Diss. ETH № 18821

Modeling of Gene Regulative Networks in Developmental Systems

A dissertation submitted to
ETH Zurich

for the degree of

Doctor of Sciences

presented by

Tim Hohm

Dipl.-Inform., Universität Dortmund, Germany
born April 8, 1979
citizen of Germany

accepted on the recommendation of
Prof. Dr. Eckart Zitzler, examiner
Prof. Dr. Sven Bergmann, co-examiner
Prof. Dr. Rüdiger Simon, co-examiner

2009

Institut für Technische Informatik und Kommunikationsnetze
Computer Engineering and Networks Laboratory

TIK-SCHRIFTENREIHE NR. 111

Tim Hohm

Modeling of Gene Regulative Networks in Developmental Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

A dissertation submitted to
ETH Zurich
for the degree of Doctor of Sciences

Diss. ETH № 18821

Prof. Dr. Eckart Zitzler, examiner
Prof. Dr. Sven Bergmann, co-examiner
Prof. Dr. Rüdiger Simon, co-examiner

Examination date: December 18, 2009

Acknowledgments

First of all, I would like to thank Eckart Zitzler for his support and supervision, as well as Sven Bergmann and Rüdiger Simon for their willingness to review this thesis.

Additional thanks goes to all members of the SY-STEM project for an informative and insightful time and joyous project meetings throughout Europe, and in particular to Rüdiger Simon for inspiring discussions and his endurance in answering my questions about the functioning of plant biology.

Not to forget my colleagues at the TIK for an interesting and really good time during the past four years, with a special thanks to Johannes who always took his time to help me out with whatever questions I had with respect to various graphics.

Finally, I would like to acknowledge funding of the SY-STEM Project by the European Commission under the Marie Curie Research Training Network, Project 5336.

Abstract

On their way from a fertilized egg to a fully grown organism, higher organisms run through a series of developmental stages. This development can be seen as a program that is stored in the genome of the organism and the execution of this program is determined by the genes regulating each other's activities. Up to now, the exact mechanisms for turning a one-dimensional genome into a fully three-dimensional organism remain hidden. Investigating the underlying processes, developmental biologists try to unveil these mechanisms but developmental systems constitute a difficult body of research: the regulative processes tend to be non-linear and commonly involve several genes simultaneously. At the same time, advances to register these processes experimentally are hampered by the fact that they not only depend on local information like a single cell's state at a given point in time but on spatial aspects as well, i.e., other cells' states. In result, the spatial component complicates experimental data acquisition in suitable resolution and the non-linearity of the regulative processes hampers intuitive interpretation of the available data. To aid data interpretation, mathematical models have become one of the key tools. They provide possibilities to formalize current knowledge, to validate hypotheses, and to assist in designing new experiments.

In the process of modeling developmental systems, one commonly starts from an initial hypothesis about the underlying regulative processes that is translated into a model of appropriate detail, often in form of coupled differential equations. Then, the resulting model has to be calibrated using the available experimental data before it can be used. While researchers are assisted in setting up suitable models by a range of formalisms, the process of model calibration is less understood: due to the difficulties in experimentally determining system behavior, data tends to be of qualitative nature whereas existing general purpose calibration methods rely on quantitative data.

In this context of modeling gene regulative networks underlying developmental systems, this thesis addresses two topics:

- What is a suitable method to calibrate models using qualitative data only? Investigating this question, existing custom tailored calibration techniques are reviewed and new concepts are developed in order to come up with techniques that can be used for model calibration based on qualitative data. Specifically, methods to quantify the match between qualitative data and models are proposed, suitable criteria to guide a calibration process are investigated, and methods to integrate different sources of domain knowledge are studied.
- The developed methods are used to build a model for the gene regulative network underlying the development and maintenance of the shoot apical meristem of *Arabidopsis thaliana*. The shoot apical meristem is a tissue that is responsible for the aerial growth of plants. To maintain growth, it harbors a pool of stem cells that provide necessary undifferentiated cells for the growth. The system is of interest since investigating the exact regulative mechanisms controlling stem cell homeostasis can provide further insights into stem cell regulation in general and eventually might enable influencing aspects like crop sizes. The built model is then used to explore the parameter space that allows stem cell maintenance, and to simulate the consequences of mutations, gene misexpression and cell ablations.

In result, the presented shoot apical meristem model in combination with the developed calibration methods represents a basis that can be extended to include subsystems responsible for further aspects of plant development—finally to become a virtual meristem or even plant that would be a comprehensive knowledge-base for plants. In addition, the proposed framework and calibration techniques form a useful toolkit for modeling other developmental systems.

Zusammenfassung

Auf dem Weg von einer befruchteten Eizelle hin zu einem ausgewachsenen Lebewesen durchlaufen höhere Organismen einen Entwicklungsprozess, der sich in verschiedene Phasen unterteilt. Dieser Prozess folgt einem Programm, das im Genom des Organismus gespeichert ist. Der Programmfluss wird dabei durch die gegenseitige Regulation der beteiligten Gene gesteuert. In diesem Zusammenhang sind allerdings bis dato die genauen Mechanismen unbekannt, die für die Umsetzung des eindimensionalen Genoms in ein dreidimensionales Lebewesen verantwortlich sind. Details dieser Mechanismen werden in der Entwicklungsbiologie untersucht, wobei die betrachteten regulativen Prozesse meist nichtlinear sind und von mehreren Genen gleichzeitig abhängen. Darüber hinaus sind die betrachteten Systeme nur schwer mit aktuellen experimentellen Methoden zugänglich, da die regulativen Prozesse zusätzlich zu lokalen Eigenschaften auch von räumlichen Aspekten abhängen und diese räumlichen Beziehungen nur schwer experimentell ermittelt werden können. Unter lokalen Eigenschaften zählt hierbei der Zustand einer Zelle zu einem bestimmten Zeitpunkt, während räumliche Aspekte z.B. die Abhängigkeiten der Entwicklung einer Zelle von den Zuständen benachbarter Zellen umfassen. Somit sind für entwicklungsbiologische Systeme oftmals nur unzureichende Daten verfügbar, um Rückschlüsse auf die zugrunde liegenden regulativen Prozesse zu erlauben.

Um Daten dennoch interpretieren zu können, haben sich in diversen Anwendungen mathematische Modelle bewährt: sie erlauben eine kompakte, formale Darstellung der gesammelten Erkenntnisse, und können eingesetzt werden, um Hypothesen zu testen oder neue Experimente zu entwickeln.

Im Kontext von entwicklungsbiologischen Systemen beginnt der Modellierungsprozess typischerweise mit einer ersten Hypothese bezüglich der erwarteten regulativen Prozesse und beteiligter Gene. Diese Hypothese wird dann in ein Modell geeigneter Detailstufe übersetzt. Ein in diesem Zusammenhang häufig verwendetes Modell sind Systeme von Differenzialgleichungen. Bevor diese Modelle jedoch benutzt werden können, müssen zunächst Modellparameter kalibriert werden. Während es für die Übersetzung von

Hypothesen in Modelle diverse Formalismen gibt, ist der Prozess der Modellkalibrierung weniger gut untersucht: experimentell gewonnene Daten sind üblicherweise qualitativ, während gängige Methoden zur Modellkalibrierung die Verfügbarkeit quantitativer Daten voraussetzen.

In diesem Rahmen befasst sich diese Arbeit mit den folgenden beiden Themenkomplexen:

- Wie sieht eine geeignete Methode aus, die zur Kalibrierung von Modellparametern nur qualitative Daten verwendet? Im Lauf der Untersuchung dieser Fragestellung werden in der Literatur dokumentierte, typischerweise problemspezifische Methoden analysiert, um neue Ansätze für ein generell einsetzbares Verfahren zu erarbeiten. Im Detail werden Methoden entwickelt, die es ermöglichen die Übereinstimmung zwischen Modell und qualitativen Daten zu quantifizieren. Es werden Varianten untersucht, wie diese Quantifizierung in den Kalibrierungsprozess eingebunden werden kann. Zudem werden Möglichkeiten getestet, die es erlauben, problemspezifisches Wissen in den Kalibrierungsprozess einzubeziehen.
- Die dabei entwickelten Methoden werden verwendet, um ein Modell für das genregulative Netzwerk zu erstellen, das der Entwicklung und dem Erhalt der Sprossapikalmeristeme in *Arabidopsis thaliana* zugrunde liegt. Meristeme sind Gewebe, die für das oberirdische Wachstum der Pflanze verantwortlich sind. Um das Wachstum zu ermöglichen, beherbergen diese Meristeme Stammzellen, die die Pflanze mit undifferenzierten Zellen versorgen. Die genaue Regulierung dieser Stammzellansammlungen ist dabei aus zwei Gründen interessant: zum einen hat sie einen gewissen Modellcharakter für Stammzellregulation in anderen Organismen, z.B. auch für Wirbeltiere, zum anderen könnte detailliertes Verständnis dieser Prozesse es ermöglichen, direkten Einfluss auf Aspekte wie Ernteerträge zu nehmen. Das resultierende Modell wird im Rahmen dieser Arbeit dazu benutzt, Bedingungen zu untersuchen, die den Erhalt dieser Stammzellen erlauben. Desweiteren werden mit Hilfe von Simulationen Mutationsexperimente, Misexpressionsexperimente und Zellabtragungsexperimente nachvollzogen.

Zusammen mit den entwickelten Methoden zur Modellkalibrierung, stellt das Sprossapikalmeristemmodell somit eine Basis dar, die leicht um weitere regulative Komponenten erweitert werden kann. Auf lange Sicht ist dies ein erster Schritt hin zu einem virtuellen Meristem oder gar einer virtuellen Pflanze – ein Gebilde, das Wissen bündelt und einfach abrufbar macht. Darüber hinaus bietet die Kombination aus Modellierungsumgebung und Kalibrierungsmethoden einen Ausgangspunkt, um weitere entwicklungsbiologische Systeme zu modellieren.

Contents

1

Introduction

The mystery of our own origin and mode of development has for over 100 years stimulated interest in vertebrate development. Fundamental processes of early development appear conserved among the vertebrates. By studying a particular vertebrate model organism we can therefore draw conclusions about vertebrate development in general and ultimately come up with models about our own development. (Haffter and Nüsslein-Volhard [39])

This statement concerning vertebrates can be extended to incorporate other eukarya. It reflects on the one hand the curiosity about the origin of life in general and on the other hand pronounces the interest in answering more immediate questions in smaller scale like how certain tissues or organs emerge and are maintained—events that are studied in the field of developmental biology. In detail, developmental biology is concerned with elucidating processes responsible for the intricate temporal and spatial control that takes place during the translation from genes to organisms.

Developmental systems represent a difficult body of research: processes need to be tracked with respect to both, time and space, and are taking place in a complex environment involving a multitude of genes and gene products regulating each others activity. For example, following the events resulting in the emergence of a flower in a plant necessitates to follow changes (time) in gene expression levels or respective gene product concentrations for all involved genes in all cells belonging to the considered tissue (space); a task that represents a major challenge for currently available experimental techniques [64, 97]. In consequence, available data tends to be of rather qualitative nature in form of a patterning concerning gene expression profiles and cell identities of respective tissues [35]. In addition, interactions between system components are usually non-linear which often prevents intuitive understanding of observed system dynamics [103] and complicates data interpretation.

In turn, mathematical modeling can assist in unraveling these gene regulative networks underlying system behavior given the available experimental data [64, 103]. For developmental systems, a common type of models are differential equation formalisms [80, 106] that provide the possibility to formalize hypotheses on system structure and behavior as well as they allow to test such hypotheses by providing a glimpse on the system behavior resulting from the initial hypotheses. Thereby, hypotheses can be directly compared to the often qualitative experimental results, allowing to validate and further refine working hypotheses as well as it can guide the design of new experiments, e.g., in the context of plant development [8, 29, 57],

or model organisms from the domain of animals like zebrafish [82, 108] or *Drosophila* [2, 107].

The considered scenario of developmental systems thereby differs from other scenarios in systems biology for which computational modeling is a common tool [22]: developmental systems are multi cell systems. In consequence, (i) detailed quantitative models tend to become to complex to remain computationally tractable due to the inherent parallelism, and (ii) rigorous experimental analysis of the systems is hampered by missing spatial resolution—resulting in availability of mostly qualitative data for these systems; two aspects that need to be taken into account during modeling. Still, the process of translating hypothesis into models is well investigated and researchers are assisted in this task by a range of formalisms [76, 93, 106]. Nevertheless, in order to use models for hypothesis validation, hypothesis refinement, or experimental design, the models first need to be calibrated to experimental data. In this regard, the task of model calibration for the considered type of systems remains a problem for which up to now there is no general solution: while parameter fitting for models using quantitative data is rather well explored [73, 77], as mentioned above for developmental systems often only qualitative data is available. In consequence, current approaches for model calibration under qualitative data usually are custom-tailored for certain systems and often involve tedious manual interactions [8, 57, 61, 88, 108].

1.1 · Research Complexes

In the context of modeling of gene regulative networks underlying developmental systems, this thesis is concerned with the following two topics that have been investigated simultaneously: (i) a methodology necessary for model calibration for this type of systems, and (ii) the application of mathematical models to investigate the development and maintenance of the shoot apical meristem in *Arabidopsis thaliana*. In detail, the investigated research complexes are:

1.1.1 · Parameter Estimation for Models in Developmental Biology under Qualitative Data

As pointed out above, model calibration for the systems considered in developmental biology is a problem for which up to now no general solution exists. Nevertheless, model parameters need to be calibrated before models can be used. The calibration task is difficult because for such systems mostly qualitative data is available making it hard to infer quantitative parameter values. In addition, the non-linearity of interactions between responsible genes further complicates extrapolation of suitable settings. For models in the literature, the following usually custom tailored approaches have been reported: (i) models are solved analytically to determine suitable parameter settings [61]—an approach that due to its complexity is only viable for small models. (ii) Based on analytical results for related systems, model parameters are tuned by hand [57, 108]. Again, a method only viable for small systems due to the exponentially growing parameter spaces. (iii) Heuristics are used to identify suitable parameter settings [8, 88]. Although such heuristics are promising to solve the parameter calibration task, they as well suffer from the exponential growth of a difficultly structured parameter space and from the problem that quantifying model fit with respect to a qualitatively defined target pattern remains unclear—a step that is necessary for automated parameter calibration. To come up with an effective and efficient heuristic for model calibration under qualitative data therefore comprises answering two key questions:

- how similarities between qualitatively defined, experimentally determined system behavior and model output can be measured and in what form this difference measure is made accessible for a suitable method for parameter optimization in general?
- Since available experimental data tends to be scarce with respect to time resolution, what further sources of knowledge could assist in the calibration process and how might they be incorporated? In this regard, further domain knowledge is supposed to facilitate accessibility of the parameter space structure as well as it could assist in reducing its overall size and thereby counter exponential growth.

1.1.2 · Modeling development and Maintenance of the Shoot Apical Meristem in *Arabidopsis thaliana*

The shoot apical meristem is a tissue that is responsible for the aerial growth of plants. To maintain growth it harbors a pool of stem cells, providing necessary undifferentiated cells for the growth. The key question in this context is what processes are responsible for establishing this pool and its homeostasis. A question, the answer of which can provide further insights into stem cell regulation in general and eventually might enable influence on aspects like crop sizes.

With respect to modeling the shoot apical meristem, in the literature there are two models already addressing related aspects, a model by Jönsson et al. [57] that is concerned with investigating regulation of the so called organizing center, a functional domain that is responsible for emergence and maintenance of the stem cell domain. The second model by Geier et al. [29] describes the stem cell domain and organizing center as cell pools that are connected via differentiation rates and expand due to cell proliferation, which is regarded as an externally controlled parameter. Using these models as a basis, they can be combined and extended with respect of adding explicit representation of the stem cell domain and its feedback on the organizing center and considering not only pool sizes but the spatial arrangement of the functional domains.

1.2 · Contributions and Overview

Before these two research complexes are addressed, this thesis provides an overview on techniques for modeling of biological systems with a focus on gene regulative networks (Chapter 2). In this context, the development of a spatial patterning with respect to gene expression profiles that are observed in developmental systems requires special attention. Therefore, in Chapter 3 a general mechanism for pattern formation in biological systems is introduced, namely reaction-diffusion systems. They are the most prominent approach to explain pattern formation and provide the basis for the modeling results presented later on.

Following these introductory chapters, in Chapter 4 fundamentals concerning model calibrations are presented. In this regard, the chapter contributes to answering the questions of (i) what criteria are suitable to guide an optimization method towards the identification of parameter settings that allow for the formation of a spatial patterning [46], (ii) how to present such criteria towards the optimization method [47], and (iii) how to quantify similarities between experimentally determined, qualitative tissue patterning and model outputs [49]. All three steps are investigated on basis of a framework for modeling of developmental systems and a general purpose optimization heuristic, namely an evolution strategy. Both, framework and evolution strategy are described alongside.

While the proposed techniques provide a valuable first step towards a solution of the parameter calibration task, it became apparent that an inclusion of further information sources is necessary in order to devise a reliable optimization strategy for this task. Therefore, Chapter 5 (i) investigates a method to generalize analytical results with respect to necessary conditions for pattern formation for the considered reaction-diffusion systems such that they can be incorporated in the optimization process [48], and (ii) a method is presented that allows to exploit the fact that developmental processes are organized in hierarchies of transient states.

In combination, these aspects form a method for parameter calibration for models of developmental systems using only qualitative data and its constituents are used to calibrate a model for development and homeostasis of the shoot apical meristem in *Arabidopsis thaliana* presented in Chapter 6, resulting in a robust model that is compliant with known control experiments. This model revealed that the used two combined feedback systems are sufficient to explain emergence and maintenance of the functional domains in the shoot apical meristem like the already mentioned pool of stem cells.

This thesis is concluded with a summary of the made contributions and a short discussion of implied next steps for future works with respect to both, model calibration for developmental systems in general and unraveling the

regulation underlying maintenance of shoot apical meristems in *Arabidopsis thaliana*.

2

Modeling Gene Regulative Networks

Mathematical modeling has becoming one of the key tools assisting biologists in interpreting data generated for complex systems, often studied on the level of gene regulative processes. This chapter provides a general overview on techniques for modeling gene regulative networks (GRNs). It focuses on systems in developmental biology which are at the heart of this thesis. For these systems the dynamics leading to the emergence of domains or clusters of cells showing similar gene expression profiles are studied—processes subsumed under the term 'pattern formation' (see Chapter 3).

This chapter first provides a brief introduction to the regulation of gene expression, followed by a list of commonly used modeling approaches. It is concluded with a presentation of the parameter optimization problem, one of the paramount challenges when building mathematical models for biological systems. This chapter thereby sets the stage for the work on parameter optimization for differential equation models in developmental systems (see

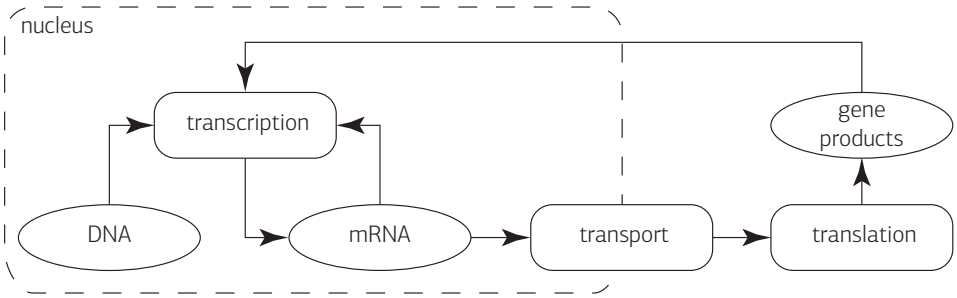

Figure 2.1. Schematic view on the process of gene expression for an eukaryotic cell. Rectangles symbolize actions during gene expression, ellipses represent biological entities in the expression process, and arrows indicate which entities have an influence on actions. The dashed line outlines the nucleus, dividing the cell in the two compartments, nucleus and cytosol.

Chapters 4 and 5) as well as the proposed model for maintenance of the shoot apical meristem in *Arabidopsis thaliana* presented in Chapter 6 of this thesis.

2.1 · Biological Background

The genome of an organism represents all its hereditary information coded in DNA or RNA. It is commonly referred to as blueprint of an organism, containing the necessary information for organism development and regulation. It can be thought of as a program, the control-flow of which is established on the level of genes, i.e., the basic units of hereditary information. The state of this program is determined by occurring gene expression profiles for a given time as well as resulting concentration levels of corresponding gene products. The regulation of this program takes place by modulating gene expression [1], i.e., the process from gene to gene product. This regulation can occur during all stages of gene expression, affecting gene transcription from DNA to mRNA, transcripts on their way to translation, or proteins resulting from translation (see Fig. 2.1).

Recent advances in experimental techniques thereby begin to allow gathering large-scale data concerning different stages of gene expression: the genome [15], the gene transcripts or transcriptome [62], the gene products in form of proteins or proteome [16, 21], or metabolites produced in biochemical reactions [23, 92]. Still, especially for developmental systems involving spatially confined domains with specific expression profiles, current high-throughput techniques lack the necessary spatial resolution. For developmental systems therefore often only qualitative data concerning gene regulative processes is available. Using either type of data, elucidating gene regulative processes remains difficult due to the overall complexity of the networks systems. In order to unravel the usually non-linear dependencies between genes, mathematical modeling can be of help.

2.2 · Modeling Approaches

Mathematical models provide compact descriptions of biological systems that allow to simulate and predict certain aspects of system structure or behavior. Modeling approaches can be data-driven, e.g., inference of GRNs from high-throughput data, or hypothesis-driven, e.g., modeling a hypothesis on putative interactions between biological identities and comparing modeling results with experimental observations. Keeping in mind the scarceness of quantitative data for developmental systems and therefore focusing on the latter case, such comparisons help to adapt understanding of the investigated developmental system and their underlying GRNs. In addition, such comparisons can guide the design of new experiments (see Fig. 2.2). This section provides a brief and by no means exhaustive overview on commonly used approaches for gene regulative processes, covering static, more descriptive methods as well as dynamic models. A more detailed review of existing methods can be found in de Jong [19].

2.2.1 · Graph Representations

The simplest model to represent GRNs is to use a directed graph $G = (V, E)$ consisting of a set of vertices V symbolizing the genes and a set of edges

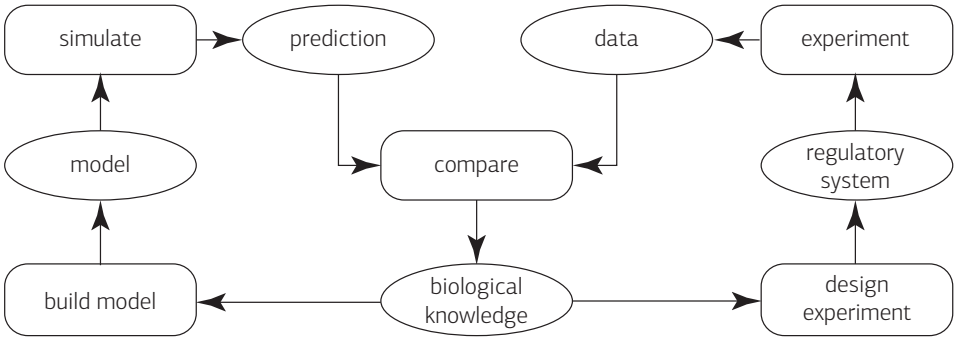

Figure 2.2. Interplay between experiments and modeling when investigating biological systems. Rectangles symbolize actions, ellipses represent information, and arrows indicate information flow.

$E \subseteq V \times V$ representing a regulating influence between genes. Each edge $e = (v_s, v_e) \in E$ is described by a tuple marking an originating vertex $v_i \in V$ and a target vertex $v_j \in V$. The tuples can be extended introducing further annotations, e.g., types of interactions like inhibition or activation. Graph representations are static models that are commonly used to visualize knowledge on regulations for considered systems but allow certain analyses as well: it can be tested if certain genes are connected, cycles in networks can be identified, or statistics concerning connectedness or general node degrees can be gathered. Thereby, regulative pathways, feedbacks necessary for homeostasis, or information on modules or network characteristics can be gained. In addition, graphs form the basis for more complex models such as state machines, Petri nets, or Bayesian networks.

2.2.2 · Bayesian Networks

Bayesian network models represent a statistical approach to give a compact representation of the joint probability distribution underlying, e.g., gene expression profiles. The considered GRNs are represented by directed acyclic graphs $G = (V, E)$ where V is a set of n biological entities, e.g., gene expression levels or gene product concentrations, represented by random variables

X_i with $1 \leq i \leq n$. The edges $e = (v_p, v_c) \in E$ represent regulative interactions between nodes and are expressed in terms of conditionals on respective nodes. The joint probability distribution underlying the expression profiles of a GRN is then given by:

$$p(X) = \prod_{i=1}^{n} p(X_i \,|parents(X_i)),$$

where $parents(X_i)$ corresponds to the nodes $v \in V$ with an incoming edge or regulating influence on node v_i. For Bayesian networks, the local Markov property is assumed, i.e., every vertex v is independent of any node apart from its immediate parents or descendants. Two graphs or Bayesian networks implying the same independences cannot be distinguished using observations of the modeled probability distribution and hence are equal [26].

Bayesian network approaches can be used to infer the structure of gene regulative networks given expression data and have the advantage that their statistical basis helps them to handle stochasticity and noise included in biological data [26]. In addition, they provide the possibility of being applicable to subsets or incomplete data. Nevertheless, Bayesian networks are in principle not designed to capture dynamics of gene regulative process but rather consider some sort of average behavior.

2.2.3 · Boolean and Generalized Logical Networks

Investigating dynamics of gene regulative networks, it becomes apparent that the dynamics in many cases describe transitions between different steady states. This leads to the notion that continuous system dynamics can be described by transitions between discrete system states where the transition rules are defined by logical functions [7, 58]. Following this idea, biological entities and their respective state can be expressed by Boolean variables, e.g., a gene can either be expressed (on or true) or not expressed (off or false), and the state of each entity is determined considering the states of a set of regulating entities following a Boolean function. Assuming a discrete notion of time, the next state of a network $X(t+1)$ containing n entities is determined evaluating the next state of each modeled entity

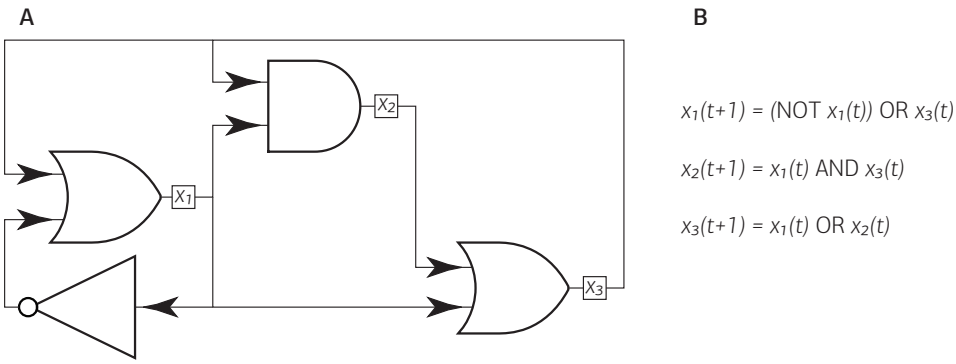

Figure 2.3. A: Boolean network consisting of three genes represented by a logic circuit. x_1, x_2, and x_3 are the Boolean variables representing the genes and the logical gates implement the regulating Boolean functions that determine the expression state of the respective genes. B: Boolean transition functions implied by the logical circuit shown.

$x_i\,(t+1)$ (with $1 \leq i \leq n$) in a synchronous fashion using their respective Boolean functions b_i and the current state of the network $X\,(t)$ as input. An example for a small Boolean network is shown in Fig. 2.3.

The discretization of the model variables brings the advantage that the state space of a given model becomes finite, e.g., in case of binary logic the state space of a model consisting of n entities has a cardinality of 2^n, resulting in the fact that the dynamics of a Boolean network always converges to a steady state or cycle and it is possible to determine all starting states that converge to a specific steady state—giving information on reachability of certain system states and partitioning the network in not necessarily disjunct modules involved in producing certain system behavior.

Still, the assumed discretization of states and time as well as the determinism with respect to transitions limits possible dynamics observable in Boolean networks. To address these limitations variants have been introduced, e.g., allowing to incorporate stochasticity in transitions [27] or allowing for still discrete but more than two states for model entities [102].

2.2.4 · Differential Equation Models

While Boolean networks discretize potentially continuous model entities like mRNA or gene product concentrations reducing them to discrete steady states, differential equation (DE) formalisms provides a way to follow GRN dynamics by modeling the time-dependent changes of concentrations of biological entities. The time development of concentrations for a model consisting of n modeled entities is modeled by rate equations:

$$\frac{\partial x_i}{\partial t} = f_i\left(x\right),$$

where x_i is the concentration of model entity i with $1 \leq i \leq n$ and $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a typically non-linear function computing the change in concentration x_i at time t taking into account the state of the model $x = [x_1, \ldots, x_n]'$ at time t. The advantages of DE models are that they are capable of representing continuous system dynamics, they allow to conveniently map biological reactions using the law of mass actions or formalisms like S-systems [93, 106] or the connectionist model [76]. In addition, DE models are highly flexible such that, e.g., constant delays in reactions, can easily be incorporated as well as they allow to take into account spatial influences of the simulated domain: instead of only time-dependent ordinary differential equation (ODE) models one then gets a system of partial differential equations (PDEs) depending on time and space. Here, the spatial component is especially important when considering gene regulative processes in developmental systems since there the expression profile of a cell in a considered tissues not just depends on itself but is influenced by external signals from neighboring cells as well.

Although DE systems are extremely versatile, this comes at the cost that analyzing them often turns out to be extremely difficult. The size of the models and the involved non-linear terms prevent analytical solutions for most systems, e.g., by means of linear stability or bifurcation analysis [101]. In order to still obtain knowledge on possible numbers and stability of steady states or robustness, two possible routes can be followed: (i) facilitating analysis by simplifying non-linear reaction equations, e.g., by approximating often occurring sigmoidal response curves by piece-wise linear equations, or

(ii) instead of analytical characterization, models can be solved numerically. Thereby, system behavior is investigated for specific simulated conditions giving an idea on system behavior for certain cases, but leaving open overall system behavior for changing conditions.

In addition, using DE formalisms one assumes continuous concentrations and that the modeled reactions take place in a well mixed compartment where each molecules can freely interact with each other, assumptions that in certain cases do not hold: when the number of reacting molecules becomes small, concentrations no longer are continuous but become discrete and reactions no longer can be reliably approximated by a deterministic scheme imposed by using DE models but one needs to resort to stochastic simulations where molecules are explicitly represented and reactions between molecules are treated as stochastic processes.

2.2.5 · Stochastic Models

In order to circumvent problems stemming from the assumption of continuity of concentrations as well as the determinism assumed for occurring reactions, stochastic models explicitly considering reacting molecules can be used. In difference to DE models, as state variables instead of concentrations discrete numbers of molecules X are used which are accompanied by a probability distribution $p(X,t)$ quantifying the probability of occurrence of these molecule numbers at a point in time t. The probability of being in a certain state X at a certain time point $t+\delta t$ is then given as [33]:

$$p(X;t+\delta t) = p(X;t)\left(1-\sum_{j=1}^{m}\alpha_j\delta t\right)+\sum_{j=1}^{m}B_i\delta t,$$

where the α_j are the probabilities that one of the m reaction takes place such that the system leaves state X and the B_i are the probabilities that the system is in another state but enters state X by means of one of the m

reactions taking place. Both probabilities are quantified with the time step δt. This can be reformulated into the master equation [33]:

$$\frac{\partial}{\partial t} p\left(X;t\right) = \sum_{j=1}^{m} \left(B_j - \alpha_j p\left(X;t\right)\right).$$

While the system description in terms of master equations provides an elegant way to describe system dynamics, they are even more difficult to solve analytically or numerically than DE models. Thereby further restrictions are imposed with respect to tractable system sizes. To ease this problem at least two variants exist: (i) by converting the stochastic system into a set of ODEs containing a stochastic term [34], a method that is only feasible for certain systems; (ii) by explicitly simulating a number of stochastic trajectories modeling the system behavior and thereby reproducing an empirical probability distribution underlying the system dynamics [33].

In addition to the increase in necessary computation time to solve such models, this type of model requires detailed and complete knowledge of reactions taking place in the considered regulatory systems. In contrast, for many of the complex systems considered in developmental biology not all involved factors are known by now but need to be approximated by supplementary factors implementing putative interactions, therefore rendering stochastic methods as hard to apply.

2.2.6 · Discussion

While static models are well suited to infer or visualize the general structure of GRNs, the key aspect of developmental systems are their dynamics leading to a sequence of transient stages these systems run through. Therefore, dynamical modeling techniques are necessary to capture these processes. In turn, structural information necessary to build dynamic models could be gathered using static models.

Considering dynamical models, the presented approaches Boolean networks, DE models, and stochastic models provide a trade-off between degree of

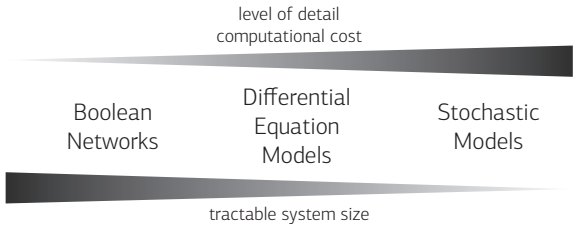

Figure 2.4 Trade-offs between levels of modeling and tractable system sizes for dynamical modeling approaches.

complexity representable by a given model on the one hand side and computational cost as well as necessary degree of structural information available beforehand on the other side (see Fig. 2.4). For models of developmental systems DE systems are the most common approach.

Another aspect of modeling not discussed so far is model parameters that need to be set encompassing two types of parameters: (i) method specific parameters like the choice of priors for Bayesian approaches, or (ii) or biological system specific parameters like thresholds used for discretization in Boolean networks or biological constants like rate constants necessary for DE or stochastic models. Thereby, especially the choice of suitable biological constants is often difficult—a topic that is discussed in the next section.

2.3 · Challenges in Modeling Dynamic Systems: Parameter Optimization of Model Parameters

Using mathematical modeling to assist research on biological systems involves a three step process: (i) the model is set up, a processes that in many case is straight forward or makes use of established formalisms, e.g., the use of the law of mass action to translate bio-chemical reactions into DEs. (ii) Model parameters need to be set, either in the form of method specific parameters or system specific parameters. Typical method specific parameters are, e.g., priors used in Bayesian approaches, while system specific parameters are, e.g., rate constants necessary to set up DE or stochastic models (see Fig. 2.5). (iii) The models are solved or simulated to generate data that can be compared to experimental observations. In addition, for such

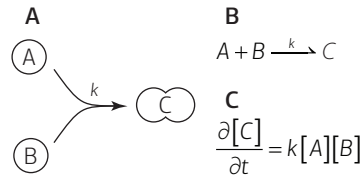

Figure 2.5 A: Complex formation of two molecules A and B to C with unknown rate constant *k*. B: Corresponding bio-chemical rate equation. C: Resulting time-dependent DE modeling concentration changes of the complex C.

comparisons metrics quantifying the degree of similarity between model and experiments need to be defined.

While the first step of setting up a suitable model is already well-studied, this section focuses on the step of tuning model parameters and in its course addresses the task of quantifying degrees of similarity between model and experiments mentioned under step three. In addition, since DE models are the best represented modeling approach to investigate processes in developmental systems and for DE models tuning of method specific parameters is well-understood, only parameters belonging to the modeled biological system are considered here.

As mentioned above such parameters involve biological constants like rate constants, dissociation constant, or reaction orders. In an ideal case, these parameters would be measured directly in experiments but in most cases experimentalists have no direct access to these factors. Still, for scenarios where high resolution quantitative data is available there exists a range of heuristics for inference of parameter settings from quantitative time course data [73, 77]. Unfortunately, for systems in developmental biology mostly only quantitative data concerning reached stable system states is available, rendering use of the former methods impossible.

In cases where only qualitative data is available, fitting or optimizing the parameters to match experimental observations becomes more difficult. For small models it is possible to solve them analytical and thereby information is gained in what parameter ranges the considered model shows a certain qualitative behavior [61, 80, 101]. In addition, such information can to some extend be transferred to bigger systems: serving as an initial guess, parameter settings are adapted by hand [57, 108]. For even larger systems or when different hypothesis are to be tested, tuning by hand becomes infeasible. In

this case, heuristics can be used [46–48, 88, 107] but the spatial component and non-linear dependences involved in developmental processes complicate this task. In addition, where for manual techniques the fit between model an experiments can be done by manual inspection, for heuristics this comparison needs to be formalized and quantified. This quantification is often achieved by using thresholds to discretize model outputs in order to get a qualitative description. This discretized model output can then be compared to qualitative experimental data on a per-entity base, e.g., qualitative gene expression profiles can be compared on a per-gene basis and the dissimilarity is then expressed in numbers of occurrences of differences [49, 107].

Formally, a parameter setting $x \in X$ minimizing the difference between model output g and experimentally determined qualitatively defined patterning p^{tar} is sought using a function h quantifying their degree of dissimilarity:

$$\underset{x \in X}{\operatorname{argmin}} \left\{ h(g\,(x)\,, p^{tar}) \right\}, \tag{2.1}$$

with $X \subseteq \mathbb{R}^n$ being the n dimensional model parameter space of in case of DE models usually real parameters. The topic of parameter optimization of model parameters is further discussed in the Chapters 4 and 5.

3

Models for Pattern Formation in Biological Systems

Fundamental aspects when studying developmental systems are the establishment of a spatial distribution of cells with different cell types, the morphogenesis ultimately resulting in the development of different tissues and organs as well as its homeostasis. In terms of gene expression, both aspects can be described by development or maintenance of patterns consisting of groups of cells exhibiting similar expression profiles. While detailed mechanisms underlying such patterns for most systems remain unknown, there exists a range of putative mechanisms that in principle could account for various phenomena in developmental processes. Agreeing on the fact that information on organisms stored in DNA is in principle identical in all cells of a single organism, information guiding pattern formation originates from the communication between cells.

In the following, one of these pattern formation mechanisms is described in detail, the so called reaction-diffusion systems. In addition, since the

resulting models tend to be to complex to be solved analytically, a brief overview is given on methods to numerically solve this type of models.

3.1 · Reaction Diffusion Systems

The idea of reaction-diffusion (RD) systems date back to Turing [105] who first proposed that it would be possible to generate timely stable and spatially heterogeneous patterns based on the interactions between to diffusing chemicals or morphohgens. Guided by experimental observations, he showed mathematically that spatial information exchange by diffusion can account for the generation of spatial heterogeneous patterns in systems that without diffusion remain spatially homogeneous. The term spatial domain hereby refers to, e.g., tissues represented by arrays of cells, and spatial heterogeneous patterns on these domains are then, e.g., domains of cells showing distinct gene expression patterns like they are observed in developmental systems. RD systems thereby provide a versatile mechanism for pattern formation [32, 51, 57, 61, 71, 82, 108].

In a general form RD systems are commonly described by differential equations (DEs) of the following form:

$$\frac{\partial \mathbf{c}}{\partial t} = D\Delta\mathbf{c} + \mathbf{f}(\mathbf{c}), \tag{3.1}$$

where $\mathbf{c}$ is a vector representing the concentrations of all considered species at time t, D is a diagonal matrix of positive diffusion constants, Δ is the Laplace operator, and $\mathbf{f}$ are the non-linear reaction kinetics. Here, DE formalisms represent a natural choice since they can conveniently be derived from underlying bio-chemical reactions using the law of mass action and they provide a suitable level of detail while remaining computationally tractable.

While diffusion processes commonly are assumed to cancel out spatial differences in concentrations, under certain conditions inclusion of diffusion can destabilize systems that otherwise are in a stable steady state—such instabilities are called diffusion-driven or Turing type instabilities. Analytical solutions to various different RD systems have shown that especially

the relation between the different diffusion constants in D is crucial for occurrence of Turing patterns as well as the interactions between the involved species [61, 80, 105]: there needs to be a difference between the involved diffusion constants and the described kinetics need to adhere to some notion of activation-inhibition interpretation, subsumed by the principle of short-range activation and long-range inhibition [31, 69, 95]. Thereby, the spatial component together with boundary conditions imposed on domain boundaries plays a crucial role in explaining the emergence of patterns.

In the most simple case, we consider a system consisting of two species in a spatial domain. The first species a is supposed to be autocatalytic and actively consuming or degrading the second species s which is constantly produced. In this scenario, a is slