



NOTE

Brownian-like deviation of neighboring cells in the early embryogenesis of the zebrafish

Juan Raphael Diaz Simões^{1,2,3} , Denis Grebenkov¹ , Paul Bourguine² and Nadine Peyri ras²¹ Condensed Matter Physics Laboratory, CNRS,  cole Polytechnique, Route de Saclay 91128 Palaiseau Cedex, France² BioEmergences Laboratory USR3695, CNRS, Universit  Paris-Saclay, 91198 Gif-sur-Yvette Cedex, France³ Author to whom any correspondence should be addressed.E-mail: juan-raphael.diaz-simoes@polytechnique.edu

Keywords: embryogenesis, migration, fluctuations, statistical physics

RECEIVED
2 July 2018REVISED
11 December 2018ACCEPTED FOR PUBLICATION
13 December 2018PUBLISHED
8 February 2019

Abstract

We investigate cell trajectories during zebrafish early embryogenesis based on 3D+time photonic microscopy imaging data. To remove the collective flow motion and focus on fluctuations, we analyze the deviations of pairs of neighboring cells. These deviations resemble Brownian motion and reveal different behaviors between pairs containing daughter cells generated by cell division and other pairs of neighboring cells. This observation justifies a common practice of using white noise fluctuations in modeling cell movement.

1. Introduction

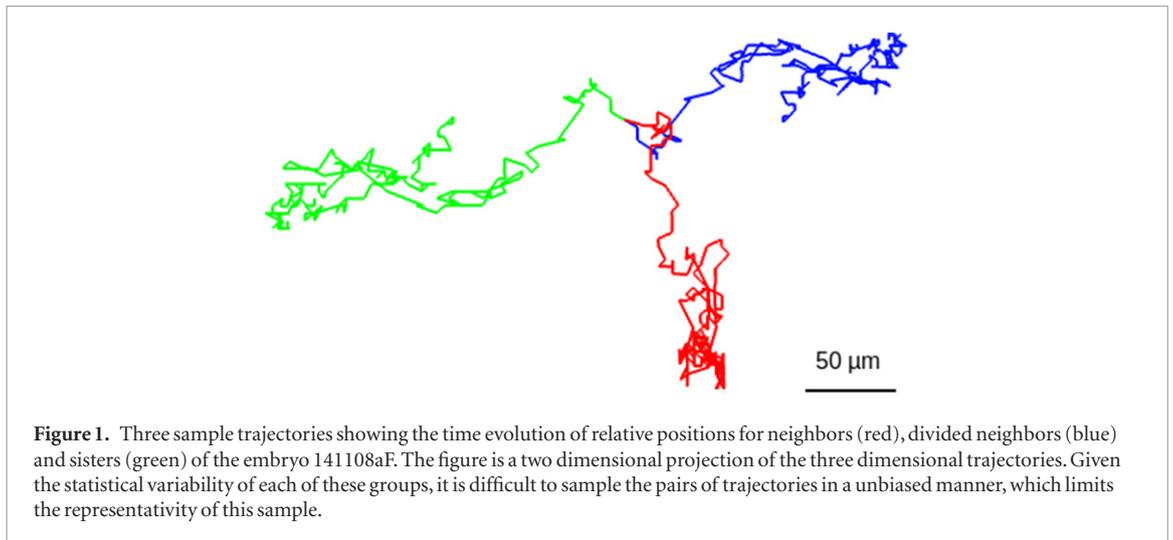
Embryogenesis is the process through which the cells of an embryo progressively organize themselves, forming functional compartments that carry out physiological activities (Fagotto 2014). The factors that contribute to this development are multiple, both genetic and epigenetic in nature (Holliday 1990). A number of models have been created aiming to describe the dynamics of groups of cells during this process (Stott *et al* 1999, Ouchi *et al* 2003, Zaman *et al* 2005, P rez and Prendergast 2007, Liedekerke *et al* 2015, Delile *et al* 2017). Many models include a stochastic component to account for *fluctuations* over a local fluid flow. These fluctuations are often modeled as a *white noise* for its conceptual and mathematical simplicity. White noise also allows one to for an internal dynamics of cells, that may contribute to autonomous, undirected displacement.

Recent progress in 3D+time imaging of embryonic development, combined to cell-tracking algorithms (Faure 2009, Amat *et al* 2014), made possible the identification of cell trajectories and genealogies in developing embryos. The zebrafish is particularly well suited for acquiring this type of data thanks to its external development and the transparency of its tissues. Other species that are studied using these methods are the sea-urchin, the rabbit and the ascidian (Faure *et al* 2016, Fabr ges *et al* 2018). The study of these trajectories allows us to corroborate the stochastic comp-

onent of the aforementioned models, provided that one manages to quantify trajectory fluctuations. These fluctuations are difficult to access without a precise model of the flow of cells during embryogenesis, which is yet not known. To overcome this problem, we use the fact that during the early embryogenesis of vertebrates, cells follow a flow that is continuous in space and neighboring cells have similar trajectories. This allows us to quantify trajectory fluctuations as the *deviations* of neighboring cells (i.e. the difference in their positions) and their evolution in time.

