the population eventually reaches a uniform distribution (Figures 9 and 10), consistent with Lemmas 5.1 and 5.2. For the weak-kernel case, all three diffusion types rapidly con-

verge to the uniform steady state, with the advection–diffusion and Fokker–Planck models exhibiting a slightly more pronounced delay compared to the Fickian diffusion model (Figure 9).

For the strong-kernel case ($k > 0$), although all models ultimately reach a uniform state, more interesting dynamics are observed. The numerical solution of the advection–diffusion equation exhibits strong nonconstant periodic patterns (Figures 10a, 10d), which persist for a long time compared to the other diffusion types, which converge eventually to uniformity. The Fickian type diffusion model consistently reaches a uniform state rapidly (Figures 10b, 10e); however, when the memory kernel has a longer temporal delay $k = 3$, more noticeable delays in movement patterns may appear (Figure 10e). The Fokker–Planck type diffusion model produces a strong delay (Figure 10c) or a "wiggling" pattern (Figure 10f) before ultimately reaching a uniform distribution.

7 Discussion

Memory plays a crucial role in determining where and how animals move, influencing their movement strategies over time. Movement models are a central tool in population ecology for studying these dynamics. However, to date, only advection–diffusion models have explicitly incorporated the effects of memory on movement. In this study, we extend this framework to consider three commonly used movement models: advection–diffusion, Fickian type diffusion, and Fokker–Planck type diffusion. We investigate how individual-level movement strategies based on accumulated memory give rise to different population-level movement patterns and spatial distributions.

Each model represents a distinct movement strategy: gradient-based movement, environment matching, and location-based movement, respectively. These models are systematically derived using two classical approaches: time–space discretization and discrete velocity-jump processes. The derivations reveal the connections between ecological movement strategies and their mathematical formulations, showing how individual-level movement rules give rise to population-level behavior. In the advection–diffusion model, animals exhibit both unbiased random movement and directional movement guided by the gradient of accumulated memory. In the Fickian type diffusion model, animals adjust their movement rates to symmetrically match those of individuals in adjacent patches based on local memory. This allows them to avoid overcrowding and, over time, the system tends to approach a uniform steady state. In contrast, the Fokker–Planck type diffusion assumes no directional preference for individuals; however, the probability of leaving a given location is influenced by memory. Spatial variations in the memory field then induce an implicit directional bias at the population level. As a result, both the advection–diffusion and Fokker–Planck models may generate spatial patterns under heterogeneous environmental conditions. These theoretical predictions are further supported by numerical simulations.

All three of these memory-induced models can be generalized into (5). Although memory has been shown to be important in animal movement, the only existing result concerns the advection–diffusion model with constant diffusion and taxis coefficients and a linear fixed-delay memory. While some studies have analyzed advection–diffu-

sion models with accumulated memory, they focus on bifurcation over equilibrium and do not prove existence results. Moreover, other types of movement models have not yet incorporated memory. To bridge this gap and validate the well-posedness of these models, we have proved the local existence of solutions for the general model (36), which allows both the diffusion coefficient and taxis function to depend on memory. The memory dependence may be nonlinear, and the temporal kernel can incorporate both a fixed-time delay and accumulated memory. The proof relies on fixed-point theory. Due to its generality, this result can be applied beyond the three models discussed above. We also explore the possibility of finite-time blow-up by providing a sufficient and necessary condition for the stability of constant steady states under a weak memory kernel, and a sufficient condition for any kernel in $L^1([0,\infty))$.

Animals may prefer or avoid crowded regions for various reasons (Ford and Swearer 2013; Matthysen 2005; Cressman and Garay 2011), leading to memory-suppressed or memory-promoted movement. We studied both cases across these three movement models. In the memory-suppressed case, the advection–diffusion and Fokker–Planck models can either exhibit aggregation or converge to a uniform state, depending on the parameters. When heterogeneous movement patterns emerge, a strong kernel produces a noticeable delay in aggregation compared to a weak kernel. These patterns are highly sensitive to the initial conditions, which aligns with the mechanism that animals rely on accumulated memory and adjust movement locally. In contrast, the Fickian diffusion model consistently evolves toward a uniform distribution. For memory-promoted movement, populations in all three models eventually become uniform. However, with a strong kernel, the advection–diffusion and Fokker–Planck models may display pronounced delays, such as periodic oscillatory movement or "wiggling" patterns, before reaching uniformity. By comparison, the Fickian diffusion model always reaches a uniform state rapidly.

