

# Mathematical models of morphogenesis

Rui Dilão<sup>a</sup>

University of Lisbon, Instituto Superior Técnico, Department of Physics, Non-Linear Dynamics Group, Av. Rovisco Pais, 1049-001 Lisbon, Portugal

**Abstract.** Morphogenesis is the ensemble of phenomena that generates the form and shape of organisms. Organisms are classified according to some of its structural characteristics, to its metabolism and to its form. In particular, the empirical classification associated with the phylum concept is related with the form and shape of organisms. In the first part of this talk, we introduce the class of mathematical models associated the Turing approach to pattern formation. In the Turing approach, morphogenesis models are described by reaction-diffusion parabolic partial differential equations. Based on this formalism, we present a mathematical model describing the first two hours of development of the fruit fly *Drosophila*. In the second part of this talk, we present results on Pareto optimality to calibrate and validate mathematical models.

## 1. Introduction

It was Turing who suggested for the first time that, in spatially extended media or biological tissues, the diffusive coupling between reacting substances can generate stable patterns (Turing patterns), [14]. These patterns are characterised by a non-homogeneous and steady spatial distribution of the concentration of some substance. In extended media, the fluctuations in the concentration of a substance in solution are eliminated by the intrinsic random movement or Brownian motion of the molecules in the media. The Brownian motion homogenises the local concentrations, preventing the emergence of steady gradients. However, with several reacting substances in a media, the situation can be completely different. For example, two colliding molecules with binding affinity can lead to a third molecule, and the local concentration of the third molecule increases. Turing called the diffusion driven reacting substances morphogens or evocators, a term used by Waddington to explain the morphological transformations induced by specific substances in the embryo of the chick, [15]. In the following we describe a possible mechanisms of morphogenesis in *Drosophila* early development and some of the mathematical problems suggested by the Turing approach.

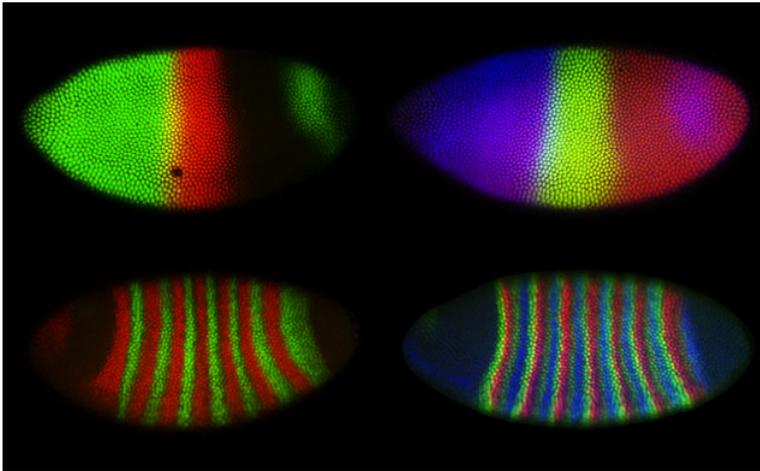
## 2. The formation of *Drosophila* segments

One of the characteristics of arthropods (insects, spiders, centipedes, shrimps, trilobites, etc.) is the formation, during the first hours of development, of protein segments along the embryo. These segments

---

<sup>a</sup> e-mail: [ruidilao@tecnico.ulisboa.pt](mailto:ruidilao@tecnico.ulisboa.pt)

This is an Open Access article distributed under the terms of the Creative Commons Attribution License 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Protein segments in the embryo of the fruit fly *Drosophila melanogaster*. (Data from the FlyEx database). On the two top figures, we see the segments of the gap-gene family of proteins. On the two bottom figures, we show the segments of the pair-rule and the segment-polarity families of proteins. These bands form during the first two hours after fertilisation of the *Drosophila* egg. In *Drosophila*, the pair-rule family shows 7 bands and, for example, in some centipedes this family develops 9 bands. The small dots in the figure show the localisation of nucleus and the different color are the markers of different proteins. At this phase of development, there are no cellular membranes and the nucleus of the embryo are not isolated from each other by the usual cellular membrane. This structure is called a syncytium. In animals, during development the first stages of development, there is also a syncytial phase.