In addition, the statistics of relative positions at short times can be used for studying the behavior of cells after division (or mitosis). It is known that immediately after division the just divided cells are rounder and less attached to their neighbors than other cells (Cadart *et al* 2014). We hypothesize that this change impacts the tissue organization. It raises also the question of the recovery of adhesion after mitosis and whether this delay can be quantified, as pattern reintegration by the cells after mitosis may be of importance for proper development. We aim to quantitatively identify this difference in attachment by analyzing the statistics of relative positions. For this purpose, we study three groups of pairs of cells:

1. neighbors: pairs of cells that are within a reference distance to each other (the control group);



2. divided neighbors: pairs composed of one cell after division and another cell within a reference distance of the first one;
3. sisters: pairs of daughter cells just after division.

From this definition, we see that group 1 contains group 2, that contains group 3. The reference distance between neighbors is chosen to match to typical distances between sisters just after division. The study of these groups also allows us to determine possible differences in the global deviation of cells depending on whether they are sisters or not. In figure 1 we show trajectories consisting of relative positions of pairs of cells sampled from the three groups.

In the following we present the data, the measurements and the results.

2. 3D + time imaging data from developing embryos

The data consist of a spatio-temporal cell lineage algorithmically extracted from 3D+time images of four wild type zebrafish specimens that have been obtained by 2-photon laser scanning microscopy. Image acquisition and image processing workflow are described in detail in Faure *et al* (2016). Cell lineages have been reconstructed using the Morphotrack cell tracking algorithm, which is a non-parametric method based on the estimation-maximization algorithm (Faure 2009). The characteristics of each data set are summarized in table 1.

Each data set consists of 3D volumes taken at regular intervals between 6h30 and 15h15 hours post fertilization. Due to the large size of the zebrafish embryo, the images do not encompass the animal *in toto*, covering however a large portion of its body. As a consequence, cells can enter and quit the field of view of the microscope at any moment. Three snapshots in figure 2 illustrate one of the data set.

The spatio-temporal lineages have been extracted from 3D images by first detecting nucleus approximate center and linking them through time with a tracking

Table 1. Characteristics of the four studied zebrafish data sets. The study is limited to the time interval between 6h30 and 15h15 hours post fertilization, encompassing gastrulation stages and early organogenesis.

ID	Time step	Voxel size (μm)	Temperature ($^{\circ}\text{C}$)
141108a	2m25	1.37	24
141108aF	2m26	1.38	28
141121a	2m30	1.21	26
170315aF	1m40	1.51	26

algorithm (Faure *et al* 2016), which is reported to perform with 98% accuracy in terms of link detection between consecutive time steps. Each data set contains between 5000 and 10 000 cells per time step, the cell number increasing in time as a consequence of cell division. The whole dataset contains an approximate cumulated number of 100 000 cells.

The reference distance for the cell neighborhood calculations has been chosen as 20% more than the average distance between sister cells, corresponding to distances around 12 μm . Larger values for the reference distance would lead to faster relative displacements due to the global cell flow. Only one third of the total number of pairs of cells for each embryo have been included into the statistics for neighbors, to reduce the computation time. Tests using different samplings of cells showed that this quantity of cells is enough for the stabilization of averages, and increasing it does not change significantly the statistics.

Given these parameters, we can count the number of cells that coexist, i.e. that are both present in the data set, for a period of time larger than a certain number t of time steps, as shown for the embryo 141108aF in table 2. By definition, this number is decreasing, and the decrease appears to be approximately exponential.

3. Measurements

We define here the measurements used in the study of deviations. In the following $\langle A_{ij} \rangle_{ij}$ denotes the average of the quantity A over the set of pairs of cells denoted

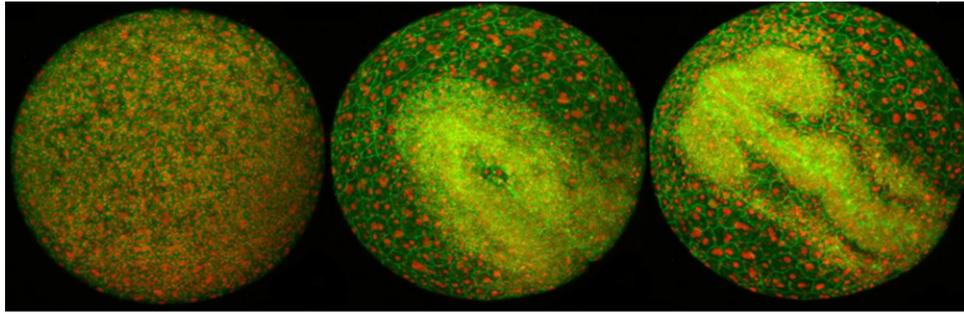


Figure 2. Three snapshots of the developing zebrafish embryo corresponding to the data set 141108aF, at 6h30 (left), 11h30 (middle) and 16h30 (right) post fertilization respectively. Animal pole view. 3D rendering of the raw data with nuclei stained by the expression of the fusion between a piece of the histone H2B and the fluorescent protein mCherry and membranes stained in green by the expression of a farnesylated form of the green fluorescent protein (eGFP). On the middle and right images one can see the result of the convergence and extension movements, showing that cells undertake overall displacements much larger than their size.

by ij . The error bars are calculated using the standard deviation of the quantity A over the same set.