This framework can be readily extended to more complex scenarios. For example, incorporating population dynamics into (5) yields:

$$u_t = \nabla \cdot (D(g *_t u)\nabla u + A(g *_t u)u\nabla(g *_t u)) + f(u),$$

where $f(u)$ represents a reaction term. For example, in the case of intraspecific competition, it can take the form $f(u) = u(1-u)$. Additionally, if movement is influenced by other spatially and temporally heterogeneous source $m(x,t)$, such as food availability, territorial markings, predation risk, and other environmental cues, instead of population density, then equation (5) becomes:

$$u_t = \nabla \cdot (D(g *_t m)\nabla u + A(g *_t m)u\nabla(g *_t m)) + f(u,m),$$

where $m(x,t)$ varies across both space and time. Here, we only studied local accumulated memory. One may also consider nonlocal accumulated memory,

$$u_t = \nabla \cdot (D(g *_t h *_x u)\,\nabla u + A(g *_t h *_x u)\,u\,\nabla(g *_t h *_x u))\,,$$

where g and h represent the temporal and spatial kernels, respectively. Such extensions provide a pathway for future work aimed at capturing more complex ecological phenomena.

8 Zero-order hold time discretization for memory in finite difference method

From exponential to Gamma/Erlang kernels. Let the causal exponential kernel be

$$g_0(t) = \frac{1}{\tau} e^{-t/\tau} \mathbf{1}_{\{t \geq 0\}}.$$

For $k \in \mathbb{N}$, define the $(k+1)$-fold self-convolution

$$g_k := \underbrace{g_0 * g_0 * \cdots * g_0}_{k+1 \text{ times}}.$$

Then g_k is the Erlang (Gamma) density with shape $k+1$ and scale τ:

$$g_k(t) = \frac{t^k e^{-t/\tau}}{\tau^{k+1}\Gamma(k+1)} \mathbf{1}_{\{t \geq 0\}}.$$

This follows from a direct time-domain calculation (e.g., for $k = 1$, $(g_0 * g_0)(t) = \frac{1}{\tau^2} e^{-t/\tau} \int_0^t ds = \frac{t}{\tau^2} e^{-t/\tau}$), or by Laplace transforms: $\mathcal{L}\{g_0\}(s) = (1 + \tau s)^{-1}$ so $\mathcal{L}\{g_k\}(s) = (1 + \tau s)^{-(k+1)}$.

Convolution over $(-\infty, t)$. For any locally integrable u and causal g_k,

$$(g_k * u)(t) = \int_{-\infty}^{t} g_k(t-s)u(s)\, ds = \big(\underbrace{g_0 * (g_0 * (\cdots * (g_0 * u) \cdots))}_{k+1}\big)(t),$$

by associativity and Tonelli/Fubini. Thus convolving once with g_k over $(-\infty, t)$ is equivalent to passing u through a cascade of $(k+1)$ identical exponential stages.

ODE cascade (continuous time). Write $z^{(1)} = g_0 * u, z^{(2)} = g_0 * z^{(1)}, \ldots, h = z^{(k+1)}$. Each stage satisfies the first-order linear ODE

$$\dot{z}^{(1)} = -\tfrac{1}{\tau} z^{(1)} + \tfrac{1}{\tau} u, \qquad \dot{z}^{(m)} = -\tfrac{1}{\tau} z^{(m)} + \tfrac{1}{\tau} z^{(m-1)} \ \ (m \geq 2),$$

and the memory variable is $h = z^{(k+1)} = g_k * u$.

