# Supplementary Information

## CONTENTS

## Appendix A: Methods

### 1. Particle simulations

The microscopic model in main text Eqs. (1) in the main text has been previously studied for fixed particle velocities $v_i = v_0$ and rotation frequencies $\Omega_i = \Omega_0$. In this scenario, particles form small clusters of aligned particles and each cluster orbits on a circle of radius $\sim v_0/\Omega_0$ [1]. To generate the microscopic test data used in Sec. 1 D (main text), we considered instead a heuristic distribution $p(v_i, \Omega_i)$ for which particles spontaneously organize into proper vortices (main text Fig. 2$A$, top). It is convenient to define and draw from this distribution using propagation speeds $v_i$ and the curvature radii $R_i = v_i/\Omega_i$ of a particle's noise-free trajectory as independent variables. In particular, we considered $\tilde{p}(v_i, R_i) \sim G(v_i; \mu_v, \sigma_v)G(R_i; \mu_R, \sigma_R)$, where $G(x; \mu_x, \sigma_x)$ represents a Gaussian normal distribution with mean $\mu_x$ and standard deviation $\sigma_x$. $\tilde{p}(v_i, R_i)$ then defines $p(v_i, \Omega_i)$ implicitly through the relation $v_i = \Omega_i R_i$. In units of the characteristic scales – mean velocity $\langle v_i \rangle_p$ and interaction radius $R$ – the particle properties $v_i$ and $\Omega_i = v_i/R_i$ used for simulating main text Eqs. (1) (see main text Fig. S1) were drawn from $\tilde{p}(v_i, R_i)$ with $\mu_v = 1$ ($\langle v_i \rangle_{\tilde{p}} = \langle v_i \rangle_p = 1$), $\sigma_v = 0.4$, $\mu_R = 2.2$ and $\sigma_R = 1.7$. From these samples, we finally removed all particles with $\Omega_i > 1.4$.

For simulations of the microscopic model in main text Eqs. (1), we set $g \simeq 0.018$ and $D_r \simeq 0.009$ ($\Rightarrow D_r \ll \langle \Omega_i \rangle_p$, $D_r \ll \langle v_i \rangle_p$) and initially placed particles randomly distributed and oriented on a domain of size $100 \times 100$ (in units of the interaction radius). Equations (1) were then numerically integrated for particles interacting within the interaction radius $R = 1$ using the Euler-Maruyama method with a time step of $dt \simeq 0.0176$. For the subsequent coarse-graining, the data were saved at time intervals of $\Delta t \simeq 0.44$.

### 2. Kernel coarse-graining with periodic and non-periodic boundaries

To coarse-grain the discrete microscopic data through main text Eqs. (2), we used a 2D Gaussian kernel

$$K(\mathbf{x}) = (2\pi\sigma^2)^{-1} \exp\left(-|\mathbf{x}|^2/2\sigma^2\right). \tag{A1}$$

Periodicity of the coarse-grained fields for the microscopic test data (main text Sec. 1 B) was ensured by placing ghost particles periodically around the domain.

Coarse-graining in non-periodic domains (main text Sec. 2) was performed by truncating and renormalizing the kernel. This was achieved by defining the integral over the non-periodic domain $\mathcal{X}$ to be $N(\mathbf{x}) = \int_{\mathcal{X}} d^2\mathbf{x}' \, K(\mathbf{x}' - \mathbf{x})$, and then replacing $K[\mathbf{x} - \mathbf{x}_i(t)]$ with $K[\mathbf{x} - \mathbf{x}_i(t)]/N(\mathbf{x})$ in main text Eqs. (2). This renormalization ensured that the coarse-grained density $\rho(t, \mathbf{x})$ integrated to the total particle number and strongly reduced artefacts near the boundary.

### 3. Sparse regression and model selection

To perform sparse regression using the sequentially thresholded least squares (STLSQ) algorithm [2], we used the same parameters when working with data from the test microscopic model (main text Sec. 1 A) as well as for the experiments using Quincke rollers (main text Sec. 2) and sunbleak fish (SI Sec. F). As the sparse regression approach used to infer hydrodynamic equations only requires the evaluation of input data within the data domain, it is independent of the system's boundary conditions. Details of the concrete steps of the learning framework are provided in the following.

**Construction of linear systems**: To construct the linear system $\mathbf{U}_t = \boldsymbol{\Theta}\boldsymbol{\xi}$, we randomly sampled the coarse-grained fields at $N_d = 5 \times 10^5$ time-space points. The explicit form of the linear systems constructed for main text
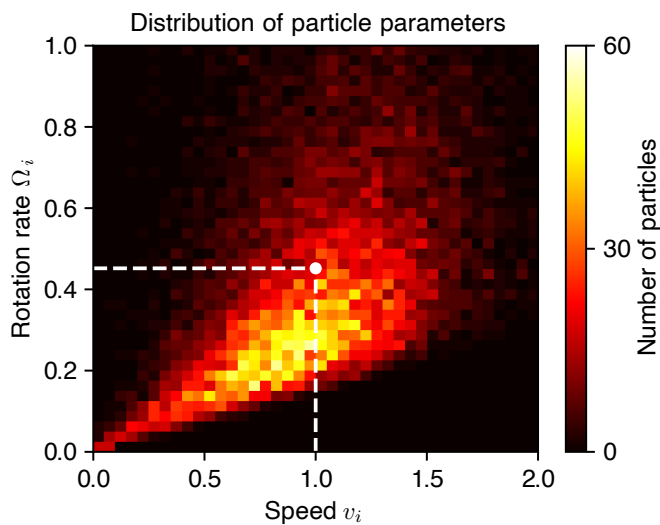


FIG. S1. Distribution of particle speeds $v_i$ and rotation rates $\Omega_i$ used to simulate main text Eqs. (1). Those parameters were drawn from a heuristic distribution $p(v_i, \Omega_i)$ that is explained in more detail in SI Sec. A 1. The white marker and dashed lines depict the mean velocity $\langle v_i \rangle_p \simeq 1$ and $\langle \Omega_i \rangle_p \simeq 0.45$.

Eqs. (4) was given by

$$
\underbrace{\begin{bmatrix} \vdots \\ \partial_t \rho \\ \vdots \end{bmatrix}}_{\mathbf{U}_t(N_d \times 1)} = \underbrace{\begin{bmatrix} \vdots & & \vdots \\ \nabla \cdot \mathbf{p} & \cdots & \nabla \cdot (\rho \mathbf{p}_\perp) \\ \vdots & & \vdots \end{bmatrix}}_{\mathbf{\Theta}(N_d \times r)} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_r \end{bmatrix}}_{\xi(r \times 1)}, \tag{A2a}
$$

$$
\underbrace{\begin{bmatrix} \partial_t p_x \\ \vdots \\ \hline \partial_t p_y \\ \vdots \end{bmatrix}}_{\mathbf{U}_t(2N_d \times 1)} = \underbrace{\begin{bmatrix} (\nabla \rho)_x & \cdots & ((\mathbf{p} \cdot \nabla)\mathbf{p})_x \\ \vdots & & \vdots \\ \hline (\nabla \rho)_y & \cdots & ((\mathbf{p} \cdot \nabla)\mathbf{p})_y \\ \vdots & & \vdots \end{bmatrix}}_{\mathbf{\Theta}(2N_d \times m)} \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}_{\boldsymbol{\xi}(m \times 1)}. \tag{A2b}
$$

Here, the subscripts denote components of the vectors, and $r, m$ are the total number of library terms in each equation. The vertical dots denote the respective terms evaluated at different time-space $(t, \mathbf{x})$ locations. The linear system in Eq. (A2b) was generated by stacking data for the $x$- and $y$-components of the time-derivatives and the library terms. Such a construction enforced the same coefficients for both the components of the polarization equation, ensuring rotational invariance (coordinate-independence) of the learned PDE. For the Quincke roller system (main text Sec. 2, main text) and the sunbleak system (SI Sec. F) linear systems analog to Eqs. (A2) were constructed for $N_d = 5 \times 10^5$ sampling points each, with term libraries described in the corresponding sections.

**Pre-processing**: Since the thresholding hyperparameter $\tau$ in STLSQ is agnostic to the scales of the library terms, as a pre-processing step, we performed transformations so that columns of the data matrix $\mathbf{\Theta}$ had zero mean and unit variance, and the time-derivative vector $\mathbf{U}_t$ had zero mean. Column standardization has been widely applied in equation discovery approaches [3–5], where it can mitigate numerical resolution limitations and may easily be extended to regularized regression techniques, such as ridge regression or LASSO.

**Stability selection** [5]: With equal spacing on a $\log_{10}$ scale, we chose 40 values for $\tau$ over the regularization path $[\tau_{\max}, \epsilon\tau_{\max}]$. The value of $\tau_{\max}$ was chosen so that all the terms get thresholded out and $\epsilon$ was set to $10^{-2}$. For every $\tau$, the data were split into 200 sub-samples each with 50% randomly selected data points. Every library term was assigned an importance score as the fraction of sub-samples in which it was learned by STLSQ; in general, this importance score was larger for smaller values of $\tau$. Along the regularization path, unique combinations of terms that had an importance score larger than 0.6 were considered and their coefficients were refitted to the full data without normalization. This procedure resulted in a small number of PDEs of increasing complexity (Figs. 2$E$, 3$D$, and 4$D$).

**Alternative approaches for model selection**: Information criteria, such as AIC or BIC, as well as Pareto analysis are alternative approaches that have previously been employed to perform model selection [6]. Typically, these methods depend on the regression residuals of the point-wise derivatives in real space to select a final model. However, for the complex active matter systems studied in this work, we observe that the addition of terms leads in some cases only to minor changes in such residuals, while it has a significant impact on the numerical long-term stability and phenomenology of the resulting models. This motivated us to use stability selection instead to first discover the terms that should appear in the model with statistical robustness to data sub-sampling and in a second step perform model selection based on a verification of numerical stability, as well as phenomenological and quantitative agreement.

**Number of sample points and importance score threshold**: To determine a suitable number of sample points $N_d$, we analyzed changes of the importance score of each term when learning the density dynamics of the microscopic active particle system described by main text Eqs. (1) (see main text Fig. A 3). As sample size increases, convergence towards the known ground truth (see SI Sec. B 2) can indeed be observed: the importance score becomes 1.0 for the expected term $\nabla \cdot \mathbf{p}$ and continuously decreases for all other terms. A similar behavior of the the Quincke roller and sunbleak fish data motivated us to fix the sample size to $5 \times 10^5$ for all problems studied in this work.
In our work, the key role of the importance score threshold is to define a sequence of candidate models with increasing complexity. We found that the sequence in which models appeared is not altered when the importance score threshold is varied within a range of 0.6–0.8, which is due to a monotonic and steep increase of the importance score for most library terms (see SI Figs. A 3 and A 3). To be consistent throughout this work, we fixed the importance score threshold to 0.6.
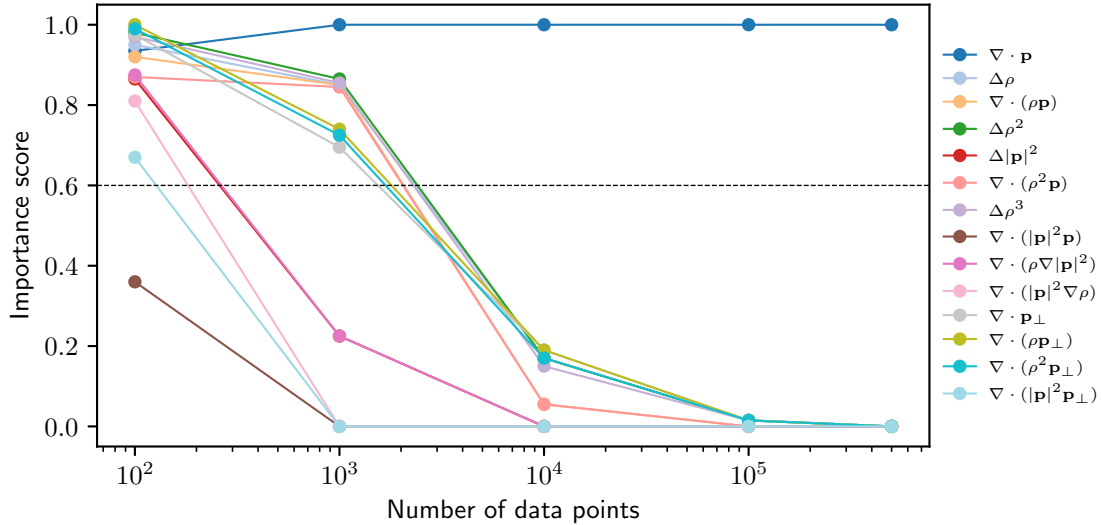
FIG. S2. Importance score evolution of each term when learning the density dynamics of the microscopic active particle system [main text Eqs. (1)] for thresholding parameter $\tau = 0.1\tau_{\max}$ at different sample sizes $N_d$. At a sample size of $N_d = 5 \times 10^5$ the expected term $\nabla \cdot \mathbf{p}$ is learned with an importance score very close to 1.0. A sample size of $N_d = 5 \times 10^5$ and and an importance score of 0.6 (black dashed line) were used throughout this work.
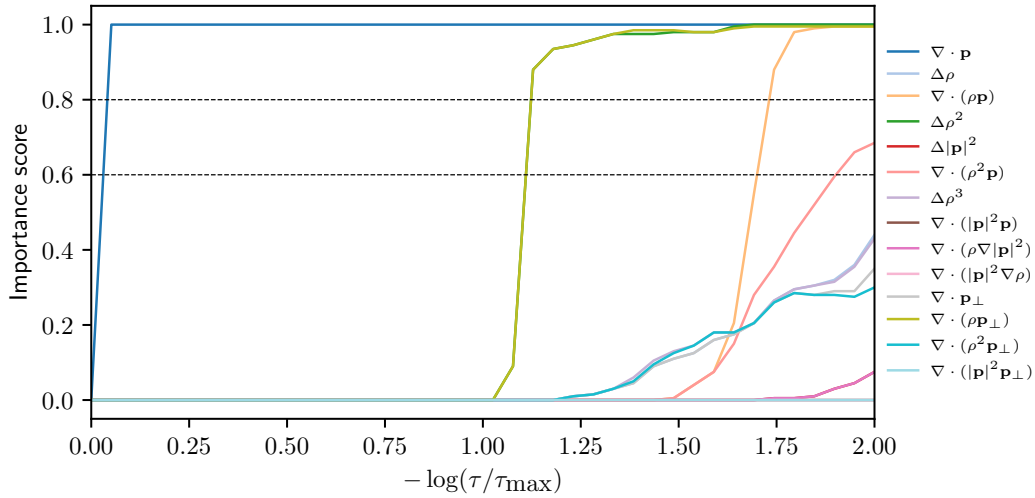


FIG. S3. The importance score obtained along the stability path $\tau \in [\tau_{\max}, \epsilon\tau_{\max}]$ (App. A 3) for each term in the library of the density equation learned from simulations of the microscopic active particle system (main text Eqs. (1)).

## 4.  Computational cost

The the computational cost of the overall learning algorithm is comprised of two computationally independent steps. The first is to represent the coarse-grained data in terms of a suitable spectral basis and in the second step the sparse regression is performed.

**Spectral representation**: The coarse-grained hydrodynamic fields [main text Eqs. (2)] were evaluated at $[N_t, N_x, N_y]$ uniformly spaced grid points in the respective directions. The resulting discrete data were projected

FIG. S4. The importance score obtained along the stability path $\tau \in [\tau_{\max}, \epsilon\tau_{\max}]$ (App. A 3 in the main manuscript) for each term in the library of the polarization dynamics equation learned from simulations of the microscopic active particle system (main text Eqs. (1)).
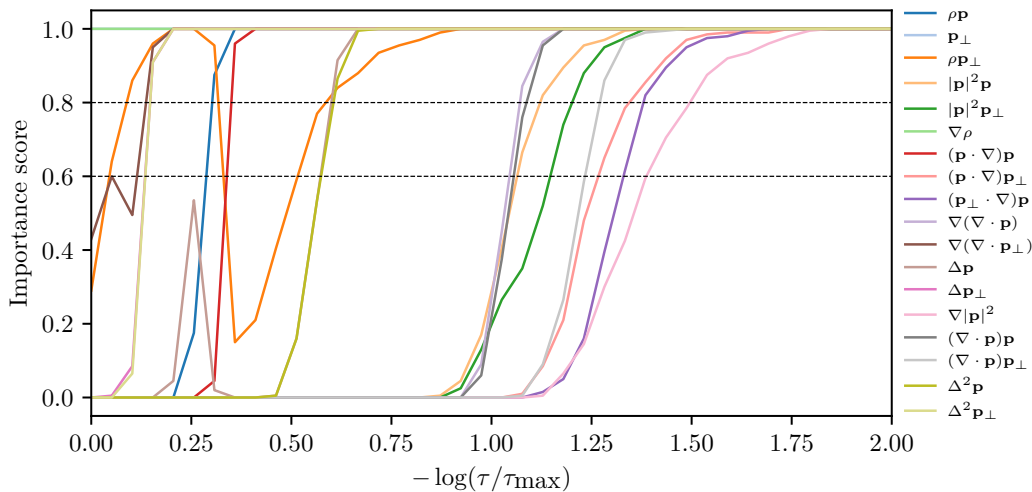
onto the spectral basis functions [main text Eqs. (3)] using multidimensional discrete cosine and Fourier transforms provided by the FFTW library [7], with an efficient time complexity of $O(N \log(N))$, where $N = N_t N_x N_y$. For the Chebyshev transforms, the data were interpolated onto the required Chebyshev extrema grid using spline functions of degree 5.

**STLSQ**: The computational cost of STLSQ is dominated by the least-square fit performed in the first thresholding step. The latter involves computing the singular value decomposition of $\boldsymbol{\Theta}$ [Eqs. (A2)] and thus scales as $\mathcal{O}(r^2 N_d)$, where $r \ll N_d$ is the number of terms in the library ($< 20$) and $N_d$ is the total number of sampled points ($5 \times 10^5$). When learning a model from a given data set, this scaling acquires a constant prefactor that contains the number of sub-samples on which STLSQ was repeatedly applied as a part of stability selection for each value of the thresholding parameter. In this work, we have used 200 subsamples and a total of 40 different values of the thresholding parameter $\tau$ (see SI Sec. A 3), amounting to a constant pre-factor of $\mathcal{O}(10^4)$.

### 5. Linear dependencies of the library terms

The procedure outlined in main text Sec. 1 D 1 leads to a few library terms for the polarization dynamics [main text Eq. (4b)] that are linearly dependent on each other. For completeness, we provide here a list of non-trivial identities that can be used to eliminate these dependencies:

$$\frac{1}{2}\nabla|\mathbf{p}|^2 = (\mathbf{p}\cdot\nabla)\mathbf{p} + (\nabla\cdot\mathbf{p}_\perp)\mathbf{p}_\perp, \tag{A3a}$$

$$(\nabla\cdot\mathbf{p})\mathbf{p}_\perp = (\nabla\mathbf{p})\cdot\mathbf{p}_\perp + (\mathbf{p}\cdot\nabla)\mathbf{p}_\perp, \tag{A3b}$$

$$(\mathbf{p}\cdot\nabla)\mathbf{p}_\perp + (\mathbf{p}_\perp\cdot\nabla)\mathbf{p}$$
$$= (\nabla\cdot\mathbf{p})\mathbf{p}_\perp + (\nabla\cdot\mathbf{p}_\perp)\mathbf{p}. \tag{A3c}$$

In Eq. (A3b), we follow the convention, $[(\nabla\mathbf{a})\cdot\mathbf{b}]_i = b_j\partial_i a_j$, with $i = x, y$ and repeated indices indicating summation. One may set $\mathbf{p} \to \mathbf{p}_\perp$ and $\mathbf{p}_\perp \to -\mathbf{p}$ in Eqs. (A3a) and (A3b) to obtain two additional identities.

To ensure in general that no linear dependencies remain after a library of terms has been constructed, it should be checked that the columns the data matrix $\boldsymbol{\Theta}$ are linearly independent, for example by using a singular value decomposition of $\boldsymbol{\Theta}$.

## 6. Numerical simulations of learned PDEs

Continuum simulations were performed using the spectral PDE solver Dedalus [8] with four-step Runge-Kutta time stepping scheme RK443. For simulation of the PDEs learned from the microscopic test data (main text Fig. $3E$) we used $256 \times 256$ Fourier modes in a doubly periodic domain with time step $4 \times 10^{-3}$. To facilitate a comparison, simulations shown in main text Fig. $3E$ were initialized using the initial density and polarization field of the coarse-grained particle data. It was verified that similar vortex patterns also form from fully random initial conditions.

For the doubly periodic simulation of Eqs. (7) (main text Fig. $4E$), we used $1024 \times 1024$ Fourier modes and time step $10^{-4}$s. The initial conditions were random with mean density 0.11, and mean horizontal and vertical velocities, $\langle v_x \rangle = 0.1\,\mathrm{mm\,s^{-1}}$ and $\langle v_y \rangle = 0$, respectively.

The simulation presented in main text Fig. $4F$ was performed on a confined square domain using the Sine/Cosine basis functions with $1024 \times 1024$ modes and time step $10^{-4}$s. The basis combinations in the $(x, y)$ directions were chosen to be (cos, cos) for density $\rho$, (sin, cos) for $v_x$ and (cos, sin) for $v_y$. These imply that normal density gradients, normal velocities and all remaining shear rates $\partial_x v_y$ and $\partial_y v_x$ vanish at the domain boundaries. The simulations were initialized with random perturbations around a mean density of 0.11. Simulations of models learned for the confined motion of sunbleak fish (see SI Sec. F) were run with the same boundary conditions and basis functions with $256 \times 256$ modes and time step $10^{-3}$ s.

Since the model in Eqs. (7) learned for the Quincke roller dynamics generates density shock waves, we added numerical diffusivities of $10^{-4}\mathrm{mm^2/s}$ (main text Fig. $4E$) and $10^{-3}\mathrm{mm^2/s}$ (main text Fig. $4F$) in both density and velocity equation to avoid Gibbs ringing.

## Appendix B: Analytic coarse-graining of the particle model

We describe in this section two approaches to analytically determine mean-field approximations of the microscopic model [main text Eqs. (1)] and compare their predictions and methodology with our learning approach. In general, analytic coarse-graining can ($i$) provide guidance for developing a physics-informed learning library, ($ii$) allows discussing our PDE learning framework as a tool to effectively infer moment closure relations, and ($iii$) predict the dependency of PDE coefficients of linear terms on distributions of microscopic parameters, which can be used to validate learned hydrodynamic models. However, to produce interpretable mean-field equations, analytic coarse-graining procedures typically have to

(i) factorize pair-correlations;

(ii) impose closure relations; and

(iii) neglect microscopic parameter variability.

Such approximations particularly affect non-linear terms and terms with higher-order derivatives in the resulting coarse-grained equations. It is, therefore, instructive to directly compare the coefficients from analytic approximations against the results obtained by hydrodynamic model learning, where the latter effectively infers a suitable expansion of pair-correlations, as well as a closure relation directly from the underlying data and naturally integrates microscopic parameter variability.

### 1. Dynamic equation of the one-particle probability density

We first describe a common analytic coarse-graining procedure that is often applied to microscopic models like main text Eqs. (1) *for constant and homogeneous parameters* $v_i = v_0$ and $\Omega_i = \Omega_0$. Specifically, this approach aims to find an approximate dynamic equation for the one-particle probability density [1, 9–14]

$$f(t, \theta, \mathbf{x}) = \sum_{i=1}^{N} \langle \delta \left( \theta - \theta_i(t) \right) \delta \left( \mathbf{x} - \mathbf{x}_i \right) \rangle, \tag{B1}$$

where $\langle \cdot \rangle$ denotes a Gaussian white noise average. To find such a dynamic equation, this approach proceeds in two steps: First, a dynamic equation for the angular moments of $f(t, \theta, \mathbf{x})$ is derived and second, to truncate the infinite hierarchy of dynamic equations, a closure relation is imposed.

Neglecting multiplicative noise and factorizing pair-correlations gives rise to a nonlinear integro-differential equation [1, 11] that can be transformed into an infinite hierarchy of coupled PDEs for the angular moments $f_n(t, \mathbf{x})$ defined by

$$f_n(t, \mathbf{x}) = \int_0^{2\pi} d\theta \, f(t, \theta, \mathbf{x}) e^{in\theta}. \tag{B2}$$

For the microscopic model [main text Eqs. (1)] and equal microscopic parameters, $v_i = v_0$ and $\Omega_i = \Omega_0$ for all particles, this procedure leads to [1, 11]

$$\partial_t f_n + \frac{v_0}{2} \left[ \partial_x \left( f_{n+1} + f_{n-1} \right) - i \partial_y \left( f_{n+1} - f_{n-1} \right) \right] = n(i\Omega_0 - D_r n) f_n + \frac{g n \pi}{2} \left( f_{n-1} f_1 - f_{n+1} f_{-1} \right). \tag{B3}$$

Each complex angular moment $f_n$ can be identified as a mean-field variable that represents different orientational order parameters encoded by the probability density $f(t, \theta, \mathbf{x})$ [12]. In particular, $f_0$ represents the particle number density $\rho$ and $f_1 =: p_x + i p_y$ represents the polarization density $\mathbf{p} = (p_x, p_y)^\top$. These fields correspond to the coarse-graining information in main text Eqs. (2) that our learning framework extracts explicitly from given microscopic data. For $n = 0$ and $n = 1$, we can therefore write Eqs. (B3) as

$$\partial_t \rho + v_0 \nabla \cdot \mathbf{p} = 0, \tag{B4a}$$

$$\partial_t \mathbf{p} + \frac{v_0}{2} (\nabla \rho + \nabla \cdot \mathbf{Q}) = \Omega_0 \mathbf{p}_\perp - D_r \mathbf{p} + \frac{g\pi}{2} (\rho \mathbb{I} - \mathbf{Q}) \cdot \mathbf{p}, \tag{B4b}$$

which also shows the coupling to the next higher mode $f_2 =: Q_{xx} + i Q_{xy}$, corresponding to the independent degrees of freedom of a nematic tensor. The chiral term $\sim \Omega_0 \mathbf{p}_\perp$ with $\mathbf{p}_\perp = (-p_y, p_x)^\top$ breaks the mirror symmetry. Terms constructed from $\mathbf{p}_\perp$ are therefore generally allowed in chiral systems and consequently included into the library in main text Eqs. (4).

The final step that is key to analytically close the infinite hierarchy of Eqs. (B3) requires the introduction of moment closure assumptions [10, 11, 15]. Depending on the structure of the mode coupling, the resulting closure relation allows to express the nearest coupled modes with $|n| = k$ in terms of modes $|n| < k$ and neglects the remaining modes. For example, in the case of Eq. (B4), a moment closure assumption must provide an expression $\mathbf{Q}(\rho, \mathbf{p})$ [11, 15]. As modes with higher mode numbers $n$ are increasingly suppressed by rotational noise, which can be seen by the prefactor $-n^2 D_r$ in Eq. (B3), a common closure assumption is $\partial_t \mathbf{Q} \approx 0$ [1, 12, 14]. Equation (B3) for $n = 2$ then implies the desired closure relation $\mathbf{Q}(\rho, \mathbf{p})$ that can be used in Eq. (B4b) and leads to

$$\begin{aligned}
\partial_t \mathbf{p} = &-\frac{v_0}{2} \nabla \rho + \frac{\bar{g}\rho}{2} \mathbf{p} + \Omega \mathbf{p}_\perp - D_r \mathbf{p} \\
&+ \frac{1}{8} \frac{D_r}{\Omega_0^2 + 4 D_r^2} \left[ 2 v_0^2 \Delta \mathbf{p} + \bar{g} v_0 \left( 5 \nabla |\mathbf{p}|^2 - 6 \mathbf{p} \cdot \nabla \mathbf{p} - 10 \mathbf{p} \nabla \cdot \mathbf{p} \right) - 4 \bar{g}^2 \mathbf{p}^2 \mathbf{p} \right] \\
&+ \frac{1}{8} \frac{\Omega_0}{\Omega_0^2 + 4 D_r^2} \left[ v_0^2 \Delta \mathbf{p}_\perp - \bar{g} v_0 \left( 3 \mathbf{p} \cdot \nabla \mathbf{p}_\perp + 5 \mathbf{p}_\perp \cdot \nabla \mathbf{p} \right) - 2 \bar{g}^2 \mathbf{p}^2 \mathbf{p}_\perp \right],
\end{aligned} \tag{B5}$$

where we defined $\bar{g} = \pi g$. Equation (B5) is equivalent to the result given in Ref. [1]. The coefficients in Eq. (B5) are listed and computed in Tab. S I, where we used $\Omega_0 = \langle v_i \Omega_i \rangle_p / \langle v_i \rangle_p = 0.5$ and $v_0 = \langle v_i^2 \rangle_p / \langle v_i \rangle_p = 1.14$. These values for $\Omega_0$ and $v_0$ are suggested by an analytic kernel coarse-graining (see SI Sec. B 2) of the microscopic model with a distribution $p(v_i, \Omega_i)$ of kinetic particle parameters, as considered in this work (see SI Sec. A 1).

As discussed in the main text (see Sec. 1 E and Tab. I), coefficients associated with terms linear in the fields and derivatives inferred from the model learning agree well with analytic predictions. However, Tab. S I shows that coefficients associated with terms that are non-linear in the fields or derivatives can substantially differ between these two approaches (e.g. $b_5$, $b_8$ or $b_{13}$ in Tab. S I) or they do not even appear in the analytically coarse-grained dynamics (e.g. $b_4$, $b_{11}$ or $b_{12}$ in Tab. S I). At the same time, the analytically coarse-grained values of these non-linear terms are most strongly affected by the various approximations listed in the introduction of SI Sec. B 1. Moreover, simulations of the analytically coarse-grained model (Eq. (B5) with coefficients given in Tab. S I, simulation

TABLE S I. Comparison of analytic coarse-graining (CG) results with learned coefficients suggest a limited validity of common analytic coarse-graining approximations in systems with microscopic parameter variability and mesoscopic pattern formation. Analytic coefficients (CG coefficient) have been obtained by coarse-graining the model in main text Eqs. (1) for constant microscopic parameters $v_i = v_0$ and $\Omega_i = \Omega_0$ [see SI Sec. B 1 and Eq. (B5)] using a common closure relation [1]. Models were learned (PDE 1, PDE 4 and PDE 8 shown here; see Tab. S III for the complete list) from coarse-grained fields (see main text Sec. 1 B and Fig. 3) of the microscopic dynamics in main text Eqs. (1) with a distribution of parameters $v_i$ and $\Omega_i$ (see SI Sec. A 1 and Fig. S1). Analytic coefficients were calculated (CG value) using $\Omega_0 = \langle v_i\Omega_i\rangle_p/\langle v_i\rangle_p = 0.5$, $v_0 = \langle v_i^2\rangle_p/\langle v_i\rangle_p = 1.14$, $D_r = 0.009$, $\bar{g} = \pi g = 1.1$ and $c_0 = (8\Omega_0^2 + 32D_r^2)^{-1}$. While the analytic coefficients for linear terms agree well with the learned models (see main text Sec. 1 E and Tab. I), coefficients of higher order terms in the fields and derivatives can differ significantly (e.g. $b_5$, $b_8$ or $b_{13}$) or are not predicted by the analytic coarse-graining to contribute to the mean-field dynamics (e.g. $b_4$, $b_{11}$ or $b_{12}$).

| Term | CG coefficient | CG value | PDE 1 | PDE 4 | PDE 8 ◄ |
|---|---|---|---|---|---|
| $b_1\mathbf{p}$ | $-D_r$ | -0.009 | -0.009 | -0.009 | -0.009 |
| $b_2\rho\mathbf{p}$ | $\bar{g}/2$ | 0.0283 | – | 0.013 | 0.009 |
| $b_3\mathbf{p}_\perp$ | $\Omega_0$ | 0.5 | 0.414 | 0.428 | 0.440 |
| $b_4\rho\mathbf{p}_\perp$ | – | – | – | – | -0.010 |
| $b_5|\mathbf{p}|^2\mathbf{p}$ | $-4c_0\bar{g}^2 D_r$ | $-6\times10^{-5}$ | – | – | -0.080 |
| $b_6|\mathbf{p}|^2\mathbf{p}_\perp$ | $-2c_0\bar{g}^2\Omega_0$ | -0.0016 | – | – | – |
| $b_7\nabla\rho$ | $-v_0/2$ | -0.57 | -0.638 | -595 | -0.595 |
| $b_8(\mathbf{p}\cdot\nabla)\mathbf{p}$ | $-6c_0\bar{g}v_0 D_r$ | -0.0017 | – | -0.536 | -0.463 |
| $b_9(\mathbf{p}\cdot\nabla)\mathbf{p}_\perp$ | $-3c_0\bar{g}v_0\Omega_0$ | -0.0483 | – | – | – |
| $b_{10}(\mathbf{p}_\perp\cdot\nabla)\mathbf{p}$ | $-5c_0\bar{g}v_0\Omega_0$ | -0.0805 | – | – | – |
| $b_{11}\nabla(\nabla\cdot\mathbf{p})$ | – | – | – | – | 0.078 |
| $b_{12}\nabla(\nabla\cdot\mathbf{p}_\perp)$ | – | – | – | 0.265 | 0.277 |
| $b_{13}\Delta\mathbf{p}$ | $2c_0 v_0^2 D_r$ | 0.0117 | – | – | -0.155 |
| $b_{14}\Delta\mathbf{p}_\perp$ | $c_0 v_0^2\Omega_0$ | 0.3245 | – | 0.202 | 0.196 |
| $b_{15}\nabla|\mathbf{p}|^2$ | $5c_0\bar{g}v_0 D_r$ | 0.0014 | – | – | – |
| $b_{16}(\nabla\cdot\mathbf{p})\mathbf{p}$ | $-10c_0\bar{g}v_0 D_r$ | -0.0029 | – | – | -0.225 |
| $b_{17}(\nabla\cdot\mathbf{p})\mathbf{p}_\perp$ | – | – | – | – | – |
| $b_{18}\Delta^2\mathbf{p}$ | – | – | – | – | -0.483 |
| $b_{19}\Delta^2\mathbf{p}_\perp$ | – | – | – | 1.197 | 1.235 |

parameters same as used for the learned model, see SI Sec. A 6) reveal a dynamics that does not exhibit any steady, saturated vortex patterns (see spatial snapshots in Fig. S5). Instead, the analytic model becomes quickly unstable when starting from a small, random density and polarization perturbation (see 'spatial maxima' plots in Fig. S5). Notably, our learning framework takes a different route and determines in a data-driven manner an effective closure relation that best explains the observed dynamics. Thus, a key approximation made in analytic coarse-graining approaches is circumvented and closures are determined that can account quantitatively for non-trivial effects from correlations and microscopic parameter variability.

## 2. Dynamic equations from conventional kernel coarse-graining

While the previous approach provides a clear coarse-graining strategy to find a closed set of PDEs from a system of stochastic ODEs with homogeneous microscopic parameters, it is more challenging to understand how the phenomenological coefficients will depend on the distribution $p(v_i, \Omega_i)$ of microscopic kinetic parameters described in SI Sec. A 1. We therefore consider an alternative strategy, for which we write main text Eqs. (1) as

$$\frac{d\mathbf{x}_i}{dt} = v_i\mathbf{p}_i, \tag{B6a}$$

$$\frac{d\mathbf{p}_i}{dt} = \Omega_i\boldsymbol{\epsilon}\cdot\mathbf{p}_i + \mathbf{F}_i, \tag{B6b}$$

where $\boldsymbol{\epsilon}\cdot\mathbf{p}_i = \mathbf{p}_{i,\perp} = (-\sin\theta_i, \cos\theta_i)^\top$, and $\mathbf{F}_i$ contains forces from interactions and rotational diffusion. Taking directly the time derivative of the coarse-graining prescription in main text Eq. (2a) and using Eq. (B6a), we find

$$\partial_t\rho(t,\mathbf{x}) + \nabla\cdot\mathbf{J}(t,\mathbf{x}) = 0, \tag{B7}$$
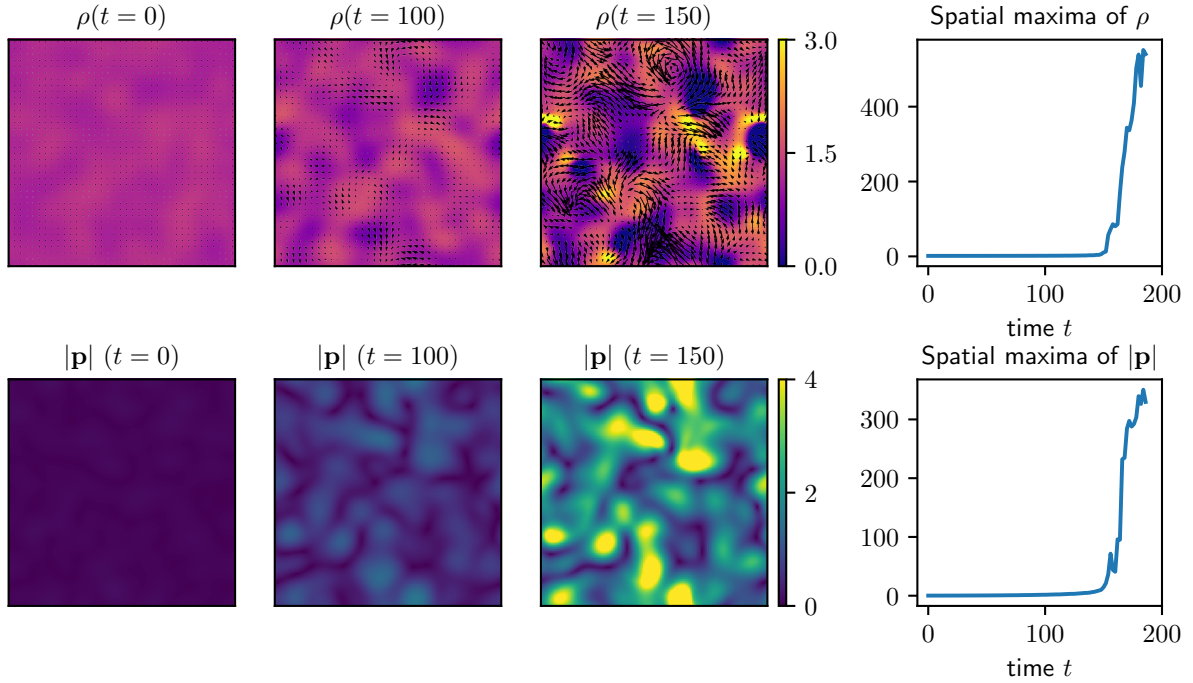
FIG. S5. Numerical simulation of the analytically coarse-grained model Eq. (B5) for parameters listed in Tab. S I ('CG value'), along with $\partial_t \rho = -\nabla \cdot \mathbf{p}$, with snapshots of the density field $\rho$ (top row) and snapshots of the polarization magnitude $|\mathbf{p}|$ (bottom row). The transient dynamics does not exhibit any coherent pattern formation and the simulation becomes unstable around times $t \sim 150$. The latter is illustrated in right-most panels that show the spatial maxima of density $\rho$ and $|\mathbf{p}|$, which both grow over two orders of magnitude before numerical values become intractable. Numerical parameters (see SI Sec. A 6) and box size are the same as for the simulations used in Fig. 3 of the main text.

where we have defined a flux

$$\mathbf{J}(t, \mathbf{x}) = \sum_i K\left[\mathbf{x} - \mathbf{x}_i(t)\right] v_i \mathbf{p}_i(t). \tag{B8}$$

Using this definition and Eq. (B6b), we find a dynamic equation for $\mathbf{J}$ of the form

$$\partial_t \mathbf{J}(t, \mathbf{x}) + \nabla \cdot \boldsymbol{\sigma}(t, \mathbf{x}) = \mathbf{T}(t, \mathbf{x}) + \boldsymbol{\Phi}(t, \mathbf{x}). \tag{B9}$$

Here, we have defined the tensor and vector fields

$$\boldsymbol{\sigma}(t, \mathbf{x}) = \sum_i K\left[\mathbf{x} - \mathbf{x}_i(t)\right] v_i^2 \mathbf{p}_i(t)\mathbf{p}_i(t), \tag{B10a}$$

$$\mathbf{T}(t, \mathbf{x}) = \boldsymbol{\epsilon} \cdot \sum_i K\left[\mathbf{x} - \mathbf{x}_i(t)\right] v_i \Omega_i \mathbf{p}_i(t), \tag{B10b}$$

$$\boldsymbol{\Phi}(t, \mathbf{x}) = \sum_i K\left[\mathbf{x} - \mathbf{x}_i(t)\right] v_i \mathbf{F}_i(t). \tag{B10c}$$

Averaging the fields in Eq. (B8) and Eqs. (B10) over the particle parameter distribution $p(v_i, \Omega_i)$ yields

$$\langle \mathbf{J}(t, \mathbf{x}) \rangle_p = \left\langle \sum_i K\left[\mathbf{x} - \mathbf{x}_i(t)\right] v_i \mathbf{p}_i(t) \right\rangle_p, \tag{B11a}$$

$$\langle \boldsymbol{\sigma}(t, \mathbf{x}) \rangle_p = \left\langle \sum_i K\left[\mathbf{x} - \mathbf{x}_i(t)\right] v_i^2 \, \mathbf{p}_i(t)\mathbf{p}_i(t) \right\rangle_p, \tag{B11b}$$

$$\langle \mathbf{T}(t, \mathbf{x}) \rangle_p = \left\langle \boldsymbol{\epsilon} \cdot \sum_i K\left[\mathbf{x} - \mathbf{x}_i(t)\right] v_i \Omega_i \, \mathbf{p}_i(t) \right\rangle_p, \tag{B11c}$$

$$\langle \boldsymbol{\Phi}(t, \mathbf{x}) \rangle_p = \left\langle \sum_i K\left[\mathbf{x} - \mathbf{x}_i(t)\right] v_i \mathbf{F}_i(t) \right\rangle_p. \tag{B11d}$$

We then adopt a moment factorization approximation

$$\langle \mathbf{J}(t, \mathbf{x}) \rangle_p \simeq \langle v_i \rangle_p \mathbf{p}, \tag{B12a}$$

$$\langle \boldsymbol{\sigma}(t, \mathbf{x}) \rangle_p \simeq \frac{1}{2} \langle v_i^2 \rangle_p \left( \rho \mathbb{I} + \mathbf{Q} \right), \tag{B12b}$$

$$\langle \mathbf{T}(t, \mathbf{x}) \rangle_p \simeq \langle v_i \Omega_i \rangle_p \mathbf{p}_\perp, \tag{B12c}$$

where we used the definition of the particle number density in main text Eq. (2a), the polarization density in main text Eq. (2b), and $|\mathbf{p}_i|^2 = 1$. Additionally, we have defined in Eq. (B12b) a nematic moment of the form

$$\mathbf{Q} = \sum_i K\left[\mathbf{x} - \mathbf{x}_i(t)\right] \left[2\mathbf{p}_i(t)\mathbf{p}_i(t) - \mathbb{I}\right]. \tag{B13}$$

Averaging Eqs. (B7) and (B9) over the microscopic parameter distributions and using Eqs. (B12), we obtain

$$\partial_t \rho + \langle v_i \rangle_p \nabla \cdot \mathbf{p} = 0, \tag{B14a}$$

$$\partial_t \mathbf{p} + \frac{\langle v_i^2 \rangle_p}{2 \langle v_i \rangle_p} \left( \nabla \rho + \nabla \cdot \mathbf{Q} \right) = \frac{\langle v_i \Omega_i \rangle_p}{\langle v_i \rangle_p} \mathbf{p}_\perp + \langle v_i \rangle_p^{-1} \langle \boldsymbol{\Phi} \rangle_p. \tag{B14b}$$

From this, we can read off predictions about the coefficients we expect to find from the learning framework for the terms $\nabla \cdot \mathbf{p}$, $\nabla \rho$ and $\mathbf{p}_\perp$ (Tab. I).

Following the same analytic coarse-graining strategy, but starting from a more general microscopic position dynamics

$$\frac{d\mathbf{x}_i}{dt} = v_i \mathbf{p}_i + v' \boldsymbol{\epsilon} \cdot \mathbf{p}_i, \tag{B15}$$

yields an additional term $\sim v' \nabla_\perp \rho$ in the polar dynamics SI Eq. (B14b). It is therefore a simple, non-interacting term – for which analytic coarse-graining reliably predicts coefficients (see Tab. SI) – that would result from a chiral propagation of particles. However, such a chiral propagation is absent in the microscopic model used in our work (see main text Eq. (1a), where $v_i' = 0$), such that the term $\nabla_\perp \rho$ was omitted from the library, which is equivalent to omitting the Levi-Civita tensor $\boldsymbol{\epsilon}$ from the set $\mathcal{S}$ given in Eq. (5).

### 3. Physical interpretation of higher-order terms and relation to microscopic particle properties

The analytic coarse-graining results in Tab. SI are based on the simplifying assumption of constant microscopic parameters $v_i = v_0$ and $\Omega_i = \Omega_0$ and invoke the various other approximations described in SI Sec. B 1. Nevertheless, a more detailed comparison with learning results is also instructive for terms that are non-linear in the fields or derivatives. In the following, we discuss this comparison for three different groups of terms.

- Terms that are present in the learned models and also expected by the analytic coarse-graining

$$\{b_5 |\mathbf{p}^2| \mathbf{p}, b_8 (\mathbf{p} \cdot \nabla)\mathbf{p}, b_{13} \Delta \mathbf{p}, b_{13} \Delta \mathbf{p}_\perp, b_{16} (\nabla \cdot \mathbf{p})\mathbf{p}\}$$

TABLE S II. Parameters $a_l$ of the density dynamics PDE (main text Fig. 2$D$) learned from simulations of the microscopic active particle system main text Eqs. (1). The sparsest model (◄) agrees well with the analytic coarse-graining prediction (Tab. I).

| Term | PDE 1◄ | PDE 2 | PDE 3 | PDE 4 |
|---|---|---|---|---|
| $a_1\nabla\cdot\mathbf{p}$ | -0.991 | -0.991 | -0.972 | -0.957 |
| $a_2\Delta\rho$ | – | – | – | – |
| $a_3\nabla\cdot(\rho\mathbf{p})$ | – | – | -0.015 | -0.037 |
| $a_4\Delta\rho^2$ | – | 0.022 | 0.022 | 0.022 |
| $a_5\Delta|\mathbf{p}|^2$ | – | – | – | – |
| $a_6\nabla\cdot(\rho^2\mathbf{p})$ | – | – | – | 0.008 |
| $a_7\Delta\rho^3$ | – | – | – | – |
| $a_8\nabla\cdot(|\mathbf{p}|^2\mathbf{p})$ | – | – | – | – |
| $a_9\nabla\cdot(\rho\nabla|\mathbf{p}|^2)$ | – | – | – | – |
| $a_{10}\nabla\cdot(|\mathbf{p}|^2\nabla\rho)$ | – | – | – | – |
| $a_{11}\nabla\cdot\mathbf{p}_\perp$ | – | – | – | – |
| $a_{12}\nabla\cdot(\rho\mathbf{p}_\perp)$ | – | -0.026 | -0.026 | -0.026 |
| $a_{13}\nabla\cdot(\rho^2\mathbf{p}_\perp)$ | – | – | – | – |
| $a_{14}\nabla\cdot(|\mathbf{p}|^2\mathbf{p}_\perp)$ | – | – | – | – |

can be understood as dynamic coupling of the polar order parameter $\mathbf{p}$ to the nematic order $\mathbf{Q}$ that naturally arise from symmetry arguments [see terms $\nabla\cdot\mathbf{Q}$ and $\mathbf{Q}\cdot\mathbf{p}$ in Eq. (B4b)]. In particular, we can consider a general expansion of nematic tensors constructed in terms of the polar order parameter field

$$\mathbf{Q} = \nu_1(\mathbf{pp} - \mathbb{I}/2) + \nu_2(\nabla\mathbf{p} + \nabla\mathbf{p}^\top - \nabla\cdot\mathbf{p}\mathbb{I}) + \nu_3(\nabla\mathbf{p}_\perp + \nabla\mathbf{p}_\perp^\top - \nabla\cdot\mathbf{p}_\perp\mathbb{I}) + ..., \tag{B16}$$

which is similar to defining a closure relation, where the coefficients $\nu_1, \nu_2, \nu_3, ...$ are here inferred via a data-driven approach.

- A second group of non-linear terms can be defined as those that the learning approach identifies as relevant contributions, but which would not be predicted by the analytic coarse-graining for constant microscopic parameters

$$\{b_4\rho\mathbf{p}_\perp, b_{11}\nabla(\nabla\cdot\mathbf{p}), b_{12}\nabla(\nabla\cdot\mathbf{p}), b_{18}\Delta^2\mathbf{p}, b_{19}\Delta^2\mathbf{p}\}.$$

Here, the term $b_4\rho\mathbf{p}_\perp$ ($b_4 < 0$) describes a density-dependent reduction of average particle rotations through collective effects, which (mildly) counteracts the term $b_3\mathbf{p}_\perp$ ($b_3 > 0$) related to single particle rotations. Such a contribution is most likely a consequence of the skewed distribution of single particle rotation rates $\Omega_i$ (Fig. S1): A larger amount of particles has rotation frequencies $\Omega_i < \Omega_0$, such that the average rotation frequency of the finite subsample of particles present in a given vortex (i.e. regions with high density) tends to be reduced as compared to the mean value $\Omega_0$ of the overall distribution. The terms $b_{11}\nabla(\nabla\cdot\mathbf{p})$ and $b_{12}\nabla(\nabla\cdot\mathbf{p})$ with $b_{11}, b_{12} > 0$ correspond to non-standard diffusive terms [16] that can generally exist in any flocking-type model and in our case arise through the combination of rotational diffusion and microscopic parameter variability. The terms $b_{18}\Delta^2\mathbf{p}$ and $b_{19}\Delta^2\mathbf{p}$ with $b_{18}, b_{19} > 0$ correspond to Swift-Hohenberg-type terms, which typically appear in systems that exhibit pattern formation on mesoscopic length scales [17]. In our case, the relevant length scales are determined by the microscopic parameters as $v_0/\Omega_0$, corresponding to the approximate radius of vortices, as well as by the finite interaction range among particles, which was also recently shown to give rise to Swift-Hohenberg-type operators on the mean-field level [18].

- Lastly, we identify a third group of non-linear terms

$$\{b_6|\mathbf{p}^2|\mathbf{p}_\perp, b_9(\mathbf{p}\cdot\nabla)\mathbf{p}_\perp, b_{10}(\mathbf{p}_\perp\cdot\nabla)\mathbf{p}, b_{15}\nabla|\mathbf{p}|^2\}$$

as those that are expected by the analytic coarse-graining of a system with constant microscopic parameters, but which do not appear in the most parsimonious model that recapitulates the dynamics (PDE 8 in Tabs. S I and S III). However, these terms will be consecutively added to the learned models, when looking for those with higher complexity (Tab. S IV), but their effect on the overall dynamics and pattern formation is minute (Fig. S11, mid- and bottom-row).

TABLE S III. Parameters $b_l$ of the nine sparsest PDEs for the polarization dynamics (main text Fig. 3$C$), learned from simulations of the microscopic system main text Eqs. (1). PDE 8 (◄) reproduces the characteristic vortex dynamics as in the microscopic simulations (main text Fig. 3$A$,$B$,$E$) and the coefficients of the linear terms compare well with analytic coarse-graining predictions (Tab. I). Four additional PDEs with more terms are shown in Tab. S IV.

| Term | PDE 1 | PDE 2 | PDE 3 | PDE 4 | PDE 5 | PDE 6 | PDE 7 | PDE 8◄ | PDE 9 |
|---|---|---|---|---|---|---|---|---|---|
| $b_1\mathbf{p}$ | -0.009 | -0.009 | -0.009 | -0.009 | -0.009 | -0.009 | -0.009 | -0.009 | -0.009 |
| $b_2\rho\mathbf{p}$ | – | – | – | 0.013 | 0.013 | 0.013 | 0.007 | 0.009 | 0.009 |
| $b_3\mathbf{p}_\perp$ | 0.414 | 0.476 | 0.477 | 0.428 | 0.478 | 0.436 | 0.436 | 0.440 | 0.441 |
| $b_4\rho\mathbf{p}_\perp$ | – | -0.050 | -0.040 | – | -0.040 | -0.006 | -0.006 | -0.010 | -0.012 |
| $b_5|\mathbf{p}|^2\mathbf{p}$ | – | – | – | – | – | – | – | -0.080 | -0.080 |
| $b_6|\mathbf{p}|^2\mathbf{p}_\perp$ | – | – | – | – | – | – | – | – | 0.054 |
| $b_7\nabla\rho$ | -0.638 | -0.637 | -0.600 | -0.595 | -0.601 | -0.596 | -0.596 | -0.595 | -0.595 |
| $b_8(\mathbf{p}\cdot\nabla)\mathbf{p}$ | – | – | – | -0.536 | – | -0.510 | -0.510 | -0.463 | -0.479 |
| $b_9(\mathbf{p}\cdot\nabla)\mathbf{p}_\perp$ | – | – | – | – | – | – | – | – | – |
| $b_{10}(\mathbf{p}_\perp\cdot\nabla)\mathbf{p}$ | – | – | – | – | – | – | – | – | – |
| $b_{11}\nabla(\nabla\cdot\mathbf{p})$ | – | – | – | – | – | – | – | 0.078 | 0.077 |
| $b_{12}\nabla(\nabla\cdot\mathbf{p}_\perp)$ | – | – | 0.225 | 0.265 | 0.248 | 0.265 | 0.270 | 0.277 | 0.277 |
| $b_{13}\Delta\mathbf{p}$ | – | – | – | – | – | – | -0.151 | -0.155 | -0.156 |
| $b_{14}\Delta\mathbf{p}_\perp$ | – | – | 0.252 | 0.202 | 0.222 | 0.203 | 0.198 | 0.196 | 0.197 |
| $b_{15}\nabla|\mathbf{p}|^2$ | – | – | – | – | – | – | – | – | – |
| $b_{16}(\nabla\cdot\mathbf{p})\mathbf{p}$ | – | – | – | – | – | – | – | -0.225 | -0.213 |
| $b_{17}(\nabla\cdot\mathbf{p})\mathbf{p}_\perp$ | – | – | – | – | – | – | – | – | – |
| $b_{18}\Delta^2\mathbf{p}$ | – | – | – | – | – | – | -0.475 | -0.483 | -0.484 |
| $b_{19}\Delta^2\mathbf{p}_\perp$ | – | – | 1.100 | 1.197 | 1.085 | 1.212 | 1.215 | 1.235 | 1.243 |

TABLE S IV. Parameters $b_l$ of PDEs 10-13 for the polarization dynamics (main text Fig. 3$C$), learned from simulations of the microscopic system in main text Eqs. (1). These are learned in addition to the PDEs mentioned in Tab. S III.

| Term | PDE 10 | PDE 11 | PDE 12 | PDE 13 |
|---|---|---|---|---|
| $b_1\mathbf{p}$ | -0.009 | -0.009 | -0.009 | -0.009 |
| $b_2\rho\mathbf{p}$ | 0.010 | 0.010 | 0.010 | 0.010 |
| $b_3\mathbf{p}_\perp$ | 0.442 | 0.442 | 0.442 | 0.448 |
| $b_4\rho\mathbf{p}_\perp$ | -0.012 | -0.012 | -0.012 | -0.017 |
| $b_5|\mathbf{p}|^2\mathbf{p}$ | -0.079 | -0.060 | -0.066 | -0.065 |
| $b_6|\mathbf{p}|^2\mathbf{p}_\perp$ | 0.055 | 0.055 | 0.055 | 0.065 |
| $b_7\nabla\rho$ | -0.595 | -0.595 | -0.595 | -0.594 |
| $b_8(\mathbf{p}\cdot\nabla)\mathbf{p}$ | -0.479 | -0.479 | -0.480 | -0.461 |
| $b_9(\mathbf{p}\cdot\nabla)\mathbf{p}_\perp$ | – | 0.057 | 0.058 | 0.058 |
| $b_{10}(\mathbf{p}_\perp\cdot\nabla)\mathbf{p}$ | – | – | 0.054 | 0.054 |
| $b_{11}\nabla(\nabla\cdot\mathbf{p})$ | 0.078 | 0.078 | 0.076 | 0.076 |
| $b_{12}\nabla(\nabla\cdot\mathbf{p}_\perp)$ | 0.277 | 0.277 | 0.277 | 0.278 |
| $b_{13}\Delta\mathbf{p}$ | -0.153 | -0.150 | -0.138 | -0.138 |
| $b_{14}\Delta\mathbf{p}_\perp$ | 0.197 | 0.197 | 0.197 | 0.195 |
| $b_{15}\nabla|\mathbf{p}|^2$ | – | – | – | -0.023 |
| $b_{16}(\nabla\cdot\mathbf{p})\mathbf{p}$ | -0.213 | -0.215 | -0.218 | -0.202 |
| $b_{17}(\nabla\cdot\mathbf{p})\mathbf{p}_\perp$ | -0.117 | -0.151 | -0.171 | -0.171 |
| $b_{18}\Delta^2\mathbf{p}$ | -0.489 | -0.454 | -0.403 | -0.403 |
| $b_{19}\Delta^2\mathbf{p}_\perp$ | 1.244 | 1.244 | 1.245 | 1.236 |

**Appendix C: Parameters and parameter robustness of learned models**

The parameters of the PDEs learned from simulations of the active polar particle model in main text Eqs. (1) are summarized in Tab. S II (density dynamics) and Tab. S III (polarization dynamics). For the experimental Quincke roller system [15], the learned hydrodynamic model parameters are given in Tab. S V (density dynamics) and Tab. S VI (velocity dynamics).

The robustness of the sparse regression through STLSQ (SI Sec. A 3) is demonstrated in Fig. S6 for the identification of the polarization dynamics [main text Eq. (4b)]. As specified in the stability selection procedure (SI Sec. A 3), selected terms have non-zero coefficients in at least 60 % of the subsamples. To further quantify the uncertainty in
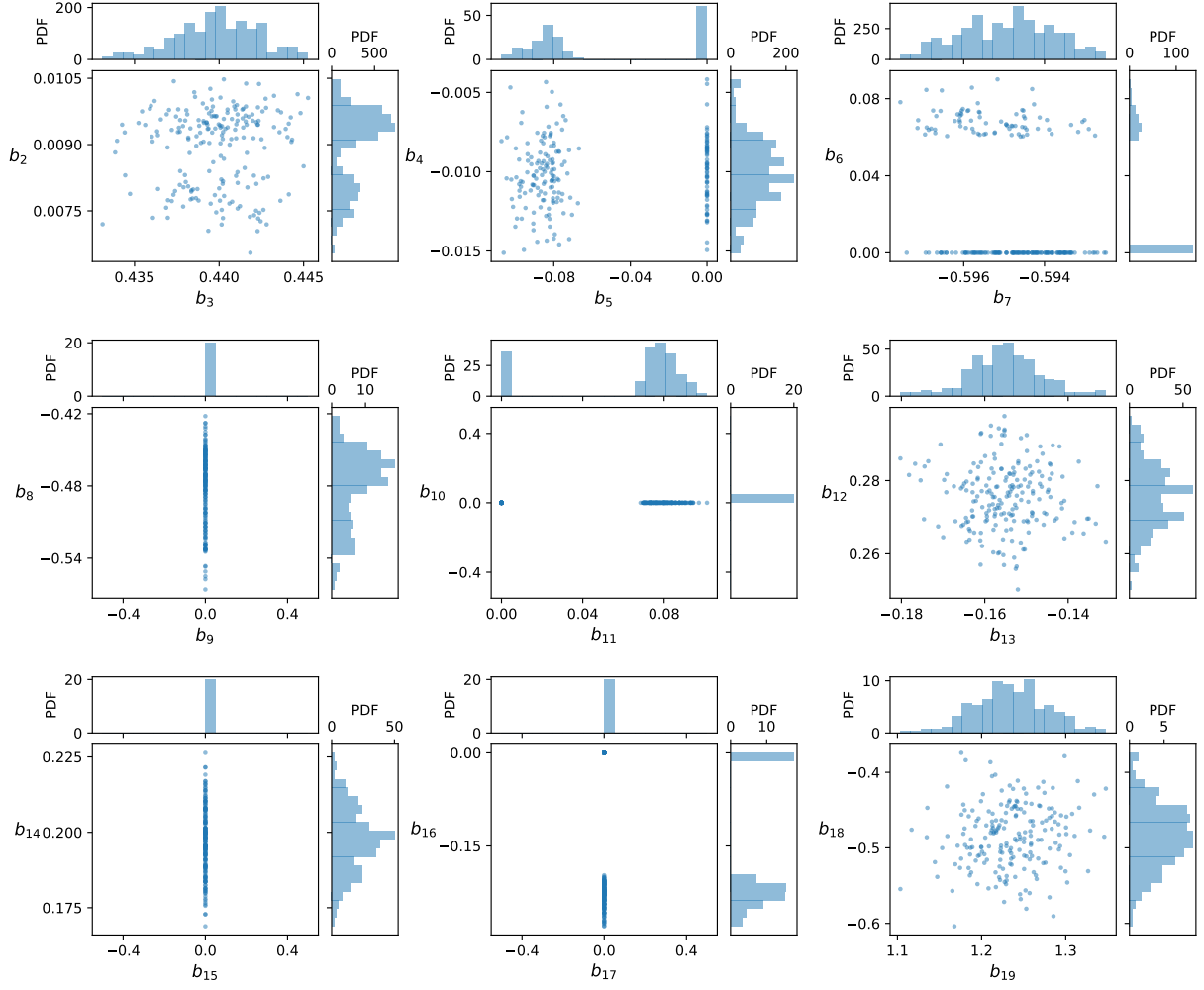
TABLE S V. Parameters $c_l$ of the PDE for the density dynamics (main text Fig. $4C$) learned from experimental data for self-propelled Quincke rollers (Supplementary Movie S2 of Ref. [15]). The dimensions of the coefficients are such that $[\mathbf{v}]$ =mm/s and $[\rho] = 1$, where the density $\rho$ represents the area fraction of rollers of diameter $D_c = 4.8\,\mu$m. The four sparsest PDEs are shown corresponding to the cut-off $n_0 \in \{50, 100\}$ above which the temporal Chebyshev modes in main text Eqs. (3) are set to zero to ignore high frequencies. The sparsest PDEs (◄) have coefficients close to each other and agree well with the mass conservation equation obtained from analytic coarse-graining (Ref. [15]).

| Term | Unit | $n_0 = 50$ | | | | $n_0 = 100$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | PDE 1◄ | PDE 2 | PDE 3 | PDE 4 | PDE 1◄ | PDE 2 | PDE 3 | PDE 4 |
| $c_1 \nabla \cdot \mathbf{v}$ | − | − | − | -0.052 | -0.052 | − | -0.051 | -0.051 | -0.055 |
| $c_2 \Delta\rho$ | mm$^2$ s$^{-1}$ | − | 0.016 | 0.040 | 0.023 | − | 0.055 | 0.040 | 0.041 |
| $c_3 \nabla \cdot (\rho\mathbf{v})$ | − | -0.950 | -0.950 | -1.068 | -1.067 | -0.945 | -1.057 | -1.054 | -0.985 |
| $c_4 \Delta\rho^2$ | mm$^2$ s$^{-1}$ | − | 0.047 | 0.080 | 0.081 | − | -0.076 | -0.051 | -0.062 |
| $c_5 \Delta|\mathbf{v}|^2$ | s | − | − | − | 0.001 | − | − | 0.001 | 0.001 |
| $c_6 \nabla \cdot (\rho^2\mathbf{v})$ | − | − | − | − | − | − | − | − | -0.313 |
| $c_7 \Delta\rho^3$ | mm$^2$ s$^{-1}$ | − | − | − | − | − | 0.427 | 0.341 | 0.366 |
| $c_8 \nabla \cdot (|\mathbf{v}|^2\mathbf{v})$ | mm$^{-2}$ s$^2$ | − | − | 0.035 | 0.035 | − | 0.034 | 0.034 | 0.034 |
| $c_9 \nabla \cdot (\rho\nabla|\mathbf{v}|^2)$ | s | − | − | − | -0.013 | − | − | -0.008 | -0.007 |
| $c_{10} \nabla \cdot (|\mathbf{v}|^2\nabla\rho)$ | s | − | -0.018 | -0.039 | -0.028 | − | -0.036 | -0.028 | -0.027 |

the values of the coefficients, we performed a bootstrapping procedure by performing least-squares regression only on the terms identified. This leads to empirical probability density functions whose standard deviations can be used to quantify uncertainties in the coefficients (Fig. S7). Similar results for the velocity dynamics for the Quincke roller system [Eq. (6b)] are presented in Figs. S8 and S9.

TABLE S VI. Parameters $d_l$ of the PDE for the velocity dynamics (main text Fig. $4C$) learned from experimental data for self-propelled Quincke rollers (Supplementary Movie S2 of Ref. [15]). The dimensions of the coefficients are such that $[\mathbf{v}]$ =mm/s and $[\rho] = 1$, where the density $\rho$ represents the area fraction of rollers of diameter $D_c = 4.8\,\mu$m. The four sparsest PDEs are shown corresponding to the cut-off $n_0 \in \{50, 100\}$ above which the temporal Chebyshev modes in main text Eqs. (3) are set to zero to ignore high frequencies. The sparsest PDEs which reproduce the experimental observations (◄) have coefficients that are close to each other for different values of $n_0$, and they agree well with corresponding values reported in Ref. [15] (Tab. II).

| Term | Unit | $n_0 = 50$ | | | | $n_0 = 100$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | PDE 1 | PDE 2◄ | PDE 3 | PDE 4 | PDE 1 | PDE 2◄ | PDE 3 | PDE 4 |
| $d_1 \mathbf{v}$ | s$^{-1}$ | − | 2.281 | 1.524 | 1.491 | − | 1.825 | 1.252 | 1.122 |
| $d_2 \rho\mathbf{v}$ | s$^{-1}$ | − | 8.356 | 5.156 | 4.745 | − | 6.135 | 3.143 | 3.083 |
| $d_3 |\mathbf{v}|^2\mathbf{v}$ | mm$^{-2}$ s | − | -2.194 | -1.436 | -1.382 | − | -1.710 | -1.095 | -0.999 |
| $d_4 \nabla\rho$ | mm$^2$ s$^{-2}$ | − | -1.620 | -1.711 | -2.074 | − | -1.689 | -2.438 | -2.430 |
| $d_5 (\mathbf{v} \cdot \nabla)\mathbf{v}$ | − | -0.639 | -0.674 | -0.678 | -0.679 | -0.662 | -0.696 | -0.702 | -0.702 |
| $d_6 \nabla(\nabla \cdot \mathbf{v})$ | mm$^2$ s$^{-1}$ | − | − | − | − | − | − | − | − |
| $d_7 \Delta\mathbf{v}$ | mm$^2$ s$^{-1}$ | − | − | − | − | − | − | − | 0.002 |
| $d_8 \nabla(|\mathbf{v}|^2)$ | − | − | − | − | 0.090 | − | − | 0.169 | 0.168 |
| $d_9 (\nabla \cdot \mathbf{v})\mathbf{v}$ | − | − | − | -0.189 | -0.190 | − | − | -0.178 | -0.179 |
| $d_{10} \Delta^2\mathbf{v}$ | mm$^4$ s$^{-1}$ | − | − | − | − | − | − | − | 0.000 |

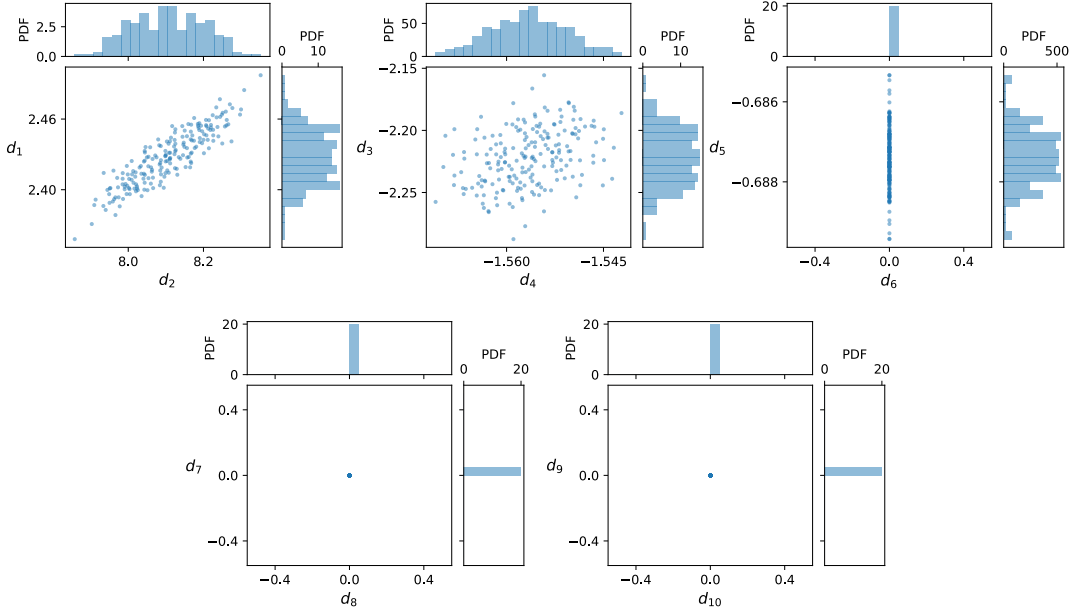FIG. S6. Parameters $b_l$ of the polarization dynamics (main text Eq. (4b) and Fig. 3$C$) obtained by applying the STLSQ algorithm on 200 data sub-samples with 50 % randomly selected data points (see SI Sec. A 3). The units of the coefficients are the same as in Tab. S III. This plot is generated for the thresholding parameter $\tau = 1.34 \times 10^{-4}$, at which PDE 8 (Tab. S III) is found in more than 60 % of the sub-samples. Histograms indicate the marginal probability density functions (PDFs) of the corresponding coefficients.
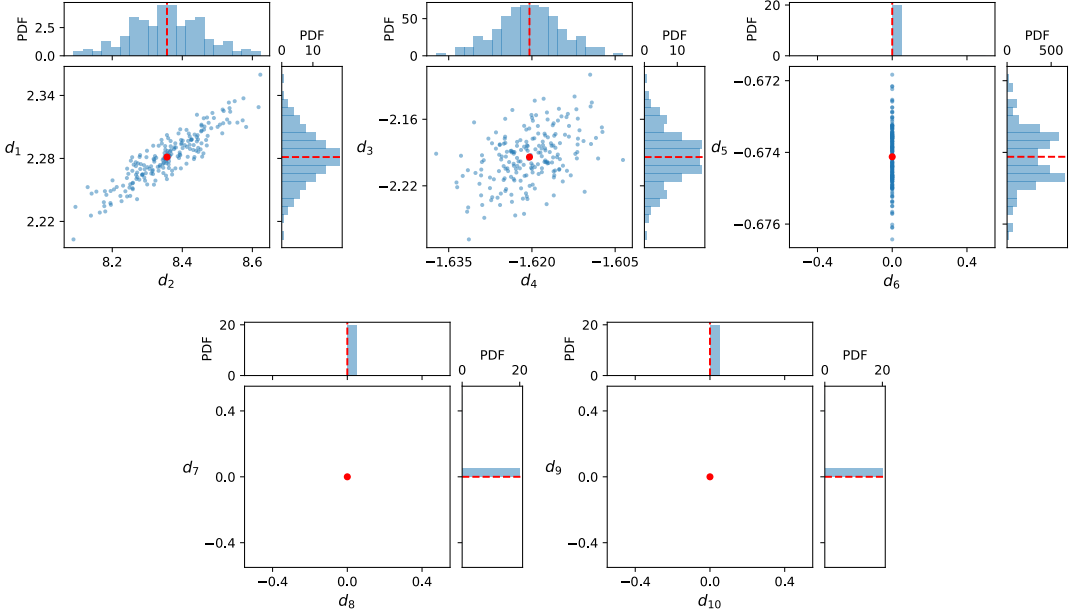
FIG. S7. Parameters $b_l$ of the polarization dynamics (main text Eq. (4b) and Fig. 3$C$) obtained from the least-squares method by setting the thresholded coefficients in Fig. S6 to zero. The units of the coefficients are the same as in Tab. S III. A total of 200 points (blue) are presented from 200 sub-samples with 50% randomly chosen data points. The red points indicate the fitted coefficients on the entire data set, which are the same as for PDE 8 in Tab. S III. The histograms indicate the marginal probability density functions (PDFs) of the corresponding coefficients, which can be used to find uncertainties for the coefficients. The fitted values and standard deviations of the coefficients are (mean $\pm$ standard deviation): $b_2 = 0.009 \pm 0.0004$, $b_3 = 0.440 \pm 0.0020$, $b_4 = -0.010 \pm 0.0016$, $b_5 = -0.080 \pm 0.0111$, $b_7 = -0.595 \pm 0.0009$, $b_8 = -0.463 \pm 0.0124$, $b_{11} = 0.078 \pm 0.0083$, $b_{12} = 0.277 \pm 0.0092$, $b_{13} = -0.155 \pm 0.0082$, $b_{14} = 0.196 \pm 0.0125$, $b_{16} = -0.225 \pm 0.0204$, $b_{18} = -0.483 \pm 0.0360$, $b_{19} = 1.235 \pm 0.0434$.

FIG. S8. Parameters $d_l$ of the velocity dynamics (main text Eq. (6b) and Fig. $4C$) obtained by applying the STLSQ algorithm on 200 data sub-samples with 50% randomly selected data points (see SI Sec. A 3). The units of the coefficients are the same as in Tab. S VI. This plot is generated for the thresholding parameter $\tau = 9.59 \times 10^{-2}$, at which PDE 2 (Tab. S VI and $n_0 = 50$) is found in more than 60% of the sub-samples. The histograms indicate the marginal probability density functions (PDFs) of the corresponding coefficients.



FIG. S9. Parameters $d_l$ of the velocity dynamics (main text Eq. (6b) and Fig. $4C$) obtained from the least-squares method by setting the thresholded coefficients in Fig. S8 to zero. The units of the coefficients are the same as in Tab. S VI. A total of 200 points (blue) are presented from 200 sub-samples with 50% randomly chosen data points. The red points indicate the fitted coefficients on the entire data set, which are the same as for PDE 2 in Tab. S VI. The histograms indicate the marginal probability density functions (PDFs) of the corresponding coefficients. The fitted values and the standard deviations of the coefficients are (mean $\pm$ standard deviation): $d_1 = 2.281 \pm 0.0242$, $d_2 = 8.356 \pm 0.1027$, $d_3 = -2.194 \pm 0.0239$, $d_4 = -1.620 \pm 0.0059$, $d_5 = -0.674 \pm 0.0008$.
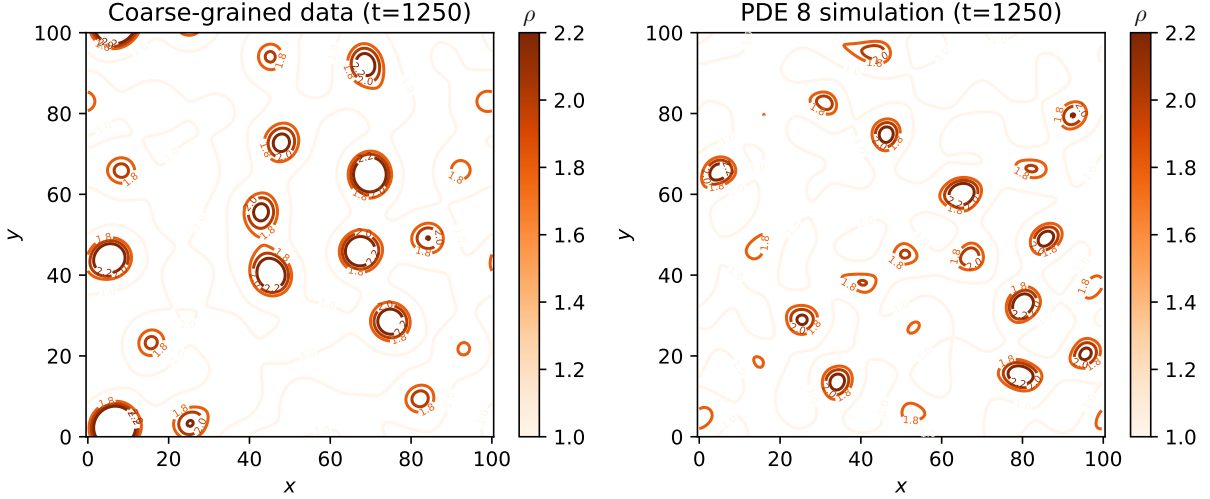
FIG. S10. Comparison of the contour plots of the density field $\rho(t, \mathbf{x})$ for the coarse-grained data (left) and the PDE simulation (right) corresponding to the chiral active Brownian model. Contour levels [1.0, 1.8, 2.0, 2.2] are the same in both plots. The contours distinctly isolate the most prominent vortices, which can be identified by the closely spaced red contours [1.8, 2.0, 2.2]. The representation shows that the number, density profiles, sizes, and disordered nature of vortices emerging in the learned model are very similar to the structures seen in the coarse-grained data.

## Appendix D: Quantitative comparison of the particle model data and continuum simulations

To facilitate a direct visual comparison between vortex patterns observed in the coarse-grained particle data and in simulations of the learned model (main text Fig. 3) the corresponding density fields are in Fig. S10 depicted as a level set representation. Key pattern characteristics, such as the number, density profiles, sizes, and disordered nature of emerging vortices, show very good agreement between patterns seen in the coarse-grained data and in the simulated model.

To further quantify this similarity between vortex patterns, we show in Fig. S11 the spatial power spectral density,

$$
S_{\rho;\mathbf{x}}(t, \mathbf{q}) = \mathcal{A}^{-1} \left| \int d^2\mathbf{x}\, \rho(t, \mathbf{x}) \exp(2\pi i \mathbf{q} \cdot \mathbf{x}) \right|^2 \tag{D1}
$$

of the long-lived vortex states at $t = 1250$ in the coarse-grained data and simulations. Here, $\mathcal{A}$ is the domain area. The simulation data correspond to PDE 1 for the density equation (Tab. S II), and PDEs 1-3 and 8-13 for the polarization equation (Tab. S III; PDEs 4-7 are found to be numerically unstable). The comparison shows that PDE 8 is the sparsest model that captures the dominant flow length scale indicated by the shape of the spectra as well as close agreement with the peak. As more terms are incorporated in the PDE (that is, the PDE index gets larger), the peak of the inferred spectrum moves closer to the maximum seen in the data.

In Fig. S12, we plot the temporal power spectral density

$$
S_{p_x;t}(\omega, \mathbf{x}) = T^{-1} \left| \int dt\, p_x(t, \mathbf{x}) \exp(2\pi i \omega t) \right|^2 \tag{D2}
$$

over a time window $[0, T]$ and averaged over a spatial window. The spectra for the coarse-grained data and simulation of PDE 8 for the momentum equation (Tab. S III) show similar overall shape with four distinct peaks at non-zero frequencies. The first set of peaks occur at $2\pi\omega = \pm\langle\Omega_i\rangle_p$ (the dotted lines), which is the average rotation frequency of the particles. This shows that the learned model reproduces the bulk temporal dynamics as seen in the input coarse-grained data.
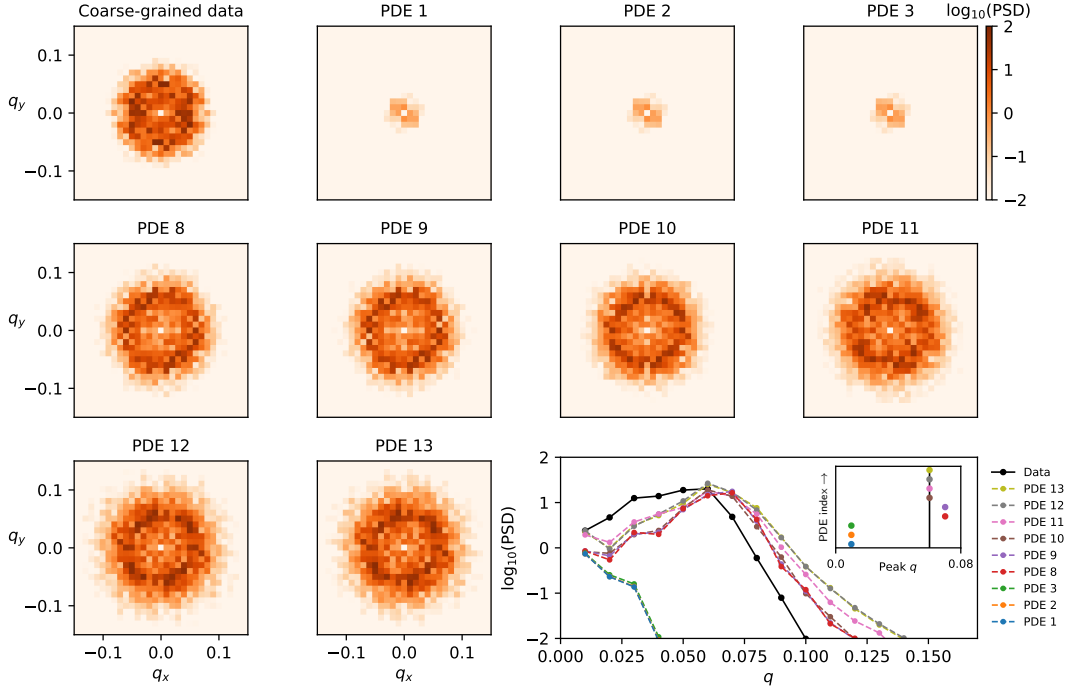
FIG. S11. Spatial power spectral density $S_{\rho;\mathbf{x}}(t,\mathbf{q}) = \mathcal{A}^{-1}\left|\int d^2\mathbf{x}\,\rho(t,\mathbf{x})\exp(2\pi i\mathbf{q}\cdot\mathbf{x})\right|^2$ averaged over 100 time points around $t = 1250$ between the coarse-grained data (top left) and simulations of the learned polarization PDEs (other color plots). Bottom right: azimuthal average of the power spectrum. For PDE 8 and above, the peak of the spectrum compares well with that of the data. Inset: wavenumber $q$ corresponding to the peak of the spectrum for each PDE. This peak gets close to the peak of the data (black line) as the PDE index gets larger, that is, the PDE incorporates more terms. To improve further the agreement in the tails of spectra, additional higher-order derivatives [19] must be accounted for in the library.
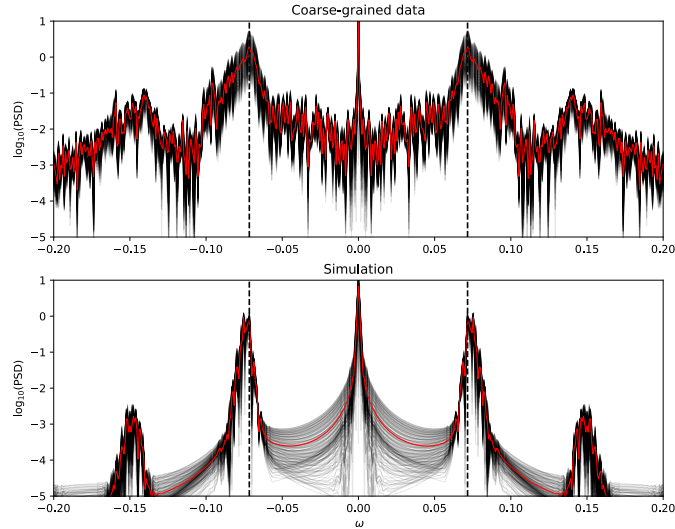


FIG. S12. Temporal power spectral density, $S_{p_x;t}(\omega,\mathbf{x}) = T^{-1}\left|\int dt\,p_x(t,\mathbf{x})\exp(2\pi i\omega t)\right|^2$ computed at grid points in the spatial window $x \in [45,50]$, $y \in [45,50]$ (black lines) for the time window $t \in [0,1700]$. The red lines show the average over the spatial locations. The non-zero peaks in the spectrum for the coarse-grained data (top) and the simulation of PDE 8 for the polarization equation (Tab. S III, bottom) compare well with $\omega = \langle\Omega_i\rangle_p/2\pi$ (black dotted lines). This implies that the simulation captures the bulk temporal dynamics in the input coarse-grained data at the average rotation frequency of the particles (see Fig. S1). The secondary peaks that result from nonlinearities also compare well between the simulations and the data. Note that the noise background in the coarse-grained particle data is several orders of magnitude smaller than the power in the characteristic peak frequencies.
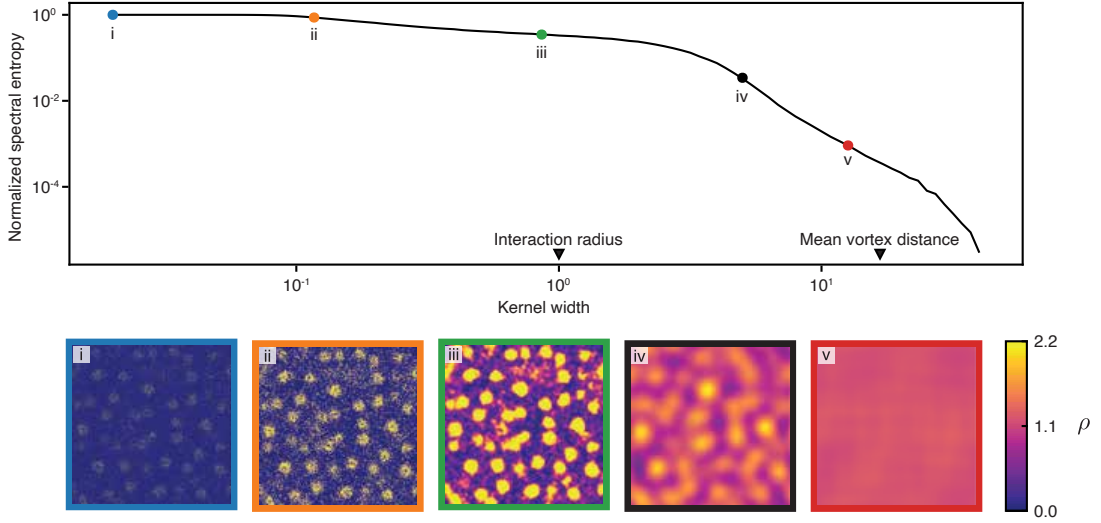
FIG. S13. Normalized spectral entropy, Eq. (E1), as a function of the Gaussian kernel width $\sigma$ (see SI Sec. A 2) for the chiral particle model data (top) quantifies the fraction of information that remains in Fourier space after coarse graining. Representative snapshots of coarse-grained fields are shown in the bottom panels. Characteristic scales in units of particle-particle interaction distance: Median vortex distance $\sim 17$ (obtained from a Delaunay triangulation of density peaks), box size 100. (i, $\sigma = 0.02$): Raw image before coarse graining. (i, $\sigma = 0.02$)–(ii, $\sigma = 0.12$): The discrete nature of the particle data remains present in the data, leading to little information loss. (ii, $\sigma = 0.12$)–(iii, $\sigma = 0.86$): As the coarse-graining scale approaches the interaction length scale, $\sigma \to 1$, coarse-grained data starts losing single particle information and vortices become more prominent. (iii, $\sigma = 0.86$)–(iv, $\sigma = 5$): Vortices start to be smoothed out as $\sigma$ exceeds the particle interaction distance and vortex size. Data from (iv, $\sigma = 5$) was used for inferring a continuum model from the chiral-particle simulation date; this choice of the coarse-graining scale ensures that the hydrodynamic fields are sufficiently smooth while still containing sufficient information about density fluctuations and vortex patterns. (v, $\sigma = 12.6$): When the kernel width $\sigma$ approaches the typical vortex-vortex distance, coarse-graining results in a constant homogeneous density and all spatially heterogeneous information is lost.

## Appendix E: Information content of coarse-grained data

To quantify the information loss due to coarse-graining as a function of the coarse-graining length scale $\sigma$, we use spectral entropy [20, 21] as a measure of the information content that remains in the coarse-grained fields. Specifically, we define the spectral entropy as

$$H(\sigma) = -\sum_{\mathbf{q}} \hat{S}_{\rho;\mathbf{x}}^{(\sigma)}(t, \mathbf{q}) \log_2 \hat{S}_{\rho;\mathbf{x}}^{(\sigma)}(t, \mathbf{q}), \tag{E1}$$

where the normalized spatial power spectral density $\hat{S}_{\rho;\mathbf{x}}^{(\sigma)}$ is defined as

$$\hat{S}_{\rho;\mathbf{x}}(t, \mathbf{q}) = S_{\rho;\mathbf{x}}(t, \mathbf{q}) \left( \int d^2 \mathbf{q} \, S_{\rho;\mathbf{x}}(t, \mathbf{q}) \right)^{-1}, \tag{E2}$$

with $S_{\rho;\mathbf{x}}(t, \mathbf{q})$ the spatial power spectral density defined in Eq. (D1). The additional index $\sigma$ indicates the Gaussian kernel smoothing width ('coarse-graining length scale') with which the underlying density field $\rho(t, \mathbf{x})$ was computed from the raw particle data. For our analysis, we rescale the spectral entropy $H$ given in Eq. (E1) by the spectral entropy of the raw particle data, yielding a normalized spectral entropy between 0 and 1 (see Figs. S13 and S14).

## Appendix F: Learning hydrodynamic equations for the collective motion of sunbleak fish

To demonstrate the generalizability of the presented learning framework to other active matter systems, we considered experimental data for the collective motion of sunbleak fish in Ref. [22]. These experiments were done in a quasi-two-dimensional tank in which the motion of $\sim 1000$ animals was influenced by a rotating dotted pattern that was projected on the bottom of the tank. This results in the fish swimming in an anti-clockwise motion (Fig. S15) that was recorded using overhead cameras.
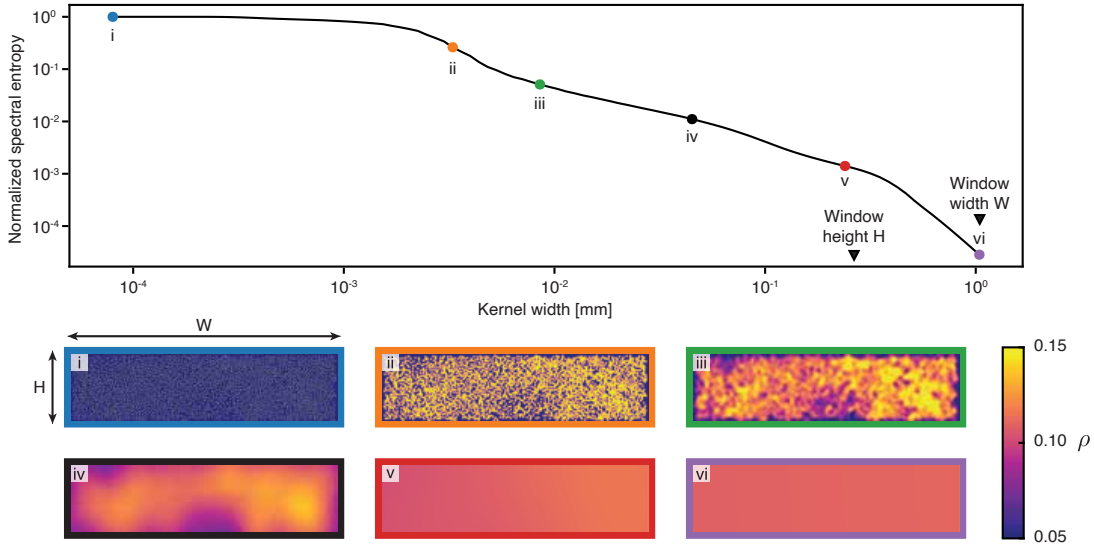
FIG. S14. Normalized spectral entropy, Eq. (E1), as a function of the Gaussian kernel width $\sigma$ (see SI Sec. A 2) for the Quincke roller system (top) quantifies the fraction of information that remains in Fourier space after coarse-graining. Representative snapshots of coarse-grained fields are shown in the bottom panels. Characteristic length scales: Roller diameter $4.8\,\mu$m, mean roller-roller centroid distance $\sim 11\mu$m, window height $H = 0.286$ mm, window width $W = 1.146$ mm. (i, $\sigma = 10^{-5}$ mm): Raw image before coarse graining. (ii, $\sigma = 0.0033$ mm)–(iii, $\sigma = 0.0085$mm): Single rollers are increasingly smoothed out, leading to an initial decrease in information. (iii, $\sigma = 0.0085$ mm)–(iv, $\sigma = 0.045$ mm): Large-scale density fluctuations become increasingly smoothed out by the coarse graining. Data from (iv, $\sigma = 0.045$ mm) was used for model learning from experimental Quincke roller data, providing a compromise between sufficiently smooth data and well-resolved details of density fluctuation in both spatial directions. (v, $\sigma = 0.24$ mm): As the coarse-graining scale $\sigma$ becomes comparable to the window height $H$, density fluctuations in the vertical direction have been smoothed out leading to an effectively one-dimensional density pattern that varies only along the horizontal direction. (vi, $\sigma = 1.04$ mm): As $\sigma$ becomes comparable to the window width $W$, all density variations disappear and the coarse-graining yields a constant homogeneous density.

## 1. Tracking of swimming trajectories

To track the fish motion between two consecutive video frames $I_n$ and $I_{n+1}$, we first detected feature points in $I_n$ using the `detectSURFFeatures` function in the MATLAB Computer Vision Toolbox, which implements the Speeded Up Robust Features (SURF) algorithm [23]. The parameters `MetricThreshold` (strongest feature threshold) and `NumScaleLevels` (number of scale levels per octave) in the SURF algorithm were set to be 1200 and 4 respectively. After getting the SURF features in $I_n$, we tracked all corresponding feature point pairs in the two frames $I_n$ and $I_{n+1}$ using the `vision.PointTracker` function in the MATLAB Computer Vision Toolbox, which implements the Kanade-Lucas-Tomasi (KLT) tracking algorithm [24, 25]. The parameter `MaxBidirectionalError` (forward-backward error threshold) in the KLT algorithm was set to be 1. To further remove the effect of outliers on the subsequent computation, we discarded all feature point pairs with displacement being outside of 2 standard deviations of the mean displacement.

## 2. Data coarse-graining and model learning

Starting from the particle data, we applied the same procedure as for the Quincke roller data (main text Sec. 2 and Fig. 4). We considered the particle positions $\mathbf{x}_i(t)$ and velocities $\mathbf{v}_i(t)$ (main text Fig. S15$A$) and applied kernel coarse-graining (main text Eqs (2); $\sigma = 0.3$ m) to obtain the density field $\rho$ and velocity field $\mathbf{v} = \mathbf{p}/\rho$ (Fig. S15$B$). These data were then projected onto Chebyshev basis functions in space and time, and a temporal mode cut-off $n > n_0$ was imposed to retain the hydrodynamically relevant time scales. Our goal was to learn equations of the form in Eq. (7) and similar physics-informed libraries as in Sec. 2 B (main text) were considered. The rotating pattern stimulus was accounted for by including an additional term $\mathbf{v}_\perp = \boldsymbol{\epsilon}^\top \cdot \mathbf{v} = (-v_y, v_x)^\top$ in the velocity equation. The complete libraries are shown in Fig. S15$C$. Application of the STLSQ algorithm along with the stability selection procedure (SI Sec. A 3) resulted in hydrodynamic models that are summarized in Fig. S15 and Tabs. S VII and S VIII. The sparsest density equation is $\partial_t \rho = e_3 \nabla \cdot (\rho \mathbf{v})$ with $e_3 \simeq -1$, which is the expected continuity equation for the

system. Along with this density equation, we simulated all the PDEs for the velocity equation starting from random initial conditions. These simulations were performed in a closed square domain with reflective boundary conditions to mimic the experimental tank with side walls (see SI Sec. A 6). We found that PDE 4 (marked by ◄ in Fig. S15) was the sparsest velocity equation that resulted in a spontaneously formed anti-clockwise vortex as seen in the coarse-grained data (Fig. S15). Furthermore, 100% of the subsamples at the corresponding thresholding parameter $\tau$ result in the same terms as in PDE 4. The uncertainties in the parameters of this PDE are quantified by a bootstrapping procedure in Fig. S16.

TABLE S VII. Phenomenological parameters $e_l$ for the density equation learned from the experimental data for the collective motion of sunbleak fish (Fig. S15). The dimensions of the coefficients are such that $[\mathbf{v}] = \mathrm{m/s}$ and $[\rho] = \mathrm{m}^{-2}$.

| Term | Unit | PDE 1◄ | PDE 2 | PDE 3 | PDE 4 |
|---|---|---|---|---|---|
| $e_1\nabla\cdot\mathbf{v}$ | m$^{-2}$ | – | – | – | – |
| $e_2\Delta\rho$ | m$^2$ s$^{-1}$ | – | $1.26\times10^{-3}$ | $1.69\times10^{-3}$ | $9.87\times10^{-4}$ |
| $e_3\nabla\cdot(\rho\mathbf{v})$ | – | $-1.02\times10^{0}$ | $-1.01\times10^{0}$ | $-1.13\times10^{0}$ | $-1.14\times10^{0}$ |
| $e_4\Delta\rho^2$ | m$^4$ s$^{-1}$ | – | $-1.57\times10^{-5}$ | $-1.81\times10^{-5}$ | $-1.66\times10^{-5}$ |
| $e_5\Delta|\mathbf{v}|^2$ | m$^{-2}$ s | – | – | $-1.68\times10^{1}$ | $-1.37\times10^{1}$ |
| $e_6\nabla\cdot(\rho^2\mathbf{v})$ | m$^2$ | – | – | $5.96\times10^{-4}$ | $6.20\times10^{-4}$ |
| $e_7\Delta\rho^3$ | m$^6$ s$^{-1}$ | – | $3.77\times10^{-8}$ | $4.10\times10^{-8}$ | $3.79\times10^{-8}$ |
| $e_8\nabla\cdot(|\mathbf{v}|^2\mathbf{v})$ | m$^{-4}$ s$^2$ | – | – | – | – |
| $e_9\nabla\cdot(\rho\nabla|\mathbf{v}|^2)$ | s | – | – | $1.62\times10^{-1}$ | $1.23\times10^{-1}$ |
| $e_{10}\nabla\cdot(|\mathbf{v}|^2\nabla\rho)$ | s | – | – | – | $4.91\times10^{-2}$ |

TABLE S VIII. Phenomenological parameters $f_l$ for the velocity equation learned from the experimental data for the collective motion of sunbleak fish (Fig. S15). Simulations the of velocity dynamics PDE 4 (◄) along with density dynamics PDE 1 in Tab. S VII show the spontaneous formation of an anti-clockwise vortex (Fig. S15$E$), recapitulating the pattern observed in the input data (Fig. S15$B$). The dimensions of the coefficients are such that $[\mathbf{v}] = \mathrm{m/s}$ and $[\rho] = \mathrm{m}^{-2}$.

| Term | Unit | PDE 1 | PDE 2 | PDE 3 | PDE 4◄ | PDE 5 |
|---|---|---|---|---|---|---|
| $f_1\mathbf{v}$ | s$^{-1}$ | – | – | $2.97\times10^{-2}$ | $3.80\times10^{-2}$ | $3.08\times10^{-2}$ |
| $f_2\rho\mathbf{v}$ | m$^2$ s$^{-1}$ | – | – | – | $-5.94\times10^{-5}$ | $-4.41\times10^{-5}$ |
| $f_3\mathbf{v}_\perp$ | s$^{-1}$ | $5.00\times10^{-2}$ | $7.49\times10^{-2}$ | $7.55\times10^{-2}$ | $7.56\times10^{-2}$ | $7.57\times10^{-2}$ |
| $f_4|\mathbf{v}|^2\mathbf{v}$ | m$^{-2}$ s | – | – | $-2.28\times10^{0}$ | $-2.40\times10^{0}$ | $-2.17\times10^{0}$ |
| $f_5\nabla\rho$ | m$^4$ s$^{-2}$ | $-3.26\times10^{-5}$ | $-3.98\times10^{-5}$ | $-3.96\times10^{-5}$ | $-3.95\times10^{-5}$ | $-3.94\times10^{-5}$ |
| $f_6(\mathbf{v}\cdot\nabla)\mathbf{v}$ | – | $-2.54\times10^{-1}$ | $-3.88\times10^{-1}$ | $-3.98\times10^{-1}$ | $-3.99\times10^{-1}$ | $-4.02\times10^{-1}$ |
| $f_7\nabla(\nabla\cdot\mathbf{v})$ | m$^2$ s$^{-1}$ | – | – | – | – | $2.60\times10^{-3}$ |
| $f_8\Delta\mathbf{v}$ | m$^2$ s$^{-1}$ | – | – | – | – | $-1.30\times10^{-3}$ |
| $f_9\nabla(|\mathbf{v}|^2)$ | – | – | $1.69\times10^{-1}$ | $1.64\times10^{-1}$ | $1.64\times10^{-1}$ | $1.66\times10^{-1}$ |
| $f_{10}(\nabla\cdot\mathbf{v})\mathbf{v}$ | – | – | – | – | – | – |
| $f_{11}\Delta^2\mathbf{v}$ | m$^4$ s$^{-1}$ | – | – | – | – | – |

**Appendix G: Temporal spectra of the coarse-grained data for the Quincke roller system and sunbleak fish**

The coarse-grained density $\rho(t,\mathbf{x})$ and velocity $\mathbf{v}(t,\mathbf{x})$ fields are projected onto Chebyshev basis functions in both space and time (see main text Sec. 2 A). To quantify how the energy decays with increasing frequency, we plot the summation of the squared mode amplitudes for each temporal mode in Fig. S17 to obtain a power spectrum similar to main text Fig. 2$B$. In all the three cases, we observe an exponential decay of the energy spectra with increasing frequency. This suggests that significant fluctuations ('noise') are absent in the coarse-grained field dynamics, typically visible in terms of slowly (algebraically) decaying temporal frequency spectra. Thus, our approach to model the dynamics using deterministic (instead of stochastic) partial differential equations is justified.

**A** Experimental data — Particle positions | Velocity component $v_x$ [m/s] | Velocity component $v_y$ [m/s]

**B** Coarse-grained fields — $\rho(x,y)$ [m$^{-2}$] | $v_x(x,y)$ [m/s] | $v_y(x,y)$ [m/s]

**C** Library

$$\partial_t \rho = e_1 \nabla \cdot \mathbf{v} + e_2 \Delta\rho + e_3 \nabla \cdot (\rho\mathbf{v}) + e_4 \Delta\rho^2 + e_5 \Delta|\mathbf{v}|^2 + e_6 \nabla \cdot (\rho^2\mathbf{v}) + e_7 \Delta\rho^3 + e_8 \nabla \cdot (|\mathbf{v}|^2\mathbf{v}) + e_9 \nabla \cdot (\rho\nabla|\mathbf{v}|^2) + e_{10}\nabla \cdot (|\mathbf{v}|^2\nabla\rho)$$

$$\partial_t \mathbf{v} = f_1\mathbf{v} + f_2\rho\mathbf{v} + f_3\mathbf{v}_\perp + f_4|\mathbf{v}|^2\mathbf{v} + f_5\nabla\rho + f_6(\mathbf{v}\cdot\nabla)\mathbf{v} + f_7\nabla(\nabla\cdot\mathbf{v}) + f_8\Delta\mathbf{v} + f_9\nabla|\mathbf{v}|^2 + f_{10}(\nabla\cdot\mathbf{v})\mathbf{v} + f_{11}\Delta^2\mathbf{v}$$

**D** Learned models — Coefficient $e_i$ (PDE 1, PDE 2, PDE 3, PDE 4) | Coefficient $f_i$ (PDE 1, PDE 2, PDE 3, PDE 4, PDE 5)

**E** Continuum simulation — $\rho(x,y)$ [m$^{-2}$] | $v_x(x,y)$ [m/s] | $v_y(x,y)$ [m/s]
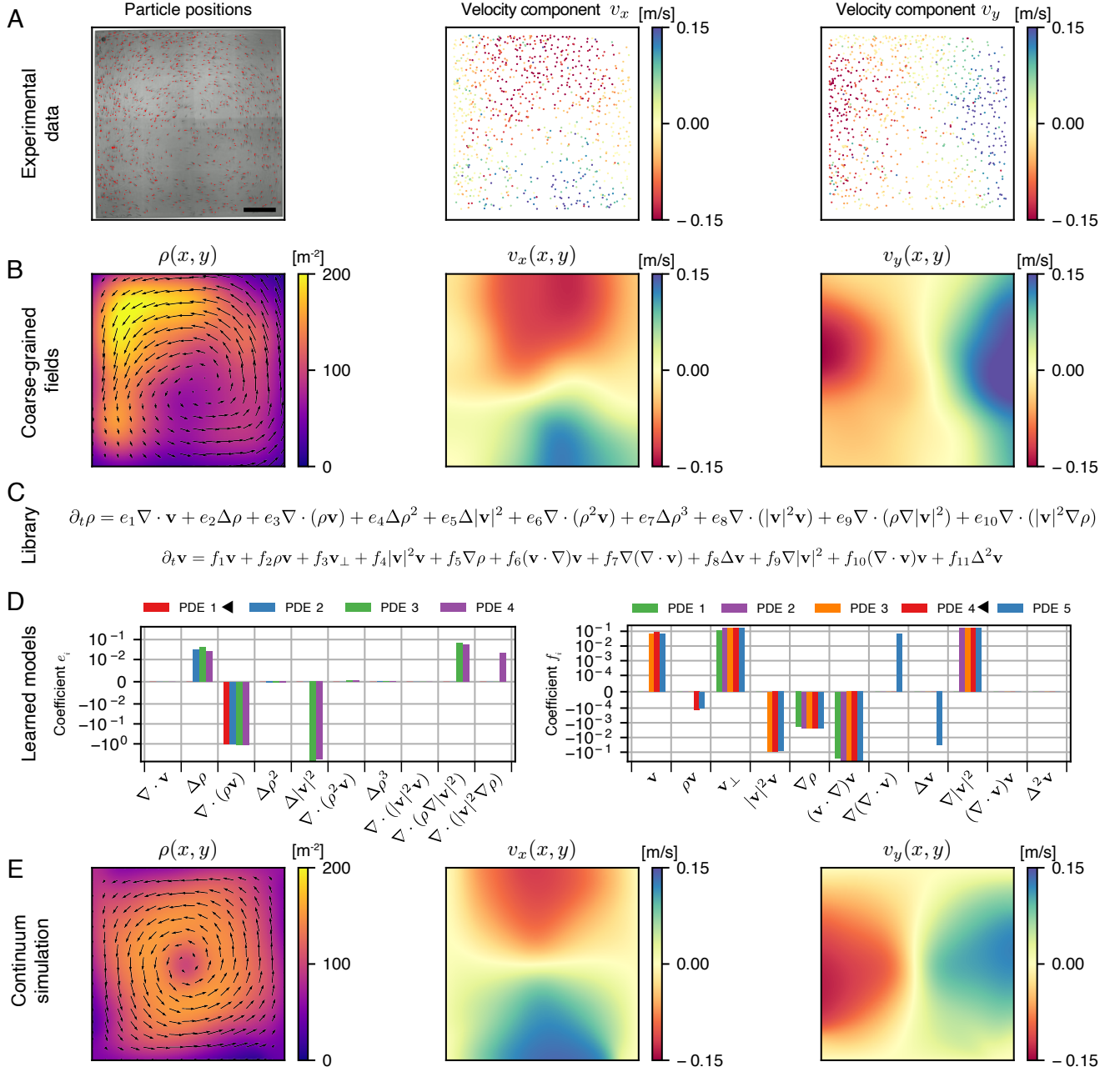
FIG. S15. Learning hydrodynamic equations for the collective motion of sunbleak fish. (A) Identified locations and velocities of approximately 1000 animals from experiments in a quasi-two-dimensional tank [22]. The fish are stimulated with a rotating pattern projected at the base of the tank, due to which they exhibit a counter-clockwise motion. Scale bar, 0.5 m. (B) Coarse-grained density $\rho(x,y)$ and velocity components, $v_x(x,y)$ and $v_y(x,y)$, at a representative time point. The coarse-graining width $\sigma$ is 0.3m. (C) Physics-informed libraries for the density and velocity dynamics. These are the same libraries as in main text Fig. 4C along with a $\mathbf{v}_\perp$ term in the velocity equation to take into account the external rotating stimulus provided to the fish by the rotating pattern. (D) Learned phenomenological coefficients $e_l$ and $f_l$ of the four sparsest PDEs for the density (left) and velocity (right) dynamics. The coefficients are non-dimensionalized with length scale $\sigma$ and time scale $\sigma/v_0$, where $v_0 = 0.13$ m/s is the average speed of the fish. (E) Simulation snapshot at $t = 500$ s of the learned hydrodynamic model (PDEs marked by ◄ in (D)) in a square domain with reflective boundary conditions. Starting from random initial conditions, this is the sparsest model for which spontaneous flow emerges and the flow settles into an anti-clockwise vortex as seen in the coarse-grained data. Furthermore, the magnitudes of the velocity components agree well with the coarse-grained data in (B).
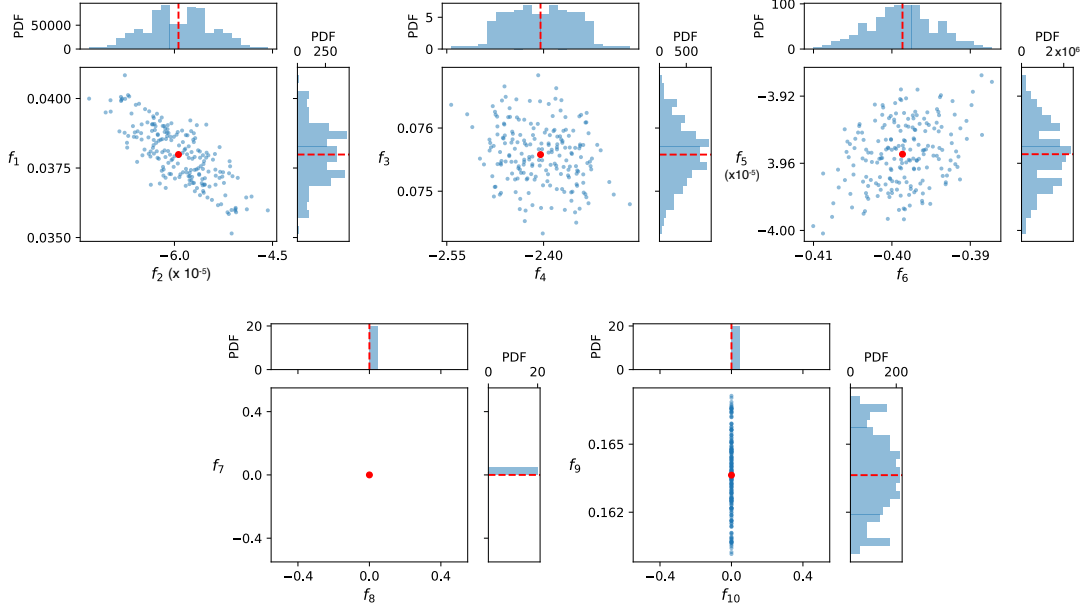
FIG. S16. Parameters $f_l$ of the velocity dynamics (Fig. S15$C$) that are obtained from the least-squares method by setting the thresholded coefficients in PDE 4 (Tab. S VIII) to zero. The units of the coefficients are the same as in Tab. S VIII. A total of 200 points (blue) are presented from 200 sub-samples with 50% randomly chosen data points. The red points indicate the fitted coefficients on the entire data set, which are the same as for PDE 4 in Tab. S VIII. The histograms indicate the marginal probability density functions (PDFs) of the corresponding coefficients. The coefficient $f_{11}$ (not shown) is thresholded out and set to zero. The fitted values and the standard deviations of the coefficients that remain in the PDE are (mean $\pm$ standard deviation): $f_1 = (3.80 \pm 0.10) \times 10^{-2}$, $f_2 = (-5.94 \pm 0.50) \times 10^{-5}$, $f_3 = (7.56 \pm 0.05) \times 10^{-2}$, $f_4 = -2.40 \pm 0.05$, $f_5 = (-3.95 \pm 0.02) \times 10^{-5}$, $f_6 = (-3.99 \pm 0.04) \times 10^{-1}$, $f_9 = (1.64 \pm 0.02) \times 10^{-1}$.
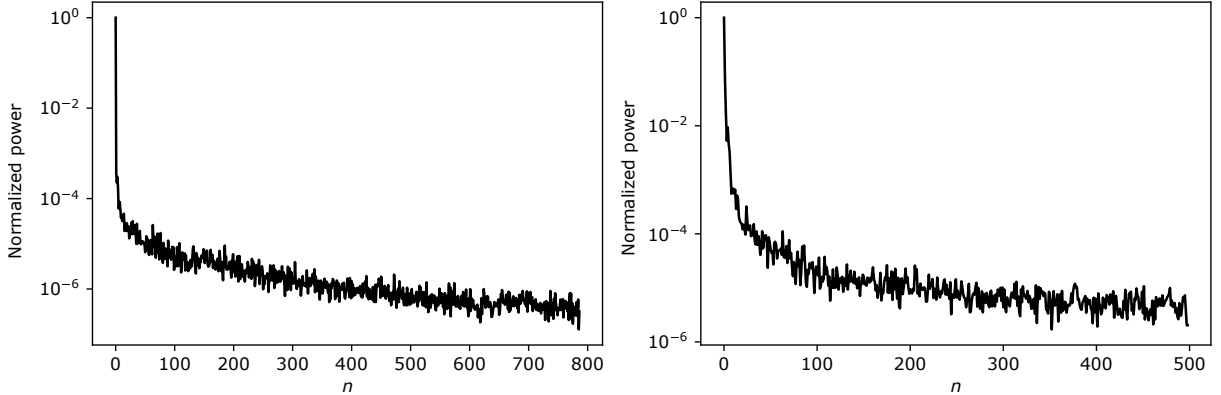


FIG. S17. Power spectra of coarse-grained data for the Quincke roller system (left; see Sec. 2 in the main text) and for sunbleak fish (right; see SI Sec. F), where $n$ denotes temporal Chebychev mode numbers. Similar to Fig. 2$B$ (main text), we define the spatio-temporal power spectrum $S_{x;n,(\alpha,\beta)} = |\mathbf{e}_x \cdot \hat{\mathbf{v}}_{n,\alpha,\beta}|^2$, where $(\alpha,\beta)$ are the spatial Chebyshev mode numbers and the total spatial spectral power is given by $S_{x;n} = \sum_{\alpha,\beta} S_{x;n,(\alpha,\beta)}$. The normalized power shown in the two panels is given by $S_{x;n}/S_{x;n=0}$. For both the Quincke roller and the sunbleak fish data spectral powers decay exponentially with increasing temporal mode number $n$. The analog spatio-temporal power spectra $S_{y;n,(\alpha,\beta)}$ show similar behavior. Note, the use of a Chebyshev basis implies that the power spectrum $S_{x;n,(\alpha,\beta)}$ is related to real-space data by $\sum_{\alpha,\beta} \sum_n S_{x;n,(\alpha,\beta)} = \int dt d^2\mathbf{x} |v_x(t,\mathbf{x})|^2 w_{\mathrm{Ch}}(t) w_{\mathrm{Ch}}(x) w_{\mathrm{Ch}}(y)$, where $w_{\mathrm{Ch}}$ is the Chebyshev weight function.

**Appendix H: Movie Legends**

**Movie 1.** Supporting Video for Figs. 2 and 3 of the main text: Comparison of discrete particle dynamics and coarse-grained density and polarization fields from simulations of an active chiral Brownian particle model (Eqs. (1), main text; see SI Sec. A 1 for simulation details and model parameters) with simulations of the learned continuum model (see SI Secs. C and D for details of the continuum model and its parameters).

**Movie 2.** Supporting Video for Fig. 4A-E of the main text: Simulations of the learned continuum model for the experimental Quincke roller system in a channel geometry with periodic boundary conditions along the horizontal direction.

**Movie 3.** Supporting Video for Fig. 4F of the main text: Simulation of the learned continuum model for the Quincke roller system in a closed-square geometry recapitulates essential characteristics of the experimental observations.

[1] B. Liebchen and D. Levis, Phys. Rev. Lett. **119**, 058002 (2017).

[2] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Proc. Natl. Acad. Sci. U.S.A. **113**, 3932 (2016).

[3] H. Schaeffer, Proc. R. Soc. A **473**, 20160446 (2017).

[4] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, Sci. Adv. **3**, e1602614 (2017).

[5] S. Maddu, B. L. Cheeseman, I. F. Sbalzarini, and C. L. Müller, Proc. R. Soc. Lond. **478**, 20210916 (2022).

[6] N. M. Mangan, J. N. Kutz, S. L. Brunton, and J. L. Proctor, Proc. R. Soc. A **473**, 20170009 (2017).

[7] M. Frigo and S. G. Johnson, Proc. IEEE **93**, 216 (2005).

[8] K. J. Burns, G. M. Vasil, J. S. Oishi, D. Lecoanet, and B. P. Brown, Phys. Rev. Res. **2**, 023068 (2020).

[9] D. S. Dean, J. Phys. A **29**, L613 (1996).

[10] E. Bertin, M. Droz, and G. Grégoire, J. Phys. A **42**, 445001 (2009).

[11] F. D. C. Farrell, M. C. Marchetti, D. Marenduzzo, and J. Tailleur, Phys. Rev. Lett. **108**, 248101 (2012).

[12] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha, Rev. Mod. Phys. **85**, 1143 (2013).

[13] E. Bertin, A. Baskaran, H. Chaté, and M. C. Marchetti, Phys. Rev. E **92**, 042141 (2015).

[14] B. Liebchen, M. E. Cates, and D. Marenduzzo, Soft Matter **12**, 7259 (2016).

[15] D. Geyer, A. Morin, and D. Bartolo, Nat. Mater. **17**, 789 (2018).

[16] J. Toner and Y. Tu, Phys. Rev. Lett. **75**, 4326 (1995).

[17] M. Cross and H. Greenside, *Pattern Formation and Dynamics in Nonequilibrium Systems* (Cambridge University Press, Cambridge, 2009).

[18] A. Mietke and J. Dunkel, Phys. Rev. X **12**, 011027 (2022).

[19] J. Słomka and J. Dunkel, Proc. Natl. Acad. Sci. U.S.A. **114**, 2119 (2017).

[20] A. Zhang, B. Yang, and L. Huang, in *2008 International Conference on BioMedical Engineering and Informatics*, Vol. 2 (2008) pp. 435–439.

[21] Y. N. Pan, J. Chen, and X. L. Li, Proc. Inst. Mech. Eng. C **223**, 1223 (2009).

[22] T. Walter and I. D. Couzin, eLife **10**, e64000 (2021).

[23] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, Comput. Vis. Image Underst. **110**, 346 (2008).

[24] B. D. Lucas and T. Kanade, in *Proceedings of the 7th International Joint Conference on Artificial Intelligence* (1981) pp. 674–679.

[25] C. Tomasi and T. Kanade, Carnegie Mellon University Technical Report CMU-CS-91-132 (1991).