Erschienen in: Coping with complexity : model reduction and data analysis / Gorban, Alexander N.; Roose, Dirk (Hrsg.). - Berlin : Springer, 2011. - (Lecture notes in computational science and engineering ; 75). - S. 299-309. - ISBN 978-3-642-14940-5
https://dx.doi.org/10.1007/978-3-642-14941-2_16

Coarse Collective Dynamics of Animal Groups

Thomas A. Frewen, Iain D. Couzin, Allison Kolpas, Jeff Moehlis, Ronald Coifman, and Ioannis G. Kevrekidis

Abstract The coarse-grained, computer-assisted analysis of models of collective dynamics in animal groups involves (a) identifying appropriate observables that best describe the state of these complex systems and (b) characterizing the dynamics of such observables. We devise "equation-free" simulation protocols for the analysis of a prototypical individual-based model of collective group dynamics. Our approach allows the extraction of information at the macroscopic level via parsimonious usage of the detailed, "microscopic" computational model. Identification of meaningful coarse observables ("reduction coordinates") is critical to the success of such an approach, and we use a recently-developed dimensionality-reduction approach (diffusion maps) to detect good observables based on data generated by

local model simulation bursts. This approach can be more generally applicable to the study of coherent behavior in a broad class of collective systems (e.g., collective cell migration).

1 Introduction

Many animal groups such as fish schools and bird flocks display remarkable collective behavior such as coherent group motion and transitions between different group configurations [1–4]. A small number of *informed* individuals in such animal groups, namely those with a preferred direction of motion, have been shown to be capable of influencing the group by facilitating the transfer of information, such as the location of a migration route, resources, or predators, to uninformed group members [5]. Understanding the mechanisms of information transfer in these and other biological systems is a problem of fundamental interest.

We consider a 1-dimensional model of animal group motion based on the 2-dimensional individual-based model in [6]. The direction of travel for each individual in the group is computed based on the occupancy of surrounding "zones of information." A single informed individual in the group has a preferred direction of travel. Model simulations indicate "stick-slip"–type dynamic behavior: the group appears to be "stuck in place" some of the time, with no net motion of its centroid; at other times the group clearly travels in the preferred direction of the informed individual.

We analyze the dynamics of this model problem through the computation of an effective free energy surface, obtained in terms of a suitable *reaction coordinate* that characterizes the state of the system. Such a reaction coordinate is first proposed by trial and error, after extended computational exploration of the dynamics. We then use *diffusion maps* [7] to extract the appropriate observable (the reaction coordinate) in an automated fashion; this involves the computation of the leading eigenvectors of the weighted Laplacian on a graph constructed from direct simulation data. Such data-mining procedures can directly link with, and facilitate, coarse-graining algorithms; in particular, there is a natural connection between these manifold-learning techniques and the equation-free framework [8] for complex/multiscale system modeling. The coarse variables identified through diffusion maps can be used to design, initialize, and process the results of short bursts of detailed simulation, whose purpose is to extract coarse-grained, macroscopic information from the fine scale, microscopic solver. We will also briefly discuss procedures for translating between physical system variables and these data-based coarse variables (observables).

2 Model Description

We consider a 1-dimensional model of collective group motion where each individual i in a group of size N is characterized by its position $c_i(t)$, speed $s_i(t) > 0$, and direction of travel $v_i(t)$. These quantities are updated for each individual at

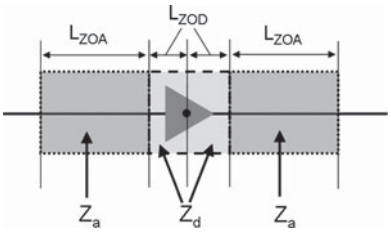

Fig. 1 Zones of information surrounding an individual in our spatially one-dimensional model. Position along the spatial axis is indicated by a *black dot*; the large *triangle* indicates direction of travel. The *zone of deflection* Z_d extends a distance L_{zod} to the *left* and *right* of the individual. The zone of attraction Z_a extends a distance L_{zoa} beyond the edges of Z_d. In our simulations we use $L_{zod} = 0.1$, $L_{zoa} = 4$, $N = 10$ individuals, and an agent response time $\Delta t = 0.1$

each time step (the time step size Δt may be interpreted as the response time of an individual). The individual update rules are based on the distribution of neighbors in spatial "zones" of information surrounding each individual as indicated in Fig. 1 (model parameter values are indicated in caption).