Exact Zero-Order Hold time discretization and the weight $1 - \alpha$. On each step $[t_n, t_{n+1})$ we assume zero-order hold (ZOH) on the input, i.e. $u(t) \equiv u^n$. Solve the linear ODE exactly:

$$h(t_{n+1}) = e^{-dt/\tau} h(t_n) + \frac{1}{\tau} \int_{t_n}^{t_{n+1}} e^{-(t_{n+1}-s)/\tau} u^n ds$$

$$= e^{-dt/\tau} h(t_n) + (1 - e^{-dt/\tau}) u^n,$$

where $dt = t_{n+1} - t_n$. Solving one stage exactly over the step gives

$$z^{n+1} \;=\; e^{-\Delta t/\tau} z^n \;+\; \left(1 - e^{-\Delta t/\tau}\right) u^n.$$

Defining $\alpha = e^{-\Delta t/\tau}$ and $\eta = 1 - \alpha$, the cascade updates are

$$z^{(1),n+1} = \alpha\, z^{(1),n} + \eta\, u^n, \qquad z^{(m),n+1} = \alpha\, z^{(m),n} + \eta\, z^{(m-1),n} \;\; (m \geq 2),$$

with $h^{n+1} = z^{(k+1),n+1}$. Here $\eta = 1 - \alpha = \int_0^{\Delta t} \frac{1}{\tau} e^{-r/\tau} dr$ is the exact weight of the held input; comparing to explicit Euler, $\eta \approx \Delta t/\tau$ when $\Delta t/\tau \ll 1$.

Computational benefits. The cascade requires $(k+1)$ auxiliary states per grid point and costs $\mathcal{O}(k+1)$ per time step (per point), avoiding the growing history sum in a naive quadrature of $\int_0^t g_k(t-s)u(s)\, ds$.

Acknowledgements We thank the reviewers for very detailed and helpful suggestions, which have significantly improved the quality of this work. The research of Hao Wang was partially supported by the Natural Sciences and Engineering Research Council of Canada (Individual Discovery Grant RGPIN-2025-05734) and the Canada Research Chairs program (Tier 1 Canada Research Chair Award).

Data Availability No data was used in this study.

Declarations

Conflict of Interest Statement The authors declare no conflicts of interest.

References

Ahn J, Yoon C (2019) Global well-posedness and stability of constant equilibria in parabolic-elliptic chemotaxis systems without gradient sensing. Nonlinearity 32(4):1327

Amélineau F, Péron C, Lescroël A, Authier M, Provost P, Grémillet D (2014) Windscape and tortuosity shape the flight costs of northern gannets. J Exp Biol 217(6):876–885

An Q, Wang C, Wang H (2020) Analysis of a spatial memory model with nonlocal maturation delay and hostile boundary condition. Discrete & Continuous Dyn Syst Ser A 40(10)

Bringuier E (2011) Particle diffusion in an inhomogeneous medium. Eur J Phys 32(4):975

Carleman T (1957) Problèmes mathématiques dans la théorie cinétique des gaz

Carrillo JA, Craig K, Yao Y (2019) Aggregation-diffusion equations: dynamics, asymptotics, and singular limits. In: Active Particles, Volume 2: Advances in Theory, Models, and Applications, Springer, pp 65–108

Cho E, Kim Y-J (2013) Starvation driven diffusion as a survival strategy of biological organisms. Bull Math Biol 75:845–870

Choi B, Kim Y-J (2019) Diffusion of biological organisms: fickian and fokker-planck type diffusions. SIAM J Appl Math 79(4):1501–1527

Choi K, Kim Y-J (2024) Chemotactic cell aggregation viewed as instability and phase separation. Nonlinear Anal Real World Appl 80:104147

Couzin ID, Krause J, Franks NR, Levin SA (2005) Effective leadership and decision-making in animal groups on the move. Nat 433(7025):513–516

Cressman R, Garay J (2011) The effects of opportunistic and intentional predators on the herding behavior of prey. Ecol 92(2):432–440