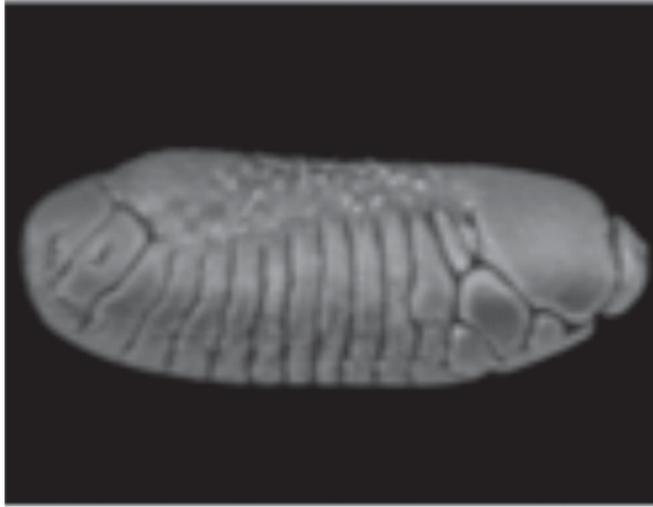
are bands formed by proteins that are expressed along the antero-posterior axis of the embryo. In Fig. 1, we show the band structure of some of the proteins of the embryo of *Drosophila*.

The number of protein segments changes from species to species and it is believed that this feature is inherited from some ancestral arthropod. The importance of these segments, formed during the first hours of development, is due to the fact that their positions determine the body plans of the adult larvae at a later stage of development (Fig. 2).

The biological events leading to the formation of the protein patterning in *Drosophila* are the following ([8] and [12]): In the *Drosophila* egg, some mRNAs of maternal origin (bicoid, nanos, caudal, hunchback, etc.) are placed near the poles of the oocyte by the mother's ovary cells, defining the antero-posterior axis of the embryo. For example, the regions of deposition of bicoid and nanos mRNAs determines the regions of the embryo where the head and the posterior organs of the larva will be developed latter. Initially, the oocyte has only one nucleus, but fertilisation triggers nuclear duplication by mitosis without the formation of cellular membranes (syncytium) (Fig. 1). The formed nucleus distribute near the internal wall of the embryo. Simultaneously, fertilisation triggers the translation of the deposited maternal mRNAs to proteins. These proteins are localised near the external nuclear membranes of the recently formed nucleus.

After fertilisation, the first 13 nuclear divisions occur without the organisation of cellular membranes around the nucleus (the embryo is a syncytium), giving rise to the syncytial blastoderm. The cytoplasmic membranes only become completely formed three hours after fertilisation, in the interphase following the 14th mitotic cycle, just before the onset of gastrulation.

During the syncytial stage, the transcribed zygotic genes are divided into three main families: gap, pair-rule and segment-polarity genes. The proteins resulting from their expression define broad segmentation patterns along the antero-posterior axis of the embryo, as shown in the two first pictures of Fig. 1. The proteins with origin in the maternal mRNAs form steady gradients along the



**Figure 2.** Segmentation of the *Drosophila* larvae. These segments begin to appear 7 hour after fertilisation and are completely formed 15 hours after fertilisation. Their positions correspond to the localisation of the expressed pair-rule proteins, expressed during the first two hours of development (Fig. 1). This suggests that the body plan of *Drosophila* is established very early during development.