In an inner *zone of deflection* Z_d, collisions with near neighbors are avoided. When this zone is occupied for individual i its desired direction of travel at the next time step $d_i(t + \Delta t)$ may be computed using

$$d_i(t + \Delta t) = -\frac{\sum_{\substack{j \neq i \\ j \in Z_d}} \frac{c_j(t) - c_i(t)}{|c_j(t) - c_i(t)|}}{\left|\sum_{\substack{j \neq i \\ j \in Z_d}} \frac{c_j(t) - c_i(t)}{|c_j(t) - c_i(t)|}\right|}. \tag{1}$$

This equation reduces to

$$d_i(t + \Delta t) = \frac{N_{L,Z_d}(i) - N_{R,Z_d}(i)}{|N_{L,Z_d}(i) - N_{R,Z_d}(i)|} \tag{2}$$

where $N_{L,Z_d}(i)$ ($N_{R,Z_d}(i)$) is the number of individuals to the left (right) of individual i within its zone of deflection. Individuals with an occupied zone of deflection prioritize avoidance by ignoring the positions and directions of travel of others outside of Z_d. Individual i only interacts with those in an outer *zone of attraction* Z_a if its zone of deflection is empty. Individuals align with and are attracted towards neighbors in this outer zone. The desired direction of travel for individual i (with no neighbors in Z_d) is calculated from

$$d_i^0(t + \Delta t) = \frac{N_{R,Z_a}(i) - N_{L,Z_a}(i)}{|N_{R,Z_a}(i) - N_{L,Z_a}(i)|} + \sum_{j \in Z_a} \frac{v_j(t)}{|v_j(t)|} \tag{3a}$$

$$d_i(t + \Delta t) = \frac{d_i^0(t + \Delta t)}{|d_i^0(t + \Delta t)|} \tag{3b}$$

where $N_{L,Z_a}(i)$ ($N_{R,Z_a}(i)$) is the number of individuals to the left (right) of individual i within its zone of attraction, and the second term on the right hand side in (3a) is a contribution based on the direction of travel of neighbor j ($v_j(t)$) at the current time step; the summation includes individual i. The desired directions of travel for each individual in the group are computed using (2) and (3).

Individuals within the group having a preferred direction of travel (here, to the right) are called *informed*; all other group members are *naive*. We consider a single informed group individual here; the "level" of information of this individual is set by the parameter $p_{inf} = 0.2$ which is the probability with which its desired direction of travel is set to its preferred direction at each time step (irrespective of the dynamics indicated by the zone rules). This parameter determines the balance between interactions with neighbors and a preferred travel direction for the informed individual.

As in [9], information uncertainty is incorporated in the model by flipping the desired travel direction of each individual with a small probability $p_{flip} = 0.001$ using a random number drawn from the uniform distribution to give the actual direction of travel for each individual at the next time step $v_i(t + \Delta t)$. The positions of the individuals are updated using

$$c_i(t + \Delta t) = c_i(t) + s_i(t + \Delta t)v_i(t + \Delta t)\Delta t. \tag{4}$$

The initial positions and directions of travel for each individual are set randomly. The speed of each individual is reset at each time step by drawing from the normal distribution $N(\mu_s = 0.85, \sigma_s^2 = 6.25 \times 10^{-4})$.

3 Stick-Slip Phenomena

The variation of the group centroid with the number of time steps, along with the position of the informed individual, simulated with the model described in the previous section, is shown in Fig. 2 for three different values of p_{inf}. The simulation results clearly visually suggest stick-slip-type behavior, with instances where the group is stuck in place (individuals "vibrate" in place) and also periods where the swarm flows in the preferred direction of the informed individual (here, to the right). In Fig. 2 flat (sloped) regions of the centroid profile correspond to *stick* (*slip*) regimes. At higher values of p_{inf} the group travels further with less stick; we analyze the $p_{inf} = 0.2$ case in greater detail since it shows a balance between instances of stick and slip. The position of the informed individual at each step (black jagged line for $p_{inf} = 0.2$) in Fig. 2 suggests that the group sticks (slips) when the informed individual is distant from (close to) the group centroid. Extensive observation of transient simulations characterized by such behavior suggests,