Dahlquist F, Lovely P, Koshland D (1972) Quantitative analysis of bacterial migration in chemotaxis. Nat New Biol 236(65):120–123
Desvillettes L, Kim Y-J, Trescases A, Yoon C (2019) A logarithmic chemotaxis model featuring global existence and aggregation. Nonlinear Anal Real World Appl 50:562–582
Ducrot A, Fu X, Magal P (2018) Turing and turing-hopf bifurcations for a reaction diffusion equation with nonlocal advection. J of Nonlinear Sci 28(5):1959–1997
Eftimie R, de Vries G, Lewis MA (2007) Complex spatial group patterns result from different animal communication mechanisms. Proc Natl Acad Sci 104(17):6974–6979
Erban R, Othmer HG (2004) From individual to collective behavior in bacterial chemotaxis. SIAM J Appl Math 65(2):361–391
Fick A (1855) Ueber diffusion. Ann Phys 170(1):59–86
Ford JR, Swearer SE (2013) Two's company, three's a crowd: food and shelter limitation outweigh the benefits of group living in a shoaling fish. Ecol 94(5):1069–1077
Fujie K, Jiang J (2020) Global existence for a kinetic model of pattern formation with density-suppressed motilities. J Differ Equ 269(6):5338–5378
Funaki T, Izuhara H, Mimura M, Urabe C (2012) A link between microscopic and macroscopic models of self-organized aggregation. Netw and Heterogeneous Media 7(4):705–740
Giunta V, Hillen T, Lewis M, Potts JR (2022) Local and global existence for nonlocal multispecies advection-diffusion models. SIAM J Appl Dyn Syst 21(3):1686–1708
Giunta V, Hillen T, Lewis MA, Potts JR (2022) Detecting minimum energy states and multi-stability in nonlocal advection-diffusion models for interacting species. J Math Biol 85(5):56
Goldstein S (1951) On diffusion by discontinuous movements, and on the telegraph equation. The Quarterly J of Mech and Appl Math 4(2):129–156
Hillen T, Othmer HG (2000) The diffusion limit of transport equations derived from velocity-jump processes. SIAM J Appl Math 61(3):751–775
Hillen T, Painter KJ (2009) A user's guide to pde models for chemotaxis. J Math Biol 58(1):183–217
Jackson DE, Ratnieks FL (2006) Communication in ants. Current Biol 16(15):R570–R574
Ji Q, Wu R, Zhang T (2024) Stability and bifurcation of a heterogeneous memory-based diffusive model. Discrete & Cont Dyn Syst: Ser S
Jiang J, Laurençot P, Zhang Y (2022) Global existence, uniform boundedness, and stabilization in a chemotaxis system with density-suppressed motility and nutrient consumption. Comm Partial Differ Equ 47(5):1024–1069
Jin H-Y, Wang Z-A (2021) Global dynamics and spatio-temporal patterns of predator-prey systems with density-dependent motion. Eur J Appl Math 32(4):652–682
Kac M (1974) A stochastic model related to the telegrapher's equation. The Rocky Mountain J of Math 4(3):497–509
Keller EF, Segel LA (1971) Model for chemotaxis. J Theor Biol 30(2):225–234
Kim H-Y, Lee KK, Firisa G, Lee J, Choi MC, Kim Y-J (2024) Fractionation by spatially heterogeneous diffusion: experiments and the two-component random walk model. J Am Chem Soc 146:25544–25551
Kwon O, Kim Y-J (2016) Evolution of dispersal with starvation measure and coexistence. Bull Math Biol 78:254–279
Laca EA (2008) Foraging in a heterogeneous environment: intake and diet choice. In: Resource Ecology, Springer, pp 81–100
Li S, Li Z, Dai B (2022) Stability and hopf bifurcation in a prey-predator model with memory-based diffusion. Discrete and Cont Dyn Syst-B 27(11):6885–6906
Li S, Yuan S, Jin Z, Wang H (2023) Bifurcation analysis in a diffusive predator-prey model with spatial memory of prey, allee effect and maturation delay of predator. J Differ Equ 357:32–63
Li Y, Simpson MJ, Wang C (2025) Lattice-based stochastic models motivate non-linear diffusion descriptions of memory-based dispersal. J Math Biol 90(5):52
Lieberman GM (1996) Second order parabolic differential equations. World Scientific
Liechti F (2006) Birds: blowin'by the wind? J Ornithol 147(2):202–211
Lim H-J, Kim Y-J, Kim H-Y (2021) Heterogeneous discrete kinetic model and its diffusion limit. Kinetic & Related Models 14:749–765
Liu C, Fu X, Liu L, Ren X, Chau CK, Li S, Xiang L, Zeng H, Chen G, Tang L-H et al (2011) Sequential establishment of stripe patterns in an expanding cell population. Sci 334(6053):238–241

Liu D, Potts JR, Salmaniw Y, Shi J, Wang H (2025) Biological aggregations from spatial memory and nonlocal advection. Physica D 476:134682
Liu D, Shi J, Jiang W (2025) Local and global bifurcation analysis of density-suppressed motility model. J Math Anal Appl 552:129810
Lunardi A (2012) Analytic semigroups and optimal regularity in parabolic problems. Birkhäuser, Basel
Lyu W, Wang Z-A (2022) Logistic damping effect in chemotaxis models with density-suppressed motility. Adv Nonlinear Anal 12(1):336–355
Ma M, Peng R, Wang Z (2020) Stationary and non-stationary patterns of the density-suppressed motility model. Physica D 402:132259
Matthysen E (2005) Density-dependent dispersal in birds and mammals. Ecography 28(3):403–416
Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, Smouse PE (2008) A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci 105(49):19052–19059
Okubo A, Levin SA (2001) Diffusion and ecological problems: modern perspectives, vol 14. Springer
Othmer HG, Dunbar SR, Alt W (1988) Models of dispersal in biological systems. J Math Biol 26(3):263–298
Othmer HG, Hillen T (2002) The diffusion limit of transport equations II: chemotaxis equations. SIAM J Appl Math 62:1222–1250
Painter KJ, Hillen T, Potts JR (2024) Biological modeling with nonlocal advection-diffusion equations. Math Models Methods Appl Sci 34(01):57–107
Patlak CS (1953) Random walk with persistence and external bias. Bull Math Biophys 15:311–338
Potts JR, Lewis MA (2019) Spatial memory and taxis-driven pattern formation in model ecosystems. Bull Math Biol 81(7):2725–2747
Salmaniw Y, Liu D, Shi J, Wang H (2025) Dynamics of a coupled nonlocal pde-ode system with spatial memory: well-posedness, stability, and bifurcation analysis. arXiv preprint arXiv:2503.11550
Sapir N, Wikelski M, McCue MD, Pinshow B, Nathan R (2010) Flight modes in migrating european bee-eaters: heart rate may indicate low metabolic rate during soaring and gliding. PLoS ONE 5(11):e13956
Shi J, Shi Q (2024) Spatial movement with temporally distributed memory and Dirichlet boundary condition. J Differ Equ 389:305–337
Shi J, Wang C, Wang H (2019) Diffusive spatial movement with memory and maturation delays. Nonlinearity 32(9):3188
Shi J, Wang C, Wang H (2021) Spatial movement with diffusion and memory-based self-diffusion and cross-diffusion. J Differ Equ 305:242–269
Shi J, Wang C, Wang H, Yan X (2020) Diffusive spatial movement with memory. J Dyn Diff Equat 32:979–1002
Shi Q (2025) Existence of the nonhomogeneous steady states in a memory-based diffusive model with nonlocal memory and dirichlet boundary. Appl Math Lett 109685
Shi Q, Shi J, Wang H (2021) Spatial movement with distributed memory. J Math Biol 82(4):33
Smith JT, Elkin JT, Reichert WM (2006) Directed cell migration on fibronectin gradients: effect of gradient slope. Exp Cell Res 312(13):2424–2432
Smith WJ (1980) The behavior of communicating: an ethological approach. Harvard University Press
Smith-Roberge J, Iron D, Kolokolnikov T (2019) Pattern formation in bacterial colonies with density-dependent diffusion. Eur J Appl Math 30(1):196–218
Song Y, Peng Y, Zhang T (2021) The spatially inhomogeneous hopf bifurcation induced by memory delay in a memory-based diffusion system. J Differ Equ 300:597–624
Song Y, Shi J, Wang H (2022) Spatiotemporal dynamics of a diffusive consumer-resource model with explicit spatial memory. Stud Appl Math 148(1):373–395
Song Y, Wang H, Wang J (2024) Cognitive consumer-resource spatiotemporal dynamics with nonlocal perception. J of Nonlinear Sci 34(1):19
Song Y, Wu S, Wang H (2019) Spatiotemporal dynamics in the single population model with memory-based diffusion and nonlocal effect. J Differ Equ 267(11):6316–6351
Song Y, Wu S, Wang H (2021) Memory-based movement with spatiotemporal distributed delays in diffusion and reaction. Appl Math Comput 404:126254
Tao Y, Winkler M (2023) Global solutions to a keller-segel-consumption system involving singularly signal-dependent motilities in domains of arbitrary dimension. J Differ Equ 343:390–418
Taylor GI (1922) Diffusion by continuous movements. Proc Lond Math Soc 2(1):196–212
Terrace HS, Metcalfe J (2005) The missing link in cognition: origins of self-reflective consciousness. Oxford University Press

Tóth Z, Jaloveczki B, Tarján G (2020) Diffusion of social information in non-grouping animals. Front Ecol Evol 8:586058
Wang C, Yuan S, Wang H (2022) Spatiotemporal patterns of a diffusive prey-predator model with spatial memory and pregnancy period in an intimidatory environment. J Math Biol 84:12
Wang H, Salmaniw Y (2023) Open problems in PDE models for knowledge-based animal movement via nonlocal perception and cognitive mapping. J Math Biol 86(5):71
Wang H, Wang K, Kim Y-J (2022) Spatial segregation in reaction-diffusion epidemic models. SIAM J Appl Math 82(5):1680–1709
Wang Y, Song Y, Wang H (2025) Bifurcations induced by nonlocal spatial memory versus nonlocal perception: Y. wang et al. J Math Biol 91(1):3
Wang Y, Wang C, Fan D, Chen Y (2023) Dynamics of a predator-prey model with memory-based diffusion. J Dyn Diff Equat 1–24
Wang Z-A, Xu X (2021) Steady states and pattern formation of the density-suppressed motility model. IMA J Appl Math 86(3):577–603
Winkler M (2010) Absense of collapse in a parabolic chemotaxis system with signal-dependent sensitivity. Math Nachr 283:1664–1673
Wu J, Mao Z, Tan H, Han L, Ren T, Gao C (2012) Gradient biomaterials and their influences on cell migration. Interface focus 2(3):337–355
Wüst M, Menzel F (2017) I smell where you walked-how chemical cues influence movement decisions in ants. Oikos 126(1):149–160
Xue S, Song Y, Wang H (2024) Spatio-temporal dynamics in a reaction-diffusion equation with nonlocal spatial memory. SIAM J Appl Dyn Syst 23(1):641–667
Yoon C, Kim Y-J (2017) Global existence and aggregation in a keller-segel model with fokker-planck diffusion. Acta Appl Math 149(1):101–123
Zein B, Long JA, Safi K, Kölzsch A, Benitez-Paez F, Wikelski M, Kruckenberg H, Demšar U (2022) Simulating geomagnetic bird navigation using novel high-resolution geomagnetic data. Eco Inform 69:101689
Zentall TR, Clement TS, Bhatt RS, Allen J (2001) Episodic-like memory in pigeons. Psychon Bulletin & Rev 8:685–690
Zhang H, Wang H, Song Y, Wei J (2023) Diffusive spatial movement with memory in an advective environment. Nonlinearity 36(9):4585