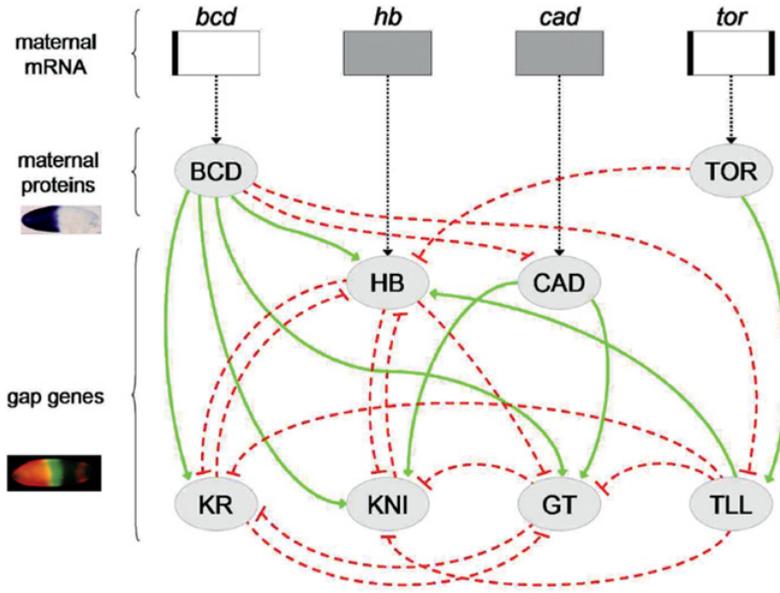
antero-posterior axis of the embryo. In the beginning of cleavage cycle 14, proteins of maternal origin act as transcription factors (activators or inhibitors) for gap-genes, pair-rule and segment polarity genes.

There are several models aiming to describe proteins steady gradients in *Drosophila* early development. Some models are based on the hypothesis of protein diffusion along the antero-posterior axis of the embryo, [2, 10]. Other models are based on the diffusion of mRNA of maternal origin, [3–7]. The protein diffusion hypothesis is sometimes justified by the absence of cellular membranes during the first 14 cleavage cycles of the embryo, and has been proposed by Nüsslein-Volhard and co-workers in the late eighties, [8]. However, this hypothesis implies several characteristics for the embryo and for the protein dynamics that have never been observed. The mRNA diffusion hypothesis is supported by the recent observation of the gradient of the bicoid mRNA of maternal origin, [13]. In order to decide which is the mechanism that best describe the segmental patterning as shown in Fig. 1, Dilão and co-workers introduced several mathematical models aiming to the describe the patterns of Fig. 1. These models are based on the hypothesis of diffusion of the maternal origin mRNA, together with localized protein production without diffusion, [3–7].

To test the pattern formation mechanisms in *Drosophila*, some details are necessary in the description of models. Optimally, these details should include measurements of protein and mRNA formation rates, degradation rates and diffusion coefficients. Also direct measurements of protein concentrations would be desirable. However, the values of these parameters change from embryo to embryo and it is not technically possible to measure simultaneously in the same embryo all these parameters.

Most of the biological information on patterning is in the form of the fluorescent images as shown in Fig. 1. The relationship between different proteins, mRNAs and genes are purely qualitative. For example, in Fig. 3, we show, the biological information of some of the known interaction between the proteins of maternal origin, the mRNAs and the gap-genes, in *Drosophila* early development.

In the double graph of Fig. 3, the green-solid arrows represent activation interactions and the red arrows inhibitory mechanisms. The symbols associated with the vertices of this double-graph represent a particular substance into one of its forms: gene, mRNA or protein. A particular formalism has been



**Figure 3.** Double graph of interactions describing the genetic network associated with the maternal and gap-gene phase of *Drosophila* early development. Green-solid represent activation of a gene by a protein and the red-dotted arrows show inhibitory interactions. Black arrows represent translation of a protein from the corresponding mRNA (Figure reprinted from [2]).

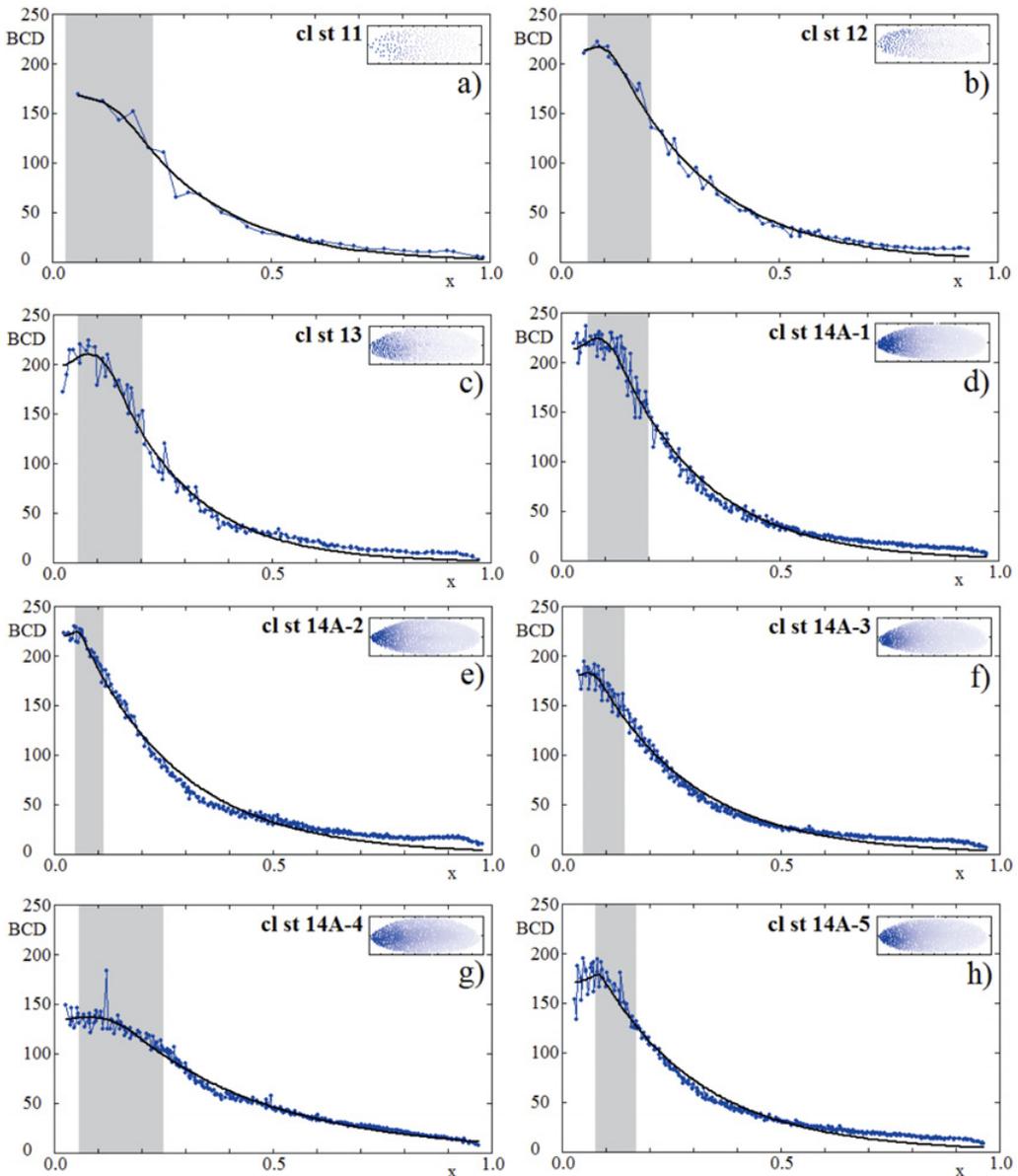
developed in order to transform graphs describing genetic regulatory networks into a mathematical model, [1] and [5].

For the particular case of the activation of Bicoid protein (BCD) by bicoid mRNA (*bcd*), represented by the sub-graph with two vertices at the top left of Fig. 3, we have assumed that bicoid mRNA diffuses along the antero-posterior axis of the embryo. The mathematical model associated to this sub-graph is described by the partial differential equation ([4, 7]),

$$\begin{aligned}\frac{\partial R}{\partial t} &= -dR + D\frac{\partial^2 R}{\partial x^2} \\ \frac{\partial B}{\partial t} &= aR\end{aligned}$$

where  $R(x, t)$  and  $B(x, t)$  are the concentration of *bcd* and BCD, respectively. These concentrations are defined along the antero-posterior axis of the embryo ( $x \in [0, L]$ ), and we assume zero-flux boundary conditions. The embryo length is  $L = 5$  mm. In the beginning of the developmental process there is the deposition of *bcd* at some (unknown) region of the antero-posterior axis of the embryo and this concentration of *bcd* ( $R(x, 0)$ ) is also unknown. This deposition region is the interval  $[\ell_1, \ell_2]$ , somewhere along the embryo with length  $L$ . In this model, the Bicoid protein does not diffuse, it is translated by the ribosomal machinery that exists outside the nuclear membranes. The diffusion of bicoid mRNA is driven by Brownian motion (molecular chaos) eventually along microtubules distributed along the embryo, [4].

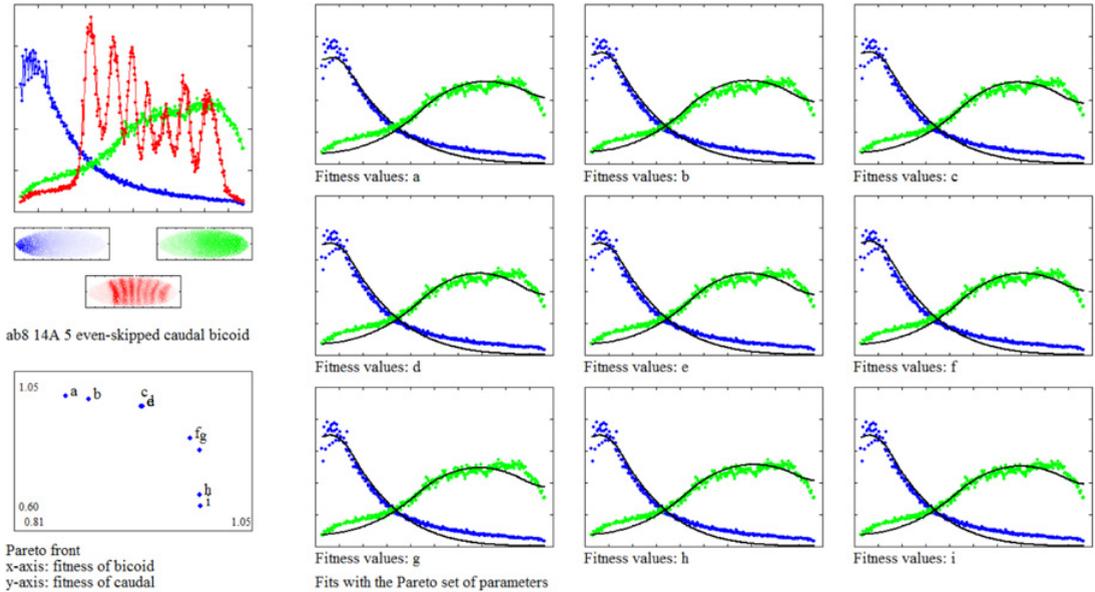
The model equations presented above with the boundary and initial conditions have seven free parameters, but theoretical analysis shows that it is only possible to determine four independent parameters, [4]. All the parameters are unknown and the experimental data is in the form of fluorescent signal proportional to protein concentration. The technique to calibrate the model equations with the



**Figure 4.** In blue, we show the experimental distribution of BCD (Bicoid protein) along the antero-posterior axis of the embryo of *Drosophila*, for different embryos in different stages of development. The small two-dimensional distribution within each frame is the two-dimensional gradient of protein BCD. The black line is the best steady solution of the model equations and the grey region is the region where *bcd* mRNA were initial deposited by the mother ovary cells. Both the grey regions and the model parameters were determined with an evolutionary genetic algorithm, [3, 4, 7]. The agreement between the diffusion model and the experimental data is very good, with relative errors in the range 5 – 8%. Figure reprinted from [4].

experimental data is to use evolutionary genetic algorithms (explained in the next section). If the fit is good, we believe that the model is sufficiently accurate to make predictions about the system.

In Fig. 4, we show the fit of the model solutions with the experimental data. The model predictions are denoted by the black lines and are the steady-state solution of the partial differential equation shown



**Figure 5.** In the picture on the top-left we show the concentration of the proteins Bicoid (blue), Caudal (green) and Even-skipped (red) along the antero-posterior axis of the embryo of *Drosophila*. On the nine figures on right we show Pareto bi-objective optimal solutions of the model equations for the same experimental data, [3, 11]. Each picture corresponds to a different set of parameter values on the Pareto front of this bi-objective optimisation problem. In the lower left we show the Pareto front (set of optimal solutions) on the Bicoid-Caudal fitness space. The parameter values have obtained with a evolutionary genetic algorithm for multi-objective optimisation problems, [3, 11].

above (Turing patterns). Each graph represents the distribution of the protein Bicoid measured in vitro and in a different embryo. Each picture corresponds to a different choice of the parameters. We don't list the choice of parameters but in the following we describe the evolutionary genetic algorithm approach used to determine the parameter values.

### 3. Evolutionary genetic algorithms

The determination of the parameters that best fit the noisy data (blue lines) shown in Fig. 4 is called an optimisation problem. In this context, we introduce now some concepts borrowed from biology. The reason for this is due to the fact that we have no guesses about parameters values and the data is not continuous, preventing us from using smoothing techniques. This is a situation very similar to the natural mechanisms leading to biological evolution: there are no preferential choices for the generation of diversity and the selection acts only a posteriori.

As we have to choose the set of parameters that characterise our particular problem, we call a population to the compact set of all the possible parameter values. In the model equations described above, we have seven parameters, and therefore the population of *Drosophilas* is represented by a compact subset of  $\mathbb{R}^7$ . A single member of *Drosophilas* is characterised by a set of parameters that corresponds to a choice of a point of this set. Within this set of points, we can reproduce, mutate, swarm, *etc.*, provided the new or modified parameter values remain in the compact set. So the question is to know which individual (set of parameters) best fits the experimental data.

A genetic algorithm is a procedure to randomly choose individuals of the population that best fit the model with the experimental data. The selection of the best candidates is done by a criterion that can be, for example, the minimisation of a least square deviation function. This procedure to be effective must be implemented in such a way that convergence to a best local or global solution is achieved.

The simpler form of a genetic evolutionary algorithm is the Covariance Matrix Adaptation Evolutionary Strategy, developed by Hansen, [9]. The implementation of this algorithm follows the following steps:

- 1 Choose an initial individual (the Darwinian ancestral)  $p_0$  of the population (the full set of parameter values for the model) and let  $C = I_n$  be a covariance matrix, where  $I_n$  is the  $n \times n$  identity matrix.
- 2 From the multivariate Gaussian distribution with covariance matrix  $C$  and mean value  $p_0$ , sample  $\lambda$  offsprings. For each offspring or set of parameter values calculate the solution of the model equations and then calculate the fitness function, a chi-squared distributions, for example. From the best  $\mu$  ( $\mu < \lambda$ ) offsprings, according to the fitness function, recalculate a new mean value  $p_0$  and a new (unbiased estimator of the) covariance matrix  $C$ .
- 3 Repeat these steps several times. After several iterations, the best individual ever found is a candidate for the best choice of the model parameters.

In general, the genetic algorithm just described has a fast convergence for a local or global optimal solution. The fits of Fig. 4 have been obtained with a similar technique, [11].

This algorithm admits generalisations for the cases where there are conflicting goals, [3, 6, 11], and the solutions of the optimisation problem is not unique. In the context of multiple goals, we have a multi-objective or Pareto type optimisation problem and evolutionary genetic algorithms are a good choice for an unbiased search of realistic solutions of multi-parameter models. For example, in Fig. 5, we show, several (non-unique) Pareto optimal solutions for the distribution of proteins Bicoid and Caudal in the embryo of *Drosophila*. In each picture, we show a different instance of the Pareto optimal solution of the model equations for the same experimental data set. In this case, the model equations do not minimize simultaneously the two experimental data sets. This is a characteristics of multi-objective optimisation techniques where non-unicity is the rule. This is a feature of all the living organisms where individuals of the same species are characterised by different sets of parameters (Pareto optimal), with different performances relative to different properties.

## 4. Conclusions

Using the the Turing morphogenesis hypothesis, we concluded that, in *Drosophila* early development, mRNA diffusion is the main morphogenesis mechanism explaining consistently the establishment of Bicoid protein gradients. Numerical simulations, [3–7, 11], have shown that the diffusion of mRNAs of maternal origin are sufficient to explain the establishment of steady non-uniform gradients of maternal and gap-gene families of proteins along the embryo of *Drosophila*.

The calibration of the morphogenesis models of *Drosophila* early development was done with evolutionary genetic algorithms. These algorithms are a class of probabilistic algorithms that can be used to search for solutions of optimisation problems, without having any a priori knowledge about the solutions or about its regularity properties.

## References

- [1] F. Alves and R. Dilão, A simple framework to describe the regulation of gene expression in prokaryotes, *C. R. Biologies*, 328 (2005) 429-444.
- [2] F. Alves and R. Dilão, Modeling segmental patterning in *Drosophila*: Maternal and gap genes, *J. Theor. Biol.* 241 (2006) 342-359.

- [3] R. Dilão, D. Muraro, M. Nicolau and M. Schoenauer, Validation of a morphogenesis model of *Drosophila* early development by a multi-objective evolutionary optimization algorithm. In: C. Pizzuti, M. D. Ritchie and M. Giacobini (Eds.): EvoBIO 2009, Lec. Notes in Computer Science, vol. 5483 (2009) 176-190.
- [4] R. Dilão and D. Muraro, mRNA diffusion explains protein gradients in *Drosophila* early development, *J. Theor. Biol.* 264 (2010) 847-853.
- [5] R. Dilão and D. Muraro, mRNA diffusion explains protein gradients in *Drosophila* early development, *PLoS ONE* 5 (2010) 5, 1-10 (e10743).
- [6] R. Dilão and D. Muraro, Calibration and validation of a genetic regulatory network model describing the production of the protein Hunchback in *Drosophila* early Development, *C. R. Biologies* 333 (2010) 779-788.
- [7] R. Dilão, Bicoid mRNA diffusion as a mechanism of morphogenesis in *Drosophila* early development, *C. R. Biologies*, 337 (2014) 679-682.
- [8] W. Driever, C. Nüsslein-Volhard, A gradient of Bicoid protein in *Drosophila* embryos, *Cell* 54 (1988) 83-93.
- [9] N. Hansen, *The CMA Evolution Strategy: A Tutorial*, 2008.
- [10] B. Houchmandzadeh, E. Wieschaus and S. Leibler, Establishment of developmental precision and proportions in the early *Drosophila* embryo, *Nature* 415 (2002) 798-802.
- [11] D. Muraro and R. Dilão, A parallel multi-objective optimization algorithm for the calibration of mathematical models, *Swarm and Evolutionary Computation*, 8 (2013) 13-25.
- [12] C. Nüsslein-Volhard and *Coming to life. How Genes Drive Development*. Yale University Press, New Haven, 2006.
- [13] A. Spirov, K. Fahmy, M. Schneider, E. Frei, M. Noll and S. Baumgartner, Formation of the bicoid morphogen gradient: an mRNA gradient dictates the protein gradient, *Development* 136 (2009) 605-614.
- [14] A. M. Turing, The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc. London, B* 237 (1952) 37-72.
- [15] C. H. Waddington, *Organisers and genes*. Cambridge: Cambridge University Press, 1940.