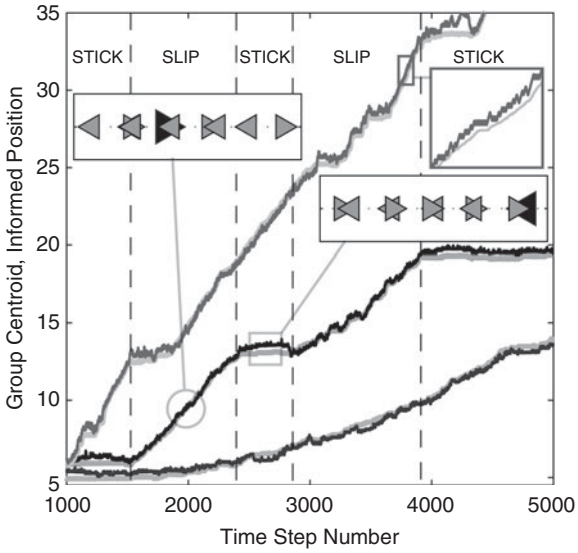

Fig. 2 Variation of group centroid and informed position as a function of simulation time step number for different values of p_{inf}, indicating stick-slip group motion. Jagged *light grey lines* (*top*, *middle*, and *bottom*) indicate centroid position, jagged *dark grey/black lines* (*top*, *middle*, and *bottom*) indicate position of the informed individual (*bottom*: $p_{inf} = 0.1$, *middle*: $p_{inf} = 0.2$, *top*: $p_{inf} = 0.4$) Stick-slip regimes for $p_{inf} = 0.2$ case are indicated in the text at the top of the figure, and separated by *dashed vertical lines*. *Grey open circle* (*square*) shows sample group configuration in slip (*stick*) regime with informed individual *shaded black*; *square* region at *top right* provides a close-up view of the group centroid and informed position for the $p_{inf} = 0.4$ case

as a possible *reaction coordinate* that characterizes the state of the system, $R(t)$ given by

$$R(t) = c_{inf}(t) - \overline{c}(t) \tag{5}$$

where $c_{inf}(t)$ is the position of the informed individual, and $\overline{c}(t) = \frac{1}{N}\sum_{i=1}^{N} c_i(t)$ is the group centroid. The evolution of this human experience-based coordinate (which we will call the "MAN" reaction coordinate) $R(t)$ with the simulation time step number (along with the corresponding centroid evolution) is plotted in Fig. 3. When the value of $R(t)$ is less than about 0.3 (middle panel) the group tends to slip – the informed individual is either at the rear or close to the centroid of the group. For values of $R(t)$ above approximately 0.3 the group is "stuck" – and the informed individual lies at the leading edge of the group.

We selected the "MAN" reaction coordinate $R(t)$ by inspecting the results of many simulations; an alternative, data-based computational procedure that automates the detection of such a coordinate (diffusion maps) will be described in Sect. 5 below.

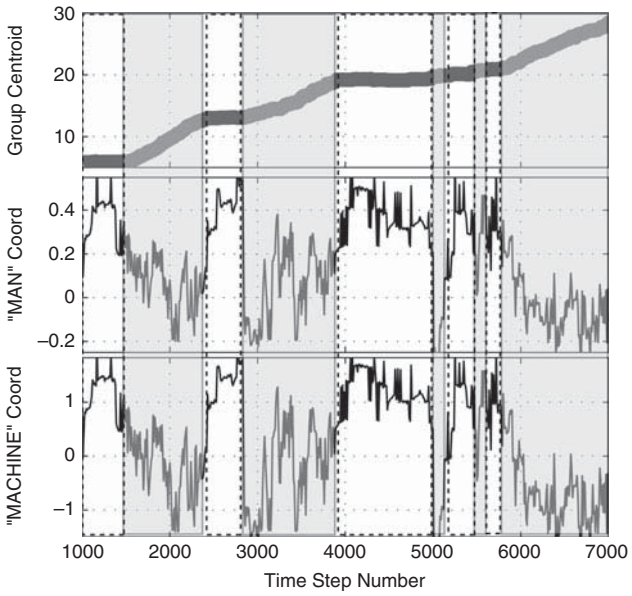

Fig. 3 Evolution of the group centroid ($p_{inf} = 0.2$) as a function of simulation time step number (*top panel*); corresponding evolution of the "MAN" coordinate $R(t)$ defined in (5) (*middle panel*); and corresponding evolution of the "MACHINE" coordinate (*bottom panel*). *Dashed boxes* mark "stick" regimes, *shaded boxes* correspond to "slip" regimes

4 Coarse Variable Analysis

We assume that the evolution of the system can be described in terms of a single, slowly evolving random variable q; the remaining system variables (or their statistics) are assumed to quickly become slaved to q. The evolution of the system can then be approximately described in terms of a time-dependent probability density function $f(q,t)$ for the slow variable q that evolves according to the following *effective* Fokker–Planck equation [10]:

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial q}\left(\frac{\partial}{\partial q}[D(q)f(q,t)] - V(q)f(q,t)\right). \tag{6}$$

If the effective drift $V(q)$ and the effective diffusion coefficient $D(q)$ can be explicitly written down as functions of q, then (6) can be used to compute interesting long-time properties of the system (e.g., the equilibrium distribution). Assuming that (6) provides a good approximation, we use $V(q) = \lim_{\Delta t\to 0} \frac{<Q(t+\Delta t)-q \mid Q(t)=q>}{\Delta t}$, and $D(q) = \lim_{\Delta t\to 0} \frac{<[Q(t+\Delta t)-q]^2 \mid Q(t)=q>}{\Delta t}$ to estimate the average drift, V, and diffusion coefficient D from short-time bursts of appropriately initialized detailed ("microscopic") simulations [11–14]. The steady solution of (6) is proportional to

$\exp[-\beta\Phi(q)]$, where the effective free energy $\Phi(q)$ is defined as

$$\beta\Phi(q) = -\int_0^q \frac{V(q')}{D(q')}\mathrm{d}q' + \ln D(q) + \text{constant}. \tag{7}$$

Consequently, computing the effective free energy and the equilibrium probability distribution could be accomplished without the need for long-time detailed simulations. Processing the results of multiple, short simulation bursts allows computational estimation of the effective drift $V(q)$ and the effective diffusion coefficient $D(q)$ (using the above definitions averaged over many replicas). $\Phi(q)$ is then computed by numerical evaluation of the formula (7). The estimation of drift and diffusion coefficients for an effective Fokker–Planck equation assumes the knowledge of suitable slow reaction coordinates that describes the system dynamics.

The effective free energy profile shown in Fig. 4 (top panel) is computed by binning long time simulation data in the (empirically determined, "MAN") coordinate $R(t)$ and using the relationship (at equilibrium) $P(R) \propto \exp(-\beta\Phi)$ where $P(R)$ is the probability distribution of R. This effective free energy profile is characterized by a number of shallow local minima, roughly equally separated by a value of $L_{zod} = 0.1$ in the reaction coordinate. When the group is "stuck", the rate of change of $\overline{c}(t)$ (the second quantity in (5)) is close to zero, and changes in $R(t)$ (for this state) largely consist of changes in the informed position ($c_{inf}(t)$) by approximately $\pm\mu_s\Delta t$. In the slip state both terms in (5) vary; $\overline{c}(t)$ changes at each time step by approximately $\frac{1}{N}(N_R - N_L)\mu_s\Delta t$, where N_R (N_L) is the total number of individuals traveling to the right (left). The speed of the group, estimated from the difference in the group centroid location over time windows of 250 and of 500 steps, is plotted as a function of the reaction coordinate $R(t)$ in the inset in (Fig. 4) (top panel); there is clearly a change in group speeds between the stick and slip regimes.

The bistable effective free energy profile shown in Fig. 4 (bottom panel) bins long time simulation data according to values of the centroid speed, and uses the equilibrium relationship between the effective free energy and the stationary distribution of centroid speed.

5 Diffusion Maps and Data-Based Analysis

We now describe a *diffusion map* approach that automates the detection of suitable coarse coordinates by data-mining detailed simulation results.

We first define a pairwise similarity matrix $\mathbf{K}$ between simulation datapoints where each "datapoint" is a vector $\mathbf{x}$ with components consisting of the (sorted) distances of all naive individuals to the informed individual in the group

$$K_{i,j} = K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \exp\left[-\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\sigma}\right)^2\right]; \tag{8}$$

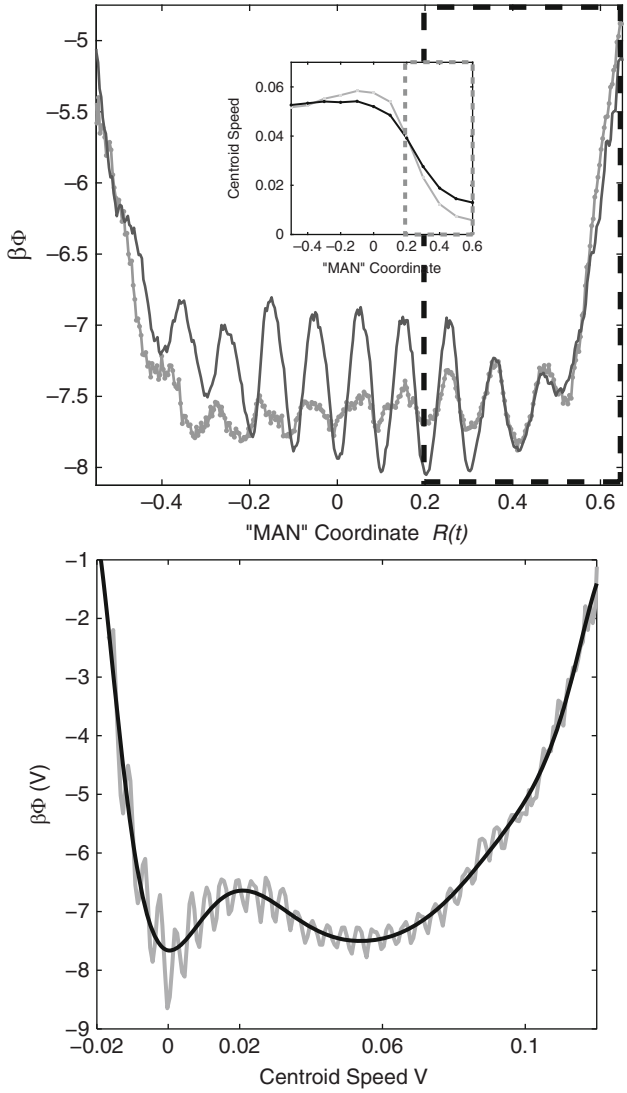

Fig. 4 *Top*: Effective free energy profile computed from long time simulations (*thick black line*) and from coarse estimation (*thin grey line*) in terms of the "MAN" coordinate $R(t)$ (5). The local minima in the effective free energy profile are separated by a value of approximately $L_{ZOD} = 0.1$. Inset: plot of centroid speed vs. the "MAN" coordinate $R(t)$; line shade distinguishes number of time steps over which centroid speed is estimated (*grey line*: 250 steps, *black line*: 500 steps). *Dashed boxes* in this figure suggest the "stick" regimes. *Bottom*: Effective free energy $\beta\Phi(V)$ (*grey line*) in terms of the centroid speed (computed from binning centroid speeds from a long time simulation and using $P(V) \propto \exp(-\beta\Phi)$). The simulation database consists of 200,000 time steps, with the centroid speed estimated over intervals of 200 steps. The *black line* is a polynomial fit to the effective potential

σ is a parameter that determines the "neighborhood size" within which the Euclidean distance between two data vectors provides a meaningful measure of their similarity. We use a small representative sample of the data in the construction of $\mathbf{K}$. Defining the diagonal normalization matrix $D_{i,i} = \sum_j K_{i,j}$, we construct the Markovian matrix $\mathbf{M} = \mathbf{D}^{-1}\mathbf{K}$. We show below that the components of the simulation data in a few of the leading eigenvectors of the matrix $\mathbf{M}$ may be used as a low dimensional representation of the data.

The matrix $\mathbf{M}$ is adjoint to the symmetric matrix $\mathbf{M_s}$

$$\mathbf{M_s} = \mathbf{D}^{1/2}\mathbf{M}\mathbf{D}^{-1/2} = \mathbf{D}^{-1/2}\mathbf{K}\mathbf{D}^{-1/2} \tag{9}$$

which shares its eigenvalues; eigenvectors of $\mathbf{M}$, denoted $\boldsymbol{\phi}_j$, are related to those of $\mathbf{M}_s$, denoted $\boldsymbol{\psi}_j$ as follows

$$\boldsymbol{\phi}_j = \mathbf{D}^{-1/2}\boldsymbol{\psi}_j \tag{10}$$

The *diffusion map* curve shown in Fig. 5 plots components of the top two, non-trivial, eigenvectors of $\mathbf{M}$ obtained by eigendecomposition of $\mathbf{M_s}$ with the use of (10). It is clear that points in this diffusion map are ordered along a curve according to the value of the diffusion map coordinate $\Phi(2)$ associated with each simulation datapoint. The diffusion map calculation has thereby provided an alternative (we will call it the "MACHINE") reaction coordinate through eigenprocessing of simulation data. We find a remarkable one-to-one correspondence between this "MACHINE" coordinate $\Phi(2)$ and the "MAN" coordinate $R(t)$, shown in the top inset in Fig. 5. Information on the coordinate $R(t)$ is not passed to the diffusion map calculation routine, yet this data-mining approach locates a coordinate that clearly and strongly correlates with $R(t)$. We also note that the Nyström formula [15] allows us to approximate the diffusion map coordinates of new simulation points, outside the original sample (used to construct $\mathbf{M_s}$).

6 Conclusions

Coarse-graining of the collective dynamics of animal group models may be enhanced through the detection of good *coarse variables* (observables). These can be proposed based on experience, obtained through extensive observation of system simulations; alternatively, they can be obtained using manifold learning techniques, such as the diffusion maps demonstrated here. In our illustrative example the essential dynamics exhibited bistability between a "slip" and a "stick" state; we note, however, that the observed bistability was especially sensitive to the "agent response time" parameter Δt. The bistable dynamics could be effectively characterized in terms of a single reaction coordinate; an effective free energy surface was obtained that succinctly summarizes the bistable dynamics. What was particularly gratifying here was the strong correlation between an empirical coordinate, based on extensive observations of the problem, and the coordinate discovered by the computer-assisted

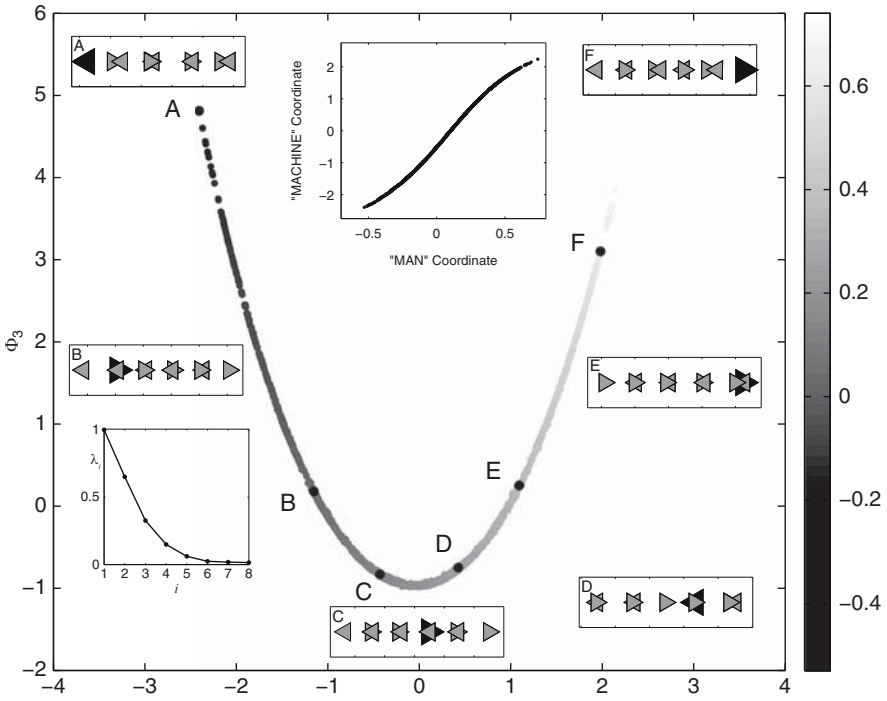

Fig. 5 Diffusion map results. The main plot shows the top two eigenvector components of the simulation data, with points shaded according to value of "MAN" coordinate $R(t)$ (scale shown in shadebar). The *bottom left* inset shows the leading eigenvalues of the corresponding Markov matrix. Individual configurations at several representative points (A,B,C,D,E,F) along the diffusion map curve are included. The instantaneous directions of the naive individuals (*grey triangles*) and of the informed individual (*black triangle*) are indicated by the triangle orientation. Top inset: plot of "MAN" coordinate $R(t)$ vs. the "MACHINE" coordinate $\Phi(2)$ for all simulation data points

Diffusion Map approach. In [16], this framework was shown to be equally effective in coarse-graining the dynamics of a 1-dimensional individual-based model of animal group formation with signaling constraints. This model has no informed individuals but for certain parameters exhibits "stick-slip" dynamics. This suggests that data-mining techniques can be naturally and efficiently linked with computational multiscale algorithms, in a way that can help accelerate the extraction of coarse-grained, system-level information from direct, individual-based models.

Acknowledgements This work was partially supported by the National Science Foundation and by the US AFOSR.

References

1. Couzin, I. D., Krause, J.: Self-organization and collective behavior in vertebrates. Adv. Study Behav. **32** (2003) 1–75
2. Deneubourg, J. L., Goss, S.: Collective patterns and decision making. Ethol. Ecol. Evol. **1** (1989) 295–311
3. Partridge, B. L.: The structure and function of fish schools. Sci. Am. **246** (1982) 114–123
4. Parrish, J. K., Edelstein-Keshet, L.: Complexity, pattern, and evolutionary trade-offs in animal aggregation. Science **284** (1999) 99–101
5. Couzin, I.: Collective Minds. Nature **445** (2007) 715
6. Couzin, I. D., Krause, J., Franks, N. R., Levin, S.: Effective leadership and decision making in animal groups on the move. Nature **433** (2005) 513–516
7. Coifman, R., Lafon, S., Lee, A., Maggioni, M., Nadler, B., Warner, F., Zucker, S. Geometric Diffusions as a tool for Harmonic Analysis and structure definition of data. Part I: Diffusion maps. Proc. Natl. Acad. Sci. USA **102** (2005) 7426–7431
8. Kevrekidis, I. G., Gear, C. W., Hummer, G.: Equation-free: The computer-aided analysis of complex multiscale systems. AIChE J. **50** (2004) 1346–1355
9. Kolpas, A., Moehlis, J., Kevrekidis, I. G.: Coarse-grained analysis of stochasticity-induced switching between collective motion states. Proc. Natl. Acad. Sci. USA **104** (2007) 5931–5935
10. Risken, H.: The Fokker-Planck Equation, Methods of solution and applications. Springer, Berlin (1989)
11. Haataja, M., Srolovitz, D., Kevrekidis, I. G.: Apparent hysteresis in a driven system with self-organized drag. Phys. Rev. Lett. **92** (2004) 160603
12. Hummer, G., Kevrekidis, I. G.: Coarse Molecular Dynamics of a Peptide Fragment: Free Energy, Kinetics and Long Time Dynamics Computations. J. Chem. Phys. **118** (2003) 10762–10773
13. Kopelevich, D., Panagiotopoulos, A. Z., Kevrekidis, I. G.: Coarse-Grained Kinetic Computations for Rare Events: Application to Micelle Formation. J. Chem. Phys. **122** (2005) 044908
14. Sriraman, S., Kevrekidis, I. G., Hummer, G.: Coarse Nonlinear Dynamics and Metastability of Filling-Emptying Transitions: Water in Carbon Nanotubes. Phys. Rev. Lett. **95** (2005) 130603
15. Baker, C.: The Numerical Treatment of Integral Equations. Clarendon Press, Oxford (1977)
16. Kolpas, A., Frewen, T. A., Moehlis, J., Kevrekidis, I. G.: Coarse analysis of collective motion with different communication mechanisms. Math. Biosci. **214** (2008) 49–57