Computerised Modelling for Developmental Biology

Bertens, L.M.F.

CHAPTER 1. MODELLING TECHNIQUES FOR DEVELOPMENTAL BIOLOGY – AN OVERVIEW

τό τε γὰρ μιμεῖσθαι σύμφυτον τοῖς ἀνθρώποις ἐκ παίδων ἐστὶ καὶ τούτῳ διαφέρουσι τῶν ἄλλων ζῴων ὅτι μιμητικώτατόν ἐστι καὶ τὰς μαθήσεις ποιεῖται διὰ μιμήσεως τὰς πρώτας

For it is an instinct of humans, from childhood, to engage in mimesis and this distinguishes them from other animals: man is the most mimetic and it is through mimesis that he gains his earliest insights.

- Aristotle, *Poetica* IV, v.4-8

1.1 MODELLING IN BIOLOGY

A first step in trying to understand the world around us is simplification. We tend to create simplified mental images of both concrete and abstract entities. Looking up at the night sky, for instance, we reduce the infinite number of stars to a set of easily recognisable constellations. We simplify things to a basic set of features, which convey all the information necessary to us. However, decisions about which information is necessary depend entirely on the situation at hand.

Furthermore we try and distinguish patterns in the information we perceive. If a pattern is underlying a series of events or objects, there is no need to remember each of the separate instances, for remembering the pattern itself suffices. Recognizing these patterns enables us to predict future events. Our knowledge of the movement of the planets, for instance, illustrates the use of understanding patterns in observed phenomena.

So, in order to comprehend concrete and abstract concepts, we build (mental) models; instead of using all information at hand we select the most relevant components and construct a simplified version of the entity we want to understand. The word 'model' is defined in the dictionary as "a small object, usually built to scale, that represents in detail another, often larger object" or "a schematic description of a system, theory or phenomenon that accounts for its properties and may be used for further study of its characteristics"[1]. Although clearly distinct, these two definitions both

[1] *The American Heritage College Dictionary,* Boston and New York, 1993.

focus on the representation of complex systems in a simplified manner. Here, as in science in general, we are particularly interested in the second definition[2].

Biological sciences are often concerned with complex systems and studying these would be unthinkable without the use of models. For centuries scientists have built models to aid the understanding of life in all its appearances. This modelling has taken on many forms in diverse fields in the biological sciences, including the 19th-century papier mâché models of Louis Thomas Jerôme Auzoux, the double helix DNA model by Watson and Crick, but also for instance statistical models in palaeontology and the modelling by taxonomic trees in evolutionary biology. Here we will focus on one particular research field, developmental biology, and the use of (computer assisted) modelling in this field. Firstly, model properties will be discussed, important in choosing a method and making modelling decisions. Subsequently a basic outline will be given of the different types of modelling that play a role in studying development.

1.2 MODEL PROPERTIES

The way we model an entity depends on the requirements of the model and the aspects of the structure or the process on which the model focuses. Which features of the studied entity are important to us and which can be ignored given the specific study? When we are interested in gene expression of immune cells in zebrafish we might omit the cells' metabolism from the model, but when studying solely the KREBS cycle the gene expression is less relevant.

Along with the topic of study, the choice of what to include and how to model the information is also dictated by the purpose of the model. Models are built for various reasons: to educate, to test hypotheses, to examine which assumptions about a system best fit the observed data, to predict future events, or to circumvent experiments that are too costly, time consuming or ethically undesirable. All these motives call for different properties of the modelling method. Here the most important sets of properties to be considered when building a model will be introduced and discussed. An overview of these properties is provided in Table 1.1.

[2] It is important to establish an unambiguous understanding of the concept 'model' in this text. The word has many uses, not only in daily life but also in the life sciences. Apart from artificial man-made models there is another very important type of model in biology: animal models. When studying human diseases and physiological or developmental processes it is rarely possible to obtain experimental data from humans; therefore animals are used that resemble the human physiological system under study. This type of model will, however, be excluded from the overview presented in this chapter and the term 'model' refers exclusively to artificial, man-made models, used to represent biological systems.

First of all, models can be either **descriptive** or **mechanistic**. Descriptive models represent observed data, but without modelling the underlying processes that produce these data. For instance, a graph of the numbers of cells in a frog embryo, plotted against time, describes the cell division but does not give us any information about the process. Therefore we cannot determine which variables play a role in the process and the model cannot be used to extrapolate and predict new data. A mechanistic model, on the other hand, describes the underlying process and includes the relevant variables and can thus be used to predict future events. Were we to model the genetic pathways underlying the cell divisions, the resulting model would be mechanistic. While mechanistic models teach us more about the workings of developmental processes, frequently too little is known about a process to construct such a model and constructing a descriptive model is the best option.

Secondly, models can be either **static** or **dynamic**. As the name implies static models describe static non-changing states, without a time component. It is evident that these models are by their very nature descriptive, as opposed to mechanistic. Dynamic models, on the other hand, describe the changes in a system or entity over time (either the simulated time of the modeled process or the running time of the model). They can be descriptive or mechanistic. The Petri net model describing the formation of biological gradients over time, presented in chapter 5 and 6, is an example of a mechanistic dynamic model.

Another defining feature of a model is whether it is **qualitative** or **quantitative**. Many processes can be understood and modelled in a qualitative way, without using exact experimental data. In the case of gradient formation for instance (*cf.* chapter 5), building a model to describe the process does not require exact quantitative data. As such, a qualitative model is a good starting point for modelling a new study, for which experimental data are still incomplete. When, on the other hand, experimental data are available, a quantitative model will allow statistical analysis and make quantitative predictions about future states. Qualitative models can be transformed into quantitative models, by including exact variable values; this is explicitly described for gradient formation in chapter 6.

This distinction between qualitative and quantitative models is directly related to the distinction between **theoretical** and **practical** (or **conceptual** and **concrete**, respectively). Theoretical models are aimed at furthering the theoretical understanding of a system or concept. They do not call for numerical accuracy and are often generic to a group of systems of similar mode of operation. Practical models are used for specific practical situations, in which decisions are made using model predictions and analyses. They are specific to a particular system and include detailed quantitative data. The distinction theoretical versus practical is linked to the trade-off between model generality and model (prediction) strength; a basic model with little

detail will be applicable to a wide range of similar processes, but cannot accurately predict future events for any of those. A highly detailed model, however, has greater predictive power, but only for a specific instance of the process under study. The choice for one or the other depends entirely on the purpose of the model, theoretical or practical.

Models representing system behaviour can do so in a **discrete** or a **continuous** manner. Discrete models represent the behaviour of systems in distinct spatial and/or temporal steps. Continuous models, on the other hand, represent space and/or time in the system's behaviour in a continuous manner, in which no separate steps are discerned. Particular attention is paid to these characteristics and the translation from one to the other in chapter 6.

Furthermore, dynamic models can be either **deterministic** or **nondeterministic**. Whereas series of events in deterministic models are determined at the onset and are predictable, nondeterministic and stochastic models allow multiple outcomes of a process. Stochastic models include the element of probability: events are not set, but depend on levels of probability. In developmental biology nondeterminism can arise through the concept of concurrency (*cf.* chapter 5 and 6) and interaction between process components, while stochastic variables can reflect environmental influences external to the system under study; for instance, the way a fertilized frog egg cell comes to lie on the substrate can be considered a stochastic factor and determines where the animal and vegetal poles of the cell will arise.

MODEL PROPERTIES		
descriptive	versus	mechanistic
static	versus	dynamic
qualitative	versus	quantitative
theoretical	versus	practical
discrete	versus	continuous
deterministic	versus	nondeterministic

Table 1.1. An overview of model properties relevant to modelling in developmental biology.

Apart from these fundamental distinctions there are several other model features that have bearing on the uses of a model. Models can be **temporal** (either discrete or continuous), like the cell division graph mentioned above. **Spatial** models include elements of space, *e.g.* 3-dimensional models of biological structures. These features are not mutually exclusive, in fact many models in developmental biology are both; for instance a model visualising the cell movement of neural crest cells during early development is both spatial and temporal.

1.3 VERBAL AND VISUAL MODELS – NON-COMPUTATIONAL MODELLING

Now that the most significant model properties have been discussed, we will take a look at the different types of models used in developmental biology. There are many ways to divide models into groups, for instance based on the features described above (static models, dynamic models, etcetera). In this overview four types of models are discussed: verbal, visual, algorithmic and equation based models. The last two types are both inherently computational and most often computer-assisted. The first two are not in essence computational and are therefore distinguished in this overview from the computational modelling techniques. However, computation is often possible and useful for these types as well and many verbal and visual models used in studying developmental biology are in fact computer-assisted and/or computational. Since computation is becoming increasingly important for developmental biology, the case studies presented in chapters 2 to 6 make use of computational and/or computer-assisted techniques.

This section will deal with verbal and visual models, *i.e.* model types that are not by definition computational. In 1.4 computational models will be discussed, divided in algorithm and equation based model types. An overview of all four categories and their model types is provided in Table 1.2. For all categories presented here emphasis will be placed on computational and computer-assisted types of modelling since these are rapidly gaining importance in developmental biology. The categorization of modelling types is of course not strict and many models can be considered hybrid and can therefore be assigned to more than one category.

1.3.1 Verbal models

On the most basic level we represent objects and concepts with words. We use language to refer to entities and to enable communication. As such, it is an elementary modelling system. Apart from our everyday, natural vocabulary, we have constructed **terminologies** for many specific research fields, including developmental biology.

VERBAL	- terminologies - ontologies
VISUAL	**SCHEMATIC REPRESENTATIONS** - 2D – schematic drawings - 3D – tangible/virtual 3D models - time – time series of schematic representations **DIAGRAMS** - informal diagrams - formal diagrams / graphical visualisations of computational models **GRAPHICAL REPRESENTATIONS OF TABULAR DATA** - charts
ALGORITHM BASED	- automata based models: • cellular automata • Lindenmayer systems - boolean networks - process algebra/calculi - Petri nets
EQUATION BASED	- differential equations: • gene regulatory networks • reaction diffusion systems • diffusion models - enzyme kinetics - organismal growth models

Table 1.2. An overview of the categories and types of models presented in this chapter.

Classification of biological entities is by no means a new phenomenon. Throughout the ages people have tried to define formal classifications and terminologies; as early as the fourth century BC, Aristotle wrote extensively about nature, classifying animals and plants (*cf. Historia Animalium*) in a system which has influenced classifications all throughout the Medieval period and the Renaissance. During the age of enlightenment the very nature of science and experimental studies changed and scholars focused their energy on systematically collecting information and, more importantly, sharing the new-found knowledge. Diderot's *Encyclopédie* (1751-1772) was as such an invaluable document, as was Linnaeus' *Systema Naturae* (1735). With modern science becoming more international, conventional terminologies and definitions are crucial in communication between scientists of a particular research field, and formal systems of knowledge are indispensible.

The field of developmental biology relies heavily on terminologies and therefore stands to gain from optimizing the efficient use of this knowledge. A useful modern method of formally representing a terminology, containing the relevant objects, concepts and entities and the relationships between these terms, is collecting this knowledge in an **ontology** (Gruber, 1993). Put another way, an ontology is a verbal model of a domain, comprising the knowledge of all concepts within that particular domain and the way these are related. The knowledge is structured in the ontology in 'classes' (analogous to encyclopaedic lemmas), which are linked in a hierarchical system by 'IS_A' relationships (*e.g.* 'cow' IS_A class in the superclass 'vertebrates'). In addition to the IS_A relationship further 'properties' (relationships between the classes) can be defined within the ontology. In this way, ontologies allow researchers to describe an area of research, using a shared standardized vocabulary, and to classify the concepts and relationships in this area into groups and hierarchies.

Ontologies are becoming increasingly important in biology in general (Dmitrieva, 2011), as exemplified by the successful Gene Ontology, GO (The Gene Ontology Consortium, 2000), and the Sequence Ontology (Eilbeck *et al.*, 2005). In an attempt to oversee and coordinate the ever growing number of biological ontologies, the Open Biomedical Ontologies consortium (OBO) was founded (Smith *et al.*, 2007; Bodenreider and Stevens, 2006). By providing shared principles, such as the use of a common representation in the OBO format or the Web Ontology Language (OWL, https://www.w3.org/TR/owl-features), the consortium helps make ontologies interoperable and of greater use to biologists. Amongst other ontologies, OBO contains a number of specific developmental ontologies, dealing with the development of model species such as *Arabidopsis thaliana*, *Caenorhabditis elegans* and *Drosophila melanogaster*. These ontologies include information about the different anatomical structures present at particular developmental stages.

The most basic goal of ontologies is to formally capture the knowledge of a specific research area and to provide researchers with a standardized vocabulary to communicate. In addition to this, they can also be used to link terms in the vocabulary to other resources in the research field, such as literature and images, as is possible with for instance the GO. Over the last few years progress has been made on implementation of a Semantic Web for the life sciences (Bodenreider and Stevens, 2006). Nowadays ontologies are an integral part of bioinformatics and developmental biology, so much so that the leading journal *Bioinformatics* now dedicates an entire section to this field.

In essence, ontologies are descriptive rather than mechanistic and static rather than dynamic. However, since ontologies by their very nature provide great freedom in relating terms using relevant properties (for instance concerning processes taking place between structures under differing circumstances), they can be used to include mechanistic and dynamic information. An example of this can be found in

chapter 4, where dynamic information on heart physiology, under differing circumstances, has been added to an ontology of heart anatomy and development. Furthermore, while ontologies can be used entirely qualitatively by including solely structure names and qualitative properties, it is also possible to add quantitative information about the structures in data properties. This also allows one to transform a theoretical model into a practical one. Since any semantic content can be modelled in an ontology, the information can be both temporal and spatial.

In chapter 4 we introduce our use of ontologies to capture knowledge about the anatomy, development and physiology of the vertebrate heart. This serves as a case study for extending the functionality of ontologies for developmental biology; as will be explained in more depth in chapter 4, it incorporates the whole range of information types, from static, qualitative knowledge to dynamic, quantitative information in both temporal and spatial domains. By connecting the ontology to 3D models (*cf.* 4.4.4) we integrate verbal modelling and visual modelling. This integration of different types of models will be further discussed in chapter 7.

1.3.2 Visual models

In the biological sciences visual information is very important, more so than in the mathematical, physical or chemical sciences, where abstract formulas can more easily be used to describe and study the systems at hand. Biological textbooks illustrate our knowledge of biological structures with drawings and diagrams and more and more journal publications refer to websites with supplementary figures and visual aids, to be downloaded or viewed online (*e.g.* Noble, 2002). Not only can visual models help clarify information, they can also lead to new insights that would be hard to achieve through other means (Iwasa, 2010); in science, as in life in general, a picture is worth more than a thousand words.

Before looking at the different types of visual models, I would like to stress that we are dealing with modelling techniques, not visualisation techniques in a broader sense. In biology the concept of 'visualisation' encompasses (in addition to visual modelling) both techniques with which biology can be made visible, *i.e.* a wide range of microscopy modalities, and techniques with which this visual information can be captured, *i.e.* imaging techniques. Whereas these techniques try and capture reality without simplification or selection, models specifically reduce information from reality to a set of essential features. These simplifications can have various purposes; they can be used in education, to describe our knowledge of biology, or to present hypotheses.

There are various types of visual models in developmental biology. Here I present a division into three basic categories: schematic representations, diagrams and graphical representations of tabular data, *cf.* Table 1.2.

Schematic representations

The first category comprises schematic representations; these are simplified yet faithful reproductions of biological reality. Examples in 2D are the **schematic drawings** of the turtle heart shown in Figs. 2.1, 2.2 and 2.3 (in the next chapter), drawings of animals or plants found in identification guides and cutaway drawings of anatomical structures. They are true to life (albeit in a simplified form) and are widely used to clarify descriptions or present hypotheses about the nature of structures in (developmental) biology. In three dimensions, schematic representations can be **tangible 3D models**, like the papier-mâché models made by Louis Auzoux in the 19th century or the plastic models used by medical specialists to help patients understand; however, they can also be **virtual 3D models**, like the 3D computer reconstructions of hearts presented in chapters 2 and 3. Virtual 3D modelling has become increasingly important over the last decennia with new advances in computer science leading to more sophisticated software, *e.g.* Mimics (www.materialise.com/mimics), Amira (www.amira.com), BioVis3D (www.biovis3d.com) and the software used to render the models in chapters 2 and 3, TDR3Dbase (Verbeek and Huijsmans, 1998). Computer graphics enable us to display 3D structures using sophisticated rendering algorithms. Two of those commonly used algorithms are volume rendering and surface rendering. Whereas volume rendering does not explicitly define the surface geometry, surface rendering represents the surface by polygons, most often using triangulation (Verbeek and Huijsmans, 1998; Eils and Athale, 2003), *cf.* chapter 2. In addition to making 3D structures more insightful, the conversion of 2D image sections to 3D surface models allows the user to perform quantitative analyses on 3D features such as volume and shape (Eils and Athale, 2003).

These 2D and 3D schematic representations are descriptive and static. They are qualitative and generally theoretical, although, as mentioned, quantitative analyses can be feasible in 3D models. Furthermore they are essentially spatial models, without a temporal aspect. However, by putting several of these representations (either 2D or 3D) of different (developmental) moments together a temporal element can be introduced. Chapter 2 presents a developmental series of turtle heart development, using models of different developmental stages. In this way the static representations are combined in a temporal model, albeit still consisting of static, discrete elements. Other examples of these types of **time series of schematic representations** would be cell or life cycle sequences. By animating the development in a continuous sequence (a movie), the model becomes dynamic. Nowadays **animations** of biological events can be found on numerous scientific websites, for instance the website of *Molecular Movies* (www.molecularmovies.com). Even so, all models in this category, including the

temporal ones, are still in essence descriptive rather than mechanistic; they merely illustrate processes and structures as opposed to explaining them.

Diagrams

The second category of visual models contains diagrams; here I will define a diagram as an abstract graphical representation of a process[3]. The keyword in this definition is 'abstract'; whereas schematic representations are true to life and directly reflect reality, diagrams reflect information about reality in an abstract, indirect way, using conventional symbols.

Two main uses of diagrams can be distinguished in developmental biology. Firstly they can be used informally to illustrate our knowledge of different steps of a process. As such, **informal diagrams** are particularly helpful in the relatively new field of systems biology. Signal transduction pathways, gene regulatory networks and metabolic pathways are frequently illustrated in diagrams. Symbolic representations of the elements in the process clarify the reactions taking place. These models are informative and help us visualise processes but cannot be used for computational analysis.

Secondly, diagrams can be used as **graphical visualisations of formal computational models**. In their most elementary form these diagrams are graphs consisting of nodes, connected by arcs. Depending on the focus of the model the nodes can represent (amongst others) classes, objects, processes or states. These diagrams can be classified as **class diagrams**, **object diagrams**, **flow charts** or **state diagrams,** respectively. As described in 1.4, there exists a wide range of computational modelling techniques which allow the user to model and analyse processes underlying biological events. Many of these techniques have a graphical component, illustrating the modelled process (Biermann *et al.*, 2004). The modelling technique of Petri nets, which is discussed in depth in chapters 5 and 6, is for instance characterized by the combination of graphical notation and exact mathematical definitions, to which we will come back in 1.4.1. Here the diagram, showing the entities and events of the process, contributes to understanding the model. In addition to this graphical component, which focuses on production and consumption of local resources and their transitions, the information contained in a Petri net can also be visualised in state diagrams, showing the different states and state changes of the system as a whole. Both visualisations belong to the category of formal diagrams.

[3] Note that in colloquial use the difference between a schematic representation and a diagram is rather hard to define. Here, I will not dispute the exact meaning of these words; the narrow definitions given are solely meant to clarify the classifications of model types presented here.

Another example of a diagram connected to an underlying formal model would be the visualisation of an ontology; here the focus lies on the classes and their mutual relationships and this type of diagram could therefore be categorized as a class diagram (we will look at the visualisation of ontologies in more depth in chapter 7).

An important modelling system consisting of formal diagrams is the object-oriented Unified Modelling Language (UML, www.uml.org). UML comprises fourteen types of diagrams, amongst which class diagrams, object diagrams and activity diagrams (which are related to flow charts). By assigning formal meanings to visual symbols, this language and formal diagrams in general allow the user to formally communicate knowledge. As is often the case with modelling techniques, UML was developed for different purposes (the modelling of computing systems), but has recently been shown to be useful in the field of biology as well (Roux-Rouquié *et al.*, 2004).

Since diagrams are most frequently used to present the development of a process, they often have a time element. This is not necessarily the case however, as exemplified by ontology visualisations and certain types of UML models. Also informal models depicting pathways and networks in systems biology cannot be said to have a true time component; rather, different processes, that are in some way linked, are shown in one diagram. On the other hand informal diagrams that represent a clear sequence of events with time indications, like life cycles, are in fact temporal models. It should be underlined that the difference with the temporal sequences mentioned under 'Schematic representations' is that diagrams consist of abstract symbols, whereas the temporal sequences discussed earlier are composed of truthful representations.

Graphical representations of tabular data

The third and last category of visual models consists of graphical representations of tabular data; tabular data are alphanumerical data (quantitative of qualitative) that can be stored in rows and columns in a database, *e.g.* concentration levels of proteins over a distance in a cell culture, but also the names of research group members in charge of cleaning the lab over a series of months. These types of data can be modelled in a variety of **charts**, such as **histograms**, **bar charts**, **graphs**, **scatter plots** or **cartograms**. Since these charts are simplified visual representations of selective information from reality, I include them in our overview of visual models. These models can be multi-dimensional; simple charts plotting two variables are two-dimensional, but more dimensions can be added for more variables. Models of tabular data are descriptive and static; they are practical (pertaining to specific situations) and can be either quantitative or qualitative, modelling spatial or temporal information, or both.

1.4 ALGORITHMIC AND EQUATION BASED MODELS – COMPUTATIONAL MODELLING

As we have seen, verbal and visual models have been used for centuries in the field of developmental biology. More recently new ways of modelling biology have been developed; integration of biology with the field of mathematics and the relatively new field of computing science has led to models that allow us to study biology through calculation and simulation. For this new area of biology the term 'computational biology' has been coined, which is defined by the NIH as "the development and application of data-analytical and theoretical methods, mathematical modelling and computational simulation techniques to the study of biological, behavioural and social systems" (NIH, 2000)[4]. This definition is followed throughout this chapter. The essential and binding element of the models discussed in the remainder of our overview is therefore computation.

For computational modelling the same holds true as for the entire field of models in developmental biology: models can be categorised in more than one way. Here a distinction is made between **algorithmic process models** and **equation based models**. While both types are concerned with biological processes, the difference between models based on algorithms and those based on equations is a difference of focus on **operation** versus **denotation** (Fischer, 2007; Priami, 2009); *i.e.* equations abstract away from the actual process and solely describe the changes in terms of variable values when a system moves from one state to another, making them denotational. They model systems by functions which transform input into output. Algorithms, on the other hand, focus on the process and describe how and why the system changes between states, which makes these models operational. The distinction between algorithmic process models and equation based models corresponds to the distinction Fischer (2007) makes between computational and mathematical models, respectively; since the terms ´mathematical´ and ´computational´ can both be used in various and sometimes interchangeable ways, these classifications are avoided here and the word ´computational´ is used only in the broad sense, as specified by the definition given in this chapter.

The distinction between algorithm and equation based models is directly related to three important pairs of polar opposites in process models :

1) **discrete versus continuous**; algorithmic models describe a system in discrete state spaces (*i.e.* the set of all states a system can be in), whereas state spaces in equation based models are generally continuous (although they can be discretised, *cf.*6.1). However, there are hybrid techniques, for instance Hybrid Petri Nets (David and Alla,

[4] This definition is slightly circular, for it does not explain the term ´computational´. Here we understand this term to refer to using either calculation or a computer.

2001; Matsuno *et al.*, 2003), in which both discrete state spaces and continuous variables are used.

2) **concurrent versus sequential**; most processes in developmental biology are characterised by concurrency (*cf.* chapters 5 and 6), making it an important factor in the choice of a modelling technique. Algorithmic models are typically applied when modelling cause and effect which can lead to concurrency or nondeterminism. Equations on the other hand lend themselves more readily for modelling sequential processes, in which input information determines global behaviour. These models are strictly functional (*i.e.* based solely on mathematical functions; specific input always yields the same output) and are therefore deterministic.

3) **averaged versus individual**; equation based models generalise the process behaviour by describing the average global behaviour of the system. By contrast, for algorithms each simulation describes the exact behaviour of one execution of the process.

With these distinctions in mind, we will now look at some algorithmic and equation based models, commonly used in developmental biology.

1.4.1 Algorithmic process models

Algorithmic process models focus on describing the dynamics of systems. As such they are particularly useful in developmental biology, where researchers are most often interested in the biological process itself. Over the past decades several algorithmic modelling techniques have been developed, often originating in the computing sciences, and have come to be used in the field of bioinformatics.

In a highly simplified form the behaviour of a system can be described as follows: at the start of the process an input state is provided and this is followed by an output state; in this context a state is a particular configuration of the information in the system. This simplification forms the basis for the development of **automata theory** (Hopcroft *et al.*, 2000). In this field processes are seen as automata, which are abstract **state machines** for which a number of states is provided along with a set of rules, which govern transitions from one state to another, taking the current state as input. In addition to the set of states and transition rules also one or more initial states and final states are specified. The process starts off in an initial state and for each next step the provided rules determine the next state of the system, until a final state is reached.

John von Neumann extended this system and developed **cellular automata** (Von Neumann, 1966), which have several biological applications. Cellular automata consist of a grid of identical, interdependent automata, referred to as cells. Again each

of the cells is in a certain state and its next state is determined by a set of rules. Here the rules take as their input not only the current state of that particular cell but also the states of a set of neighbourhood cells (and possibly external factors); this allows local interaction between cells to be modelled, making the technique suitable for modelling more complex systems, characterized by concurrent and locally determined behaviour. In developmental biology cellular automata have commonly been used to study the formation of patterns, such as pigmentation patterns in seashells (Meinhardt, 1982; Plath *et al.,* 1997).

A related modelling system was developed by Aristid Lindenmayer in 1968, the **Lindenmayer system** or **L-system** (Lindenmayer, 1968). This system resembles the system of cellular automata but while cellular automata only change the states of cells but never remove or add cells, Lindenmayer systems allow the appearance of new cells and the disappearance of existing cells. Put differently, cellular automata use *state transition rules*, while Lindenmayer systems use *rewriting rules*. This makes Lindenmayer systems interesting for biologists looking to model growth and development; due to this feature L-systems have been used successfully to model plant development, *e.g.* through arborisation (Prusinkiewicz, 1995).

Another related approach is the **boolean network**, which in a sense is a generalization of the cellular automata. The network consists of nodes, analogous to the cells in cellular automata, but while the next states of cells in cellular automata depends on the states of the direct neighbours, nodes in a boolean network can be connected to and therefore influenced by any other node in the network. The nodes can be in one of two states, one or zero, and as in cellular automata the next state of a particular node is determined by the states of the connected nodes (and the current state of that node), depending on certain logical functions. This system was originally developed by Stuart Kauffman to model genetic regulatory networks (Kauffman, 1969), in which genes are either active (in state 1) or inactive (in state 0) and interact through the network. These modelling systems can be used more generally to study the interaction of biological networks, consisting of *e.g.* proteins and genes. Although this technique strongly simplifies the biological process, it can still yield useful results regarding pathways and gene regulatory networks (Fischer and Henzinger, 2007). A downside to this technique is that the networks cannot be integrated into larger models, *i.e.* they cannot be constructed from smaller sub-parts which are combined in a hierarchical structure.

In contrast to this, **process algebras**, or **process calculi**, allow a hierarchical organization of sub-parts, *i.e.* compositional modelling. This field of modelling is concerned with studying concurrent or distributed systems by algebraic means, *i.e.* the methods of algebra are applied to the behaviour of a system (Baeten, 2005). The compositionality of the system implies that the behaviour of the system as a whole is expressed through the behaviour of its components and their interaction. This

interaction can be synchronous or asynchronous (with components changing states independently), the latter resulting in nondeterminism since different sequences of events might lead to different outcomes (Fischer and Henzinger, 2007). These aspects in particular make process algebras suitable for modelling biological behaviour; in biology, as in distributed systems, many processes are active simultaneously and compete over the use of resources while also cooperating to accomplish a common goal (Priami and Quaglia, 2005). Process algebras have been used successfully in modelling biological pathways and networks within the fields of systems biology and developmental biology (Priami and Quaglia, 2005; Fischer and Henzinger, 2007).

A strong emphasis on concurrency and local independence can also be found in **Petri nets**, which have been mentioned before in 1.3.2 and will be discussed more thoroughly in chapters 5 and 6. A Petri net is an abstract model of information flow. It has both a graphical representation and a formal mathematical definition. The model consists of places (representing passive entities like resources), transitions (representing actions or events) and the flow relation between them. Graphically the net is presented as a bipartite directed graph with places depicted as circles and transitions depicted as rectangles; these two types of nodes are connected by edges, depicting the flow relations. The state of the net is represented by the distribution of tokens (black dots) over the places and the dynamic behaviour of the system is represented by the production and removal of these tokens, governed by occurrences of the transitions. The graphical component makes the net intuitively understandable, while the formal definition enables precise analysis. Petri nets are well-established as models of concurrent and distributed systems and have recently gained importance in the field of biology, in particular biochemistry and developmental biology (Steggles *et al.*, 2007; Heiner *et al.*, 2008; Krepska *et al.*, 2008).

This list of algorithmic process models for developmental biology presented here is far from complete. In addition to lesser known models already in use, new modelling techniques are constantly being developed. Furthermore, already existing modelling techniques, used for instance in computing sciences, are being ‘re-used’ for biological purposes on a regular basis. As is obvious from their focus on processes, algorithmic models are dynamic. There are both quantitative or qualitative algorithmic models (*cf.* chapter 6), both descriptive or mechanistic (depending on whether their representation of the underlying biological process is true to life or solely meant to produce correct outcomes), theoretical or practical and deterministic or nondeterministic. Finally, they may or may not include temporal and spatial elements.

1.4.2 Equation based models

The final category of models, discussed in this overview, comprises the equation based models. These do not model the stepwise executing of natural processes the way algorithms do and allow exact quantitative analysis. As such they complement the qualitative process models. By combining the analysing strength of equation based models with the operational focus of algorithmic models, developmental processes can be studied in detail from different perspectives.

Equations can help study a great number of processes in biology in general and developmental biology in particular. Of special importance are **differential equations**, which constitute by far the most used formalism to model dynamical systems (De Jong, 2002; Priami and Quaglia, 2004; De Jong and Geiselmann, 2005) and have been applied to a variety of biological phenomena, of which gene regulatory networks, reaction-diffusion systems and diffusion models will be discussed here.

The first category of cellular processes, relevant to developmental biology and modeled using differential equations, comprises **gene regulatory networks**. The process of gene expression takes place in two steps: transcription (in which 'reading out' a string of DNA produces a complementary mRNA copy in the cell nucleus) and translation (in which proteins are produced by 'reading out' the mRNA at the ribosomes in the cell). Both processes are influenced by additional reactions and proteins; in order for transcription to take place a polymerase has to attach to the promotor site of the string of DNA and, upon transcription, to detach at the terminator site. This binding of the polymerase to promotor is regulated by other proteins, both repressor proteins (inhibiting transcription) and activator proteins (actively promoting transcription). These repressor and activator proteins are themselves products of gene transcription and translation and the combined interactions between these proteins and gene expression patterns are captured in gene regulatory networks. As is obvious from this description, these networks can become very complex, all the more so since they often include positive or negative (self) feedback loops. A simple example of such a loop is a gene which codes for a protein that actively inhibits the expression of that same gene.

These complex processes play an important role in the field of systems and developmental biology and a vast amount of models exist, mainly focused on differential equations (DEs) (De Jong, 2002; De Jong and Geiselmann, 2005). Nonlinear ordinary DEs model gene regulation by rate equations, in which the rates of change in space and/or time of one component in the system is expressed as a function of the concentrations of other components. The regulatory interactions of the network under study are captured by functional and differential relations between these concentration variables. Depending on the number of independent variables of the unknown function,

either ordinary DEs (in case of a single variable) or partial DEs (in case of multiple variables) are used.

Nonlinear ordinary DEs give an adequate description of the dynamics of gene regulation, but can be mathematically difficult to analyse. Linear ordinary DEs simplify the system and are easier to analyse, but this is countered by a reduced ability to include essential properties of the system (De Jong and Geiselmann, 2005). An intermediate solution is provided by piecewise-linear DEs, which are globally nonlinear and locally linear. They simplify the system, by abstracting biochemical details and approximating the continuous sigmoid curves of the gene behaviour in discontinuous step functions, but still retain the ability to adequately represent the system (De Jong and Geiselmann, 2002).

Apart from gene regulation many other biological processes can also be modelled using differential equations. Another type of model concerned with biochemical interactions is the **reaction-diffusion system**. This model was proposed by Alan Turing in his influential paper of 1952 (Turing, 1952), in which he showed how the interaction of two substances with different diffusion rates can lead to pattern formation. In developmental biology this corresponds to a system of chemicals, called morphogens, diffusing through a tissue and interacting, which results in stable patterns underlying morphogenesis; patterns arise, of alternating high and low concentration areas of a morphogen, accounting for instance for pigment stripes (Gilbert, 2000). The wavelength of these patterns is determined by the diffusion constants and reaction rates of the different morphogens involved. A reaction-diffusion system can be modelled by partial differential equations, consisting of a diffusion component and a reaction component.

The reaction-diffusion system can be simplified by leaving out the reaction component, which results in a **diffusion model**. This describes the spread of particles from areas of higher concentration to areas of lower concentration, like the spreading of a drop of ink in water. In 1970, Francis Crick proposed to apply this model to the biological process of gradient formation (Crick, 1970). This is a pivotal developmental process, which plays an important role in the establishment of body axes and morphogenesis. He modelled this process as a source-sink system, in which a morphogen gets produced at the source, spreads through Brownian motion and gets removed again at the sink. This model has formed the basis for many DE models describing gradient formation (Gregor *et al.*, 2005; Kicheva *et al.*, 2007; Yu *et al.*, 2009). We will come across this use of DEs or gradient formation in chapter 6, in which the modelling of this process by DEs will be complemented by a Petri net model.

In addition to differential equation models, several other equaion based models are used for biology. Essential to many processes in the cell are enzyme-

mediated biochemical reactions, in which a substrate undergoes a reaction, instigated by an enzyme. The dynamics of these are described by **models of enzyme kinetics**. As early as 1913 Leonor Michaelis and Maude Leonora Menten developed an equation to describe the rate of single-substrate enzyme kinetics, based on the reaction rate, the concentration of the substrate and the Michaelis constant of that substrate (Menten and Michaelis, 1913). The importance of this model stems from its general applicability; it describes the reaction behaviour of thousands of enzymes and therefore allows biologists to study and predict enzyme reactions in a wide range of biological processes (Jungck, 1997).

Finally, the important developmental process of growth is described by **models of organismal growth**. Here a distinction can be made between isometric and allometric growth. Isometric growth refers to an increase in volume and size, while retaining the original proportions, *i.e.* the shape is preserved because all parts grow at the same speed. This growth is for instance seen in the spiral growth of shells and can be expressed in equation models (Gilbert, 2000). Allometric growth or allometry refers to the process in which some parts grow at a different rate than others, thereby changing the overall proportions. Human allometry can easily be seen when comparing the head-limb ratio of babies to that of adults. Another clear example is the growth of the male fiddler crab; its claws are initially the same size, but as the crab grows the crushing claw becomes proportionally larger (Gilbert, 2000). The underlying laws of these types of growth can again be captured in mathematical equations.

These equation based models are first and foremost quantitative and deterministic. Their abstraction from the actual processes makes them descriptive and often practical rather than theoretical. Both spatial and temporal information can be included in these models; each of these properties will be discussed in chapter 6, when we will compare an algorithmic modelling technique with an equation based technique in order to develop new methods to combine the strengths of these approaches.

_____

This concludes the overview of models used for developmental biology. It goes without saying that this modest chapter can by no means offer an exhaustive overview; many other models are used for a wide range of developmental phenomena. The outline of the dissertation follows the theoretical overview, provided in this introductory chapter. Each of the four model categories is addressed in a case study. Chapters 2 and 3 present several 3D models, visualising aspects of heart development in the turtle *Emys orbicularis*. These visual models illustrate the importance of 3D modelling in understanding complex anatomical structures. As such they belong to the

class of schematic representations, within the category of visual models. Chapter 4 concerns the construction of a multi-species ontology of the vertebrate heart, exemplifying the category of verbal models. In chapters 5 and 6 two modelling approaches of the process of gradient formation are presented, both in the modelling framework of Petri nets; while chapter 5 focuses on a qualitative approach, chapter 6 integrates parameters of differential equations into the Petri net model. These chapters are therefore connected to the categories of algorithmic and equation based models, respectively. Apart from exemplifying the different modelling categories, each of the studies presented in the following chapters also features integration of modelling methods. In the final chapter we look into these different types of integration, encountered in the presented studies.