We denote the position of cell i at time t by $X_i(t) \in \mathbb{R}^3$ and define the *relative position* of two cells i and j by

$$X_{ij}(t) = X_i(t) - X_j(t)$$

which is well defined for t in the interval $[0, t_{ij}]$ at which the two cells co-exist. In this context, time is always counted from the first moment when both cells are present in the data set. In other words, we do not consider the absolute time of development passed after fertilization, cells at the beginning and end of the observation are analyzed in the same way. Based on the relative position we define:

- the *mean squared relative displacement* (MSRD):

$$\delta_0(t) = \langle |X_{ij}(t) - X_{ij}(0)|^2 \rangle_{ij}$$

- the *time averaged mean squared relative displacement* (TAMSRD):

$$\delta_{av}(t) = \left\langle \frac{1}{t_{ij} - t + 1} \sum_{s=0}^{t_{ij}-t} |X_{ij}(s+t) - X_{ij}(s)|^2 \right\rangle_{ij}$$

We also define the *relative increment* $V_{ij} \in \mathbb{R}^3$ by:

$$V_{ij}(t) = \frac{X_{ij}(t + \Delta t) - X_{ij}(t)}{\Delta t}$$

where Δt is the time step (see table 1). From it we derive two statistics:

- the *relative increment autocorrelation*

$$\gamma_0(t) = \frac{\langle V_{ij}(t) \cdot V_{ij}(0) \rangle_{ij}}{\langle V_{ij}(0) \cdot V_{ij}(0) \rangle_{ij}}$$

- the *time averaged relative increment autocorrelation*

$$\gamma_{av}(t) = \frac{\left\langle \frac{1}{t_{ij}-t+1} \sum_{s=0}^{t_{ij}-t} V_{ij}(s+t) \cdot V_{ij}(s) \right\rangle_{ij}}{\left\langle \frac{1}{t_{ij}+1} \sum_{s=0}^{t_{ij}} V_{ij}(s) \cdot V_{ij}(s) \right\rangle_{ij}}$$

Table 2. Number of pairs of cells per group that coexist for a time larger than t (for the embryo 141108aF).

ID	Group	# Pairs at $t = 0$	# Pairs at $t = 100$
1	Neighbors	192 795	1725
2	Divided neighbors	156 236	1308
3	Sisters	15 395	122

As one can see from the definitions, the measurements $\delta_0(t)$ and $\gamma_0(t)$ take values relative to the initial point in time (when two cells start to coexist), while $\delta_{av}(t)$ and $\gamma_{av}(t)$ include an additional time average to improve the statistics and reduce noises. Besides better statistics, the comparison of the two forms of measurements allows one to control the change in dynamics over time. In particular, when a system is stationary one should see no difference between both calculations.

4. Results

4.1. Short time dynamics

We first study the short time dynamics by plotting the evolution in time of the distribution of relative increments (figure 3). The calculation has been made for all embryos and all three groups of cells, but since their distributions are similar visually, we show only that of group 1 (neighbors) of the embryo 141108aF. This plot is done using the coordinate x of the displacement vector, the result not being qualitatively different for other coordinates. As it can be seen, the curves are fairly symmetric and do not change much in time.

An important characteristic of this distribution is that its width is compatible with the size of cells. That means that in general, the relative displacement of cells between two consecutive time steps is local and occurs in the neighborhood of the pair of cells.

Since it is difficult to evaluate the differences between the three groups visually, we compare distributions by using the Kolmogorov–Smirnov test (Corder and Foreman 2014). This test returns a p -value for the compatibility of two given distributions

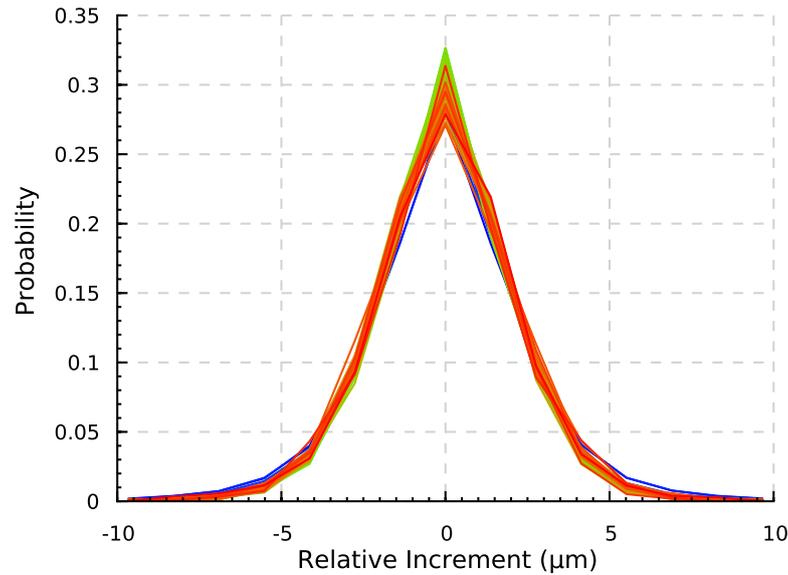


Figure 3. Distributions of the x coordinate of relative increments for pairs of group 1 (neighbors). The curves follow a color gradient from blue to red denoting the evolution of the distribution in time, for 100 time steps (4 h). The figure shows the relative stability of this distribution in time. Note the discrete character of the distribution, with a voxel size of $1.38 \mu\text{m}$.

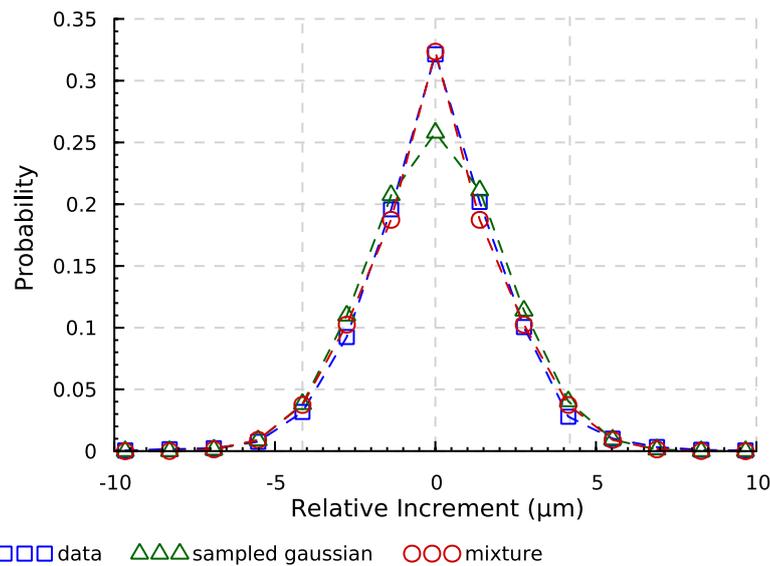


Figure 4. Distributions of the x coordinate of relative increments for pairs of neighbors at $t = 2 \text{ h}$, with its corresponding integer-sampled Gaussian distribution and the mixture model as defined in equation (2) for specimen 141108aF.

(Marsaglia *et al* 2003). For each pair of cell groups, we applied this test at each time step. The results were not conclusive neither for the temporal evolution of p -values nor the relation between the three groups, for none of the embryos. In other words, the p -values calculated both at short and long times, and for each pair of groups were neither significantly different nor stable enough in time to claim temporal or group differences.

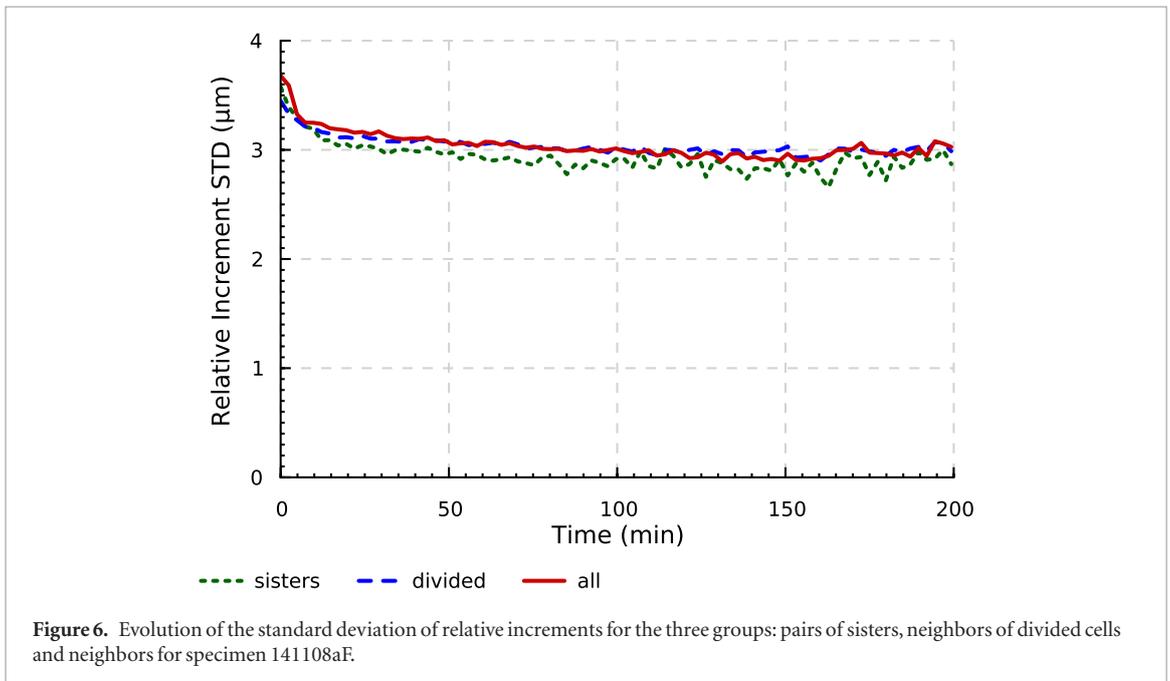
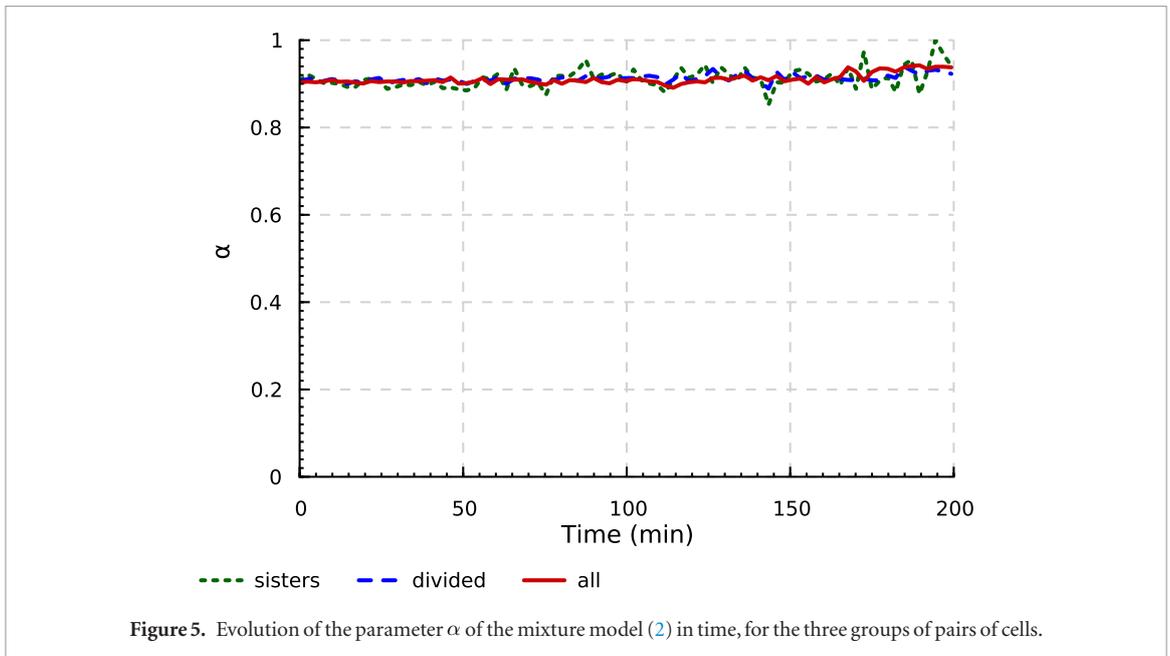
We aim now to identify the distribution of relative increments (figure 4). The first natural test is to compare this distribution with a Gaussian one whose mean and standard deviation have been calculated from the sample data. The Kolmogorov–Smirnov test distinguishes the two distributions with $p < 0.1$, which is

expected since the distribution of relative increments is discrete as a consequence of image sampling in voxels and their relatively large size, with 8 voxels covering all the distribution.

For this reason, we sample the Gaussian distribution accordingly. Given a Gaussian distribution \mathcal{N} , we define the integer-sampled Gaussian distribution $\hat{\mathcal{N}}$ by the probability distribution:

$$\mathbb{P}(\hat{\mathcal{N}} = n) = \mathbb{P}\left(-\frac{1}{2} \leq \mathcal{N} - n < \frac{1}{2}\right). \quad (1)$$

We compare the empirical distribution of relative increments with the integer-sampled Gaussian one whose mean and standard deviation parameters have



been calculated from the data. Even in this case, the Kolmogorov–Smirnov test distinguishes the two distributions for $p = 0.1$.

As one can see in figure 4, the empirical distribution and the integer-sampled Gaussian one differ around zero. Therefore, we introduce the mixture distribution:

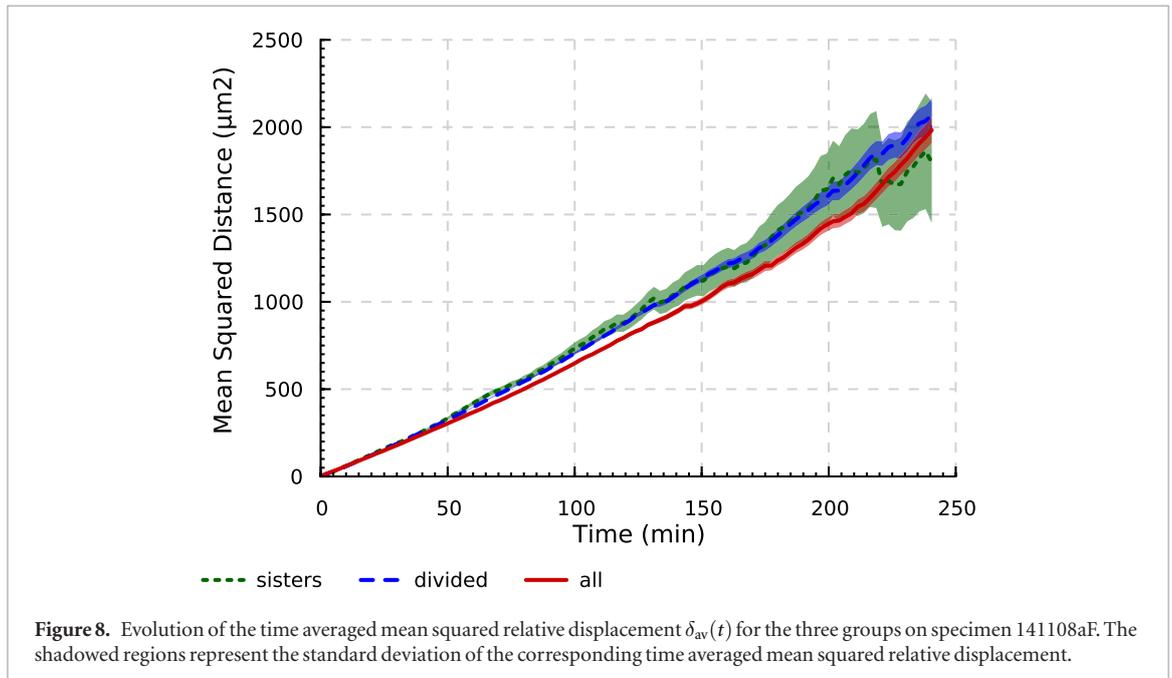
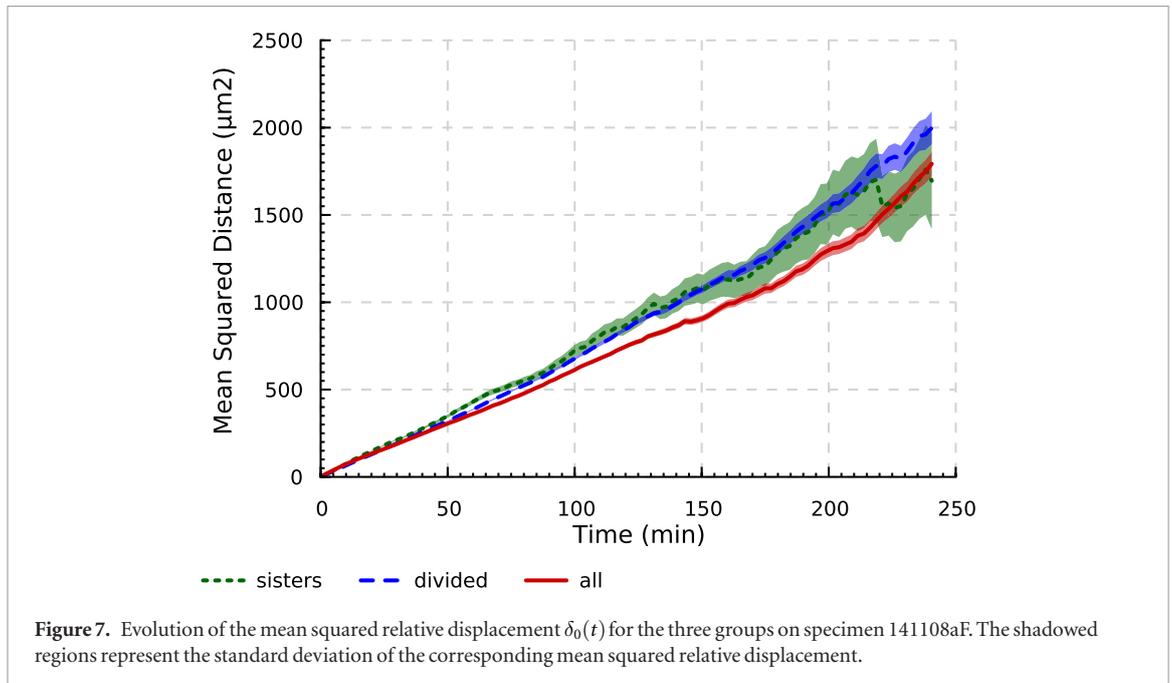
$$\mathcal{M} = \alpha \hat{\mathcal{N}} + (1 - \alpha) \delta_0 \quad (2)$$

where δ_0 is a point distribution on 0. This distribution provides an effective fit that can be interpreted as follows. There is a short-scale dynamics that cannot be accessed in our biological data, due to the spatial granularity of the images. This dynamics might be related to anomalous subdiffusive propagators (known to be peaked at 0) but without having enough

spatial resolution, we model it by a Dirac delta. In addition, there is a longer-scale dynamics which leads to a Gaussian distribution.

Using a maximum likelihood estimation, we fit the distribution of relative increments by equation (2) for all time steps and for all three groups of pairs of cells. The evolution of the parameter α is shown in figure 5. As one can see, the parameter is stable in time, fluctuating around 0.9, with no visible differences between the three groups. Note that the mixture distribution accurately fits the empirical one on figure 4.

The distribution of relative increments is centered around zero. We study now the typical size of relative increments, quantified by the standard deviation. Its time evolution is shown in figure 6. We can see that the standard deviation is slightly higher at first time



steps and decreases to a stable value around $3 \mu\text{m}$ for all three groups. Note that this standard deviation is smaller than the typical size of cells, which is around $10 \mu\text{m}$. In addition, there is no significant difference in the size of relative increments between the three groups.

4.2. Long time dynamics

We present now the evolution in time of the mean squared relative displacements δ_0 and δ_{av} for each group of cells in figures 7 and 8 for the embryo 141108aF, the result not being qualitatively different for the other specimens.

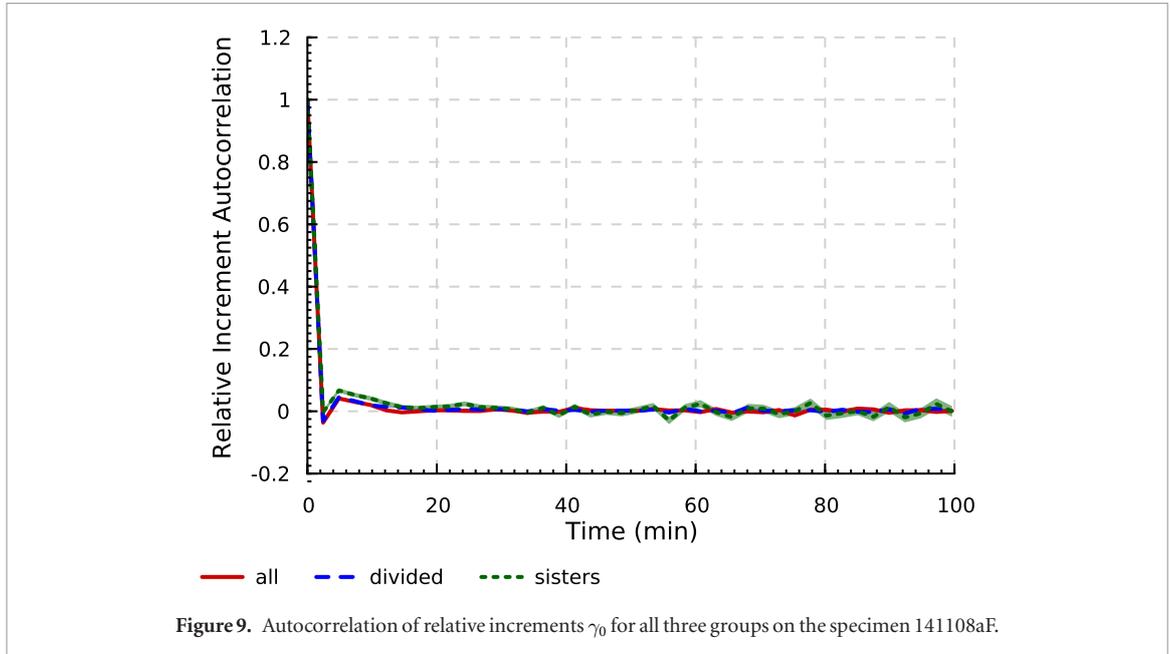
As one can see, all these curves grow almost linearly in time. The curves corresponding to group 1 (neighbors) have a slightly convex shape at the end, suggesting a faster diffusion. An interpretation of this obser-

vation is that as cells deviate from each other, effects of drift due to differences in their respective local flows become important. It is remarkable that the global linear behavior is observed at time scales up to 3 h, which shows a very low rate of divergence between neighboring cells in general. The curves corresponding to the two other groups (sisters and divided neighbors) are very similar in both cases.

Restricting the time interval to the first 200 min to avoid the high fluctuations observed in curves for group 3 (sisters), we calculate the diffusion coefficient by fitting each curve by $12Dt$ (the extra factor 2 appears due to considering the relative displacement of two cells). At 200 min we see a MSRD around $1600 \mu\text{m}^2$, corresponding to a relative distance of $40 \mu\text{m}$. This shows that even if the relative displacement between

Table 3. Diffusion coefficients ($\mu\text{m}^2/\text{min}$), calculated from the mean squared relative displacement δ_0 (MSRD) and time-averaged mean squared relative displacement δ_{av} (TAMSRD) for each group of pairs of cells.

Group Measurement	1		2		3	
	MSRD	TAMSRD	MSRD	TAMSRD	MSRD	TAMSRD
141108a	0.46	0.47	0.42	0.46	0.40	0.46
141108aF	0.56	0.66	0.59	0.67	0.50	0.60
141121a	0.45	0.48	0.40	0.45	0.37	0.43
170315aF	0.59	0.56	0.56	0.57	0.52	0.54



consecutive time steps is small, the relative displacements at long times are larger than the size of the cell, implying a change of neighborhood. An almost Gaussian character of relative displacements and the linear growth of the MSRD with time allow us to interpret fluctuations in the cells positions as (relative) diffusion. We recall that this diffusion is superimposed with the dominant coherent motion of cells.

As one can see in figure 7, the groups containing divided cells can be distinguished from the control group at time scales up to around 4 h using the mean squared relative displacement and its standard deviation. This gives a statistical significance to the difference in diffusion coefficients (see table 3), and supports the observation that the relative diffusion of cells after mitosis is slightly faster. While this observation does not depend on the specimen being studied, the diffusion coefficient values are different by a small margin between specimens. There are many possible reasons for this difference, the most probable one being the difference in temperature (table 1), since it is known that the zebrafish embryo develops faster at higher temperatures (Kimmel *et al* 1995), which can imply faster diffusion. Zebrafishes develop normally between 23 and 33 degrees celsius with a developmental speed twice higher at 33 degrees. However, we do not have enough data to establish a causal relation between these observation. Note also that the mean squared relative dis-

placement $\delta_0(t)$ and its time averaged version $\delta_{av}(t)$ yield close values of diffusion coefficients, except for the specimen 141108aF. This embryo was imaged at the highest temperature (see table 1) and has the largest deviations at later times, resulting in increased diffusion coefficients estimated from time averages.

We also present the autocorrelation of relative increments γ_0 (figure 9). For all three groups we see small negative correlation at the first time step, and very small positive one at the second time step. The time averaged version γ_{av} behaves very similarly (not shown) and thus evidences the stationarity of the increments of relative positions. We stress that this stationarity does not concern the whole process of embryogenesis.

5. Conclusion

We reported the statistical analysis of fluctuations of cell trajectories in the early embryogenesis of four wild type zebrafish specimens.

The deviations between neighboring cells are very similar to Brownian motion, and there is no visible drift effect at lag times up to 3 h. This means that, in general, neighboring cells stay close enough to each other and remain in a region that shares a common local flow. Therefore, one may consider the deviation between cells as a *fluctuation* around the flow, and this observation sup-

ports the stochastic component of the models of embryogenesis dynamics (Stott *et al* 1999, Ouchi *et al* 2003, Zaman *et al* 2005, Pérez and Prendergast 2007, Liedekerke *et al* 2015, Delile *et al* 2017). Furthermore, only small differences between the long-time behaviors of the three groups of cells are observed. Therefore, the common simplifying assumption of many models that differences in behavior of just divided cells can be neglected is justified. The characteristics that weakly deviate from the Brownian-like behavior are: (i) the relative increments are close to but not well fitted by a Gaussian distribution (figure 4), a correction term accounting for a short-time dynamics is proposed; (ii) the standard deviation of relative increments is slightly higher in the first time steps (figure 6); and (iii) relative increments are almost uncorrelated (figure 9).

The calculations took into account a large proportion of the data present in the microscope images. Therefore, the aggregated statistics gathers cells distributed both spatially and temporally within the explored developmental sequence. The study could be applied to selected cells populations to explore possible spatial and temporal variations. However, sampling the data manually and identify equivalent groups in multiple datasets, in addition to be time consuming, may be very inefficient. For this purpose, one could explore automated methods for the identification of cells with similar behavior (Diaz Simões *et al* 2017).

The data we used consisted of four different specimens of zebrafish. In the future, it would be important to apply the same methodology to other species in order to reveal if the relative displacement behavior observed in our study is universal or specific to each embryogenetical process. A potential challenge of such a study is the unequal suitability of tracking algorithms for different species.

Finally, we have been able to identify the distribution of relative increments as a mixture model composed of an integer-sampled Gaussian distribution and a point distribution around zero. The former component is related to the high discreteness of voxels, and the latter accounts for a short time dynamics for which the current spatio-temporal resolution of the microscopy is not sufficient. Finer details on the distribution can potentially be obtained by using images with a higher spatial resolution. We expect such data from recent developments in light sheet microscopy imaging (Reynaud *et al* 2014). Once combined, these measurements can provide valuable extra information for the conception and validation of more accurate models of embryonic development.

Acknowledgments

DG acknowledges the support under Grant No. ANR-13-JSV5-0006-01 of the French National Research Agency.

JRDS and NP acknowledge the support of the France BioImaging infrastructure ANR-10-INBS-04 and the Equipex Morphoscope2 ANR-11-EQPX-0029.

ORCID iDs

Juan Raphael Diaz Simões  <https://orcid.org/0000-0001-8321-0959>

Denis Grebenkov  <https://orcid.org/0000-0002-6273-9164>

References

- Amat F, Lemon W, Mossing D P, McDole K, Wan Y, Branson K, Myers E W and Keller P J 2014 Fast, accurate reconstruction of cell lineages from large-scale uorescence microscopy data *Nat. Methods* **11** 951–8
- Cadart C, Zlotek-Zlotkiewicz E, Berre M L, Piel M and Matthews H K 2014 Exploring the function of cell shape and size during mitosis *Dev. Cell* **29** 159–69
- Corder G W and Foreman D I 2014 *Nonparametric Statistics: a Step-by-Step Approach* (New York: Wiley)
- Delile J, Herrmann M, Peyri eras N and Doursat R 2017 A cell-based computational model of early embryogenesis coupling mechanical behaviour and gene regulation *Nat. Commun.* **8** 13929
- Diaz Sim oes J R, Bourguine P, Grebenkov D and Peyri eras N 2017 Cell trajectory clustering: towards the automated identification of morphogenetic fields in animal embryogenesis *Proc. of the 6th Int. Conf. on Pattern Recognition Applications and Methods (Porto, Portugal, 24–26 February 2017)* pp 746–52
- Fabr eges D, Daniel N, Duranthon V and Peyri eras N 2018 Control of the proportion of inner cells by asymmetric divisions and the ensuing resilience of cloned rabbit embryos *Development* **145** dev152041
- Fagotto F 2014 The cellular basis of tissue separation *Development* **141** 3303–18
- Faure E 2009 Reconstruction automatis ee du lignage cellulaire spatio-temporel de l'embryon *PhD Thesis* Ecole Polytechnique, Palaiseau, France
- Faure E *et al* 2016 A workflow to process 3D+time microscopy images of developing organisms and reconstruct their cell lineage *Nat. Commun.* **7** 8674
- Holliday R 1990 DNA methylation and epigenetic inheritance *Phil. Trans. R. Soc. B* **326** 329–38
- Kimmel C B, Ballard W W, Kimmel S R, Ullmann B and Schilling T F 1995 Stages of embryonic development of the zebrafish *Dev. Dyn.* **203** 253–310
- Liedekerke P V, Palm M M, Jagiella N and Drasdo D 2015 Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results *Comput. Part. Mech.* **2** 401–44
- Marsaglia G, Tsang W W and Wang J 2003 Evaluating Kolmogorov's distribution *J. Stat. Softw.* **8** 1–4
- Ouchi N B, Glazier J A, Rieu J-P, Upadhyaya A and Sawada Y 2003 Improving the realism of the cellular potts model in simulations of biological cells *Phys. A: Stat. Mech. Appl.* **329** 451–8
- P erez M and Prendergast P 2007 Random-walk models of cell dispersal included in mechanobiological simulations of tissue differentiation *J. Biomech.* **40** 2244–53
- Reynaud E G, Peychl J, Huisken J and Tomancak P 2014 Guide to light-sheet microscopy for adventurous biologists *Nat. Methods* **12** 30–4
- Stott E, Britton N, Glazier J and Zajac M 1999 Stochastic simulation of benign avascular tumour growth using the potts model *Math. Comput. Modelling* **30** 183–98
- Zaman M H, Kamm R D, Matsudaira P and Lauffenburger D A 2005 Computational model for cell migration in three-dimensional matrices *Biophys. J.* **89** 1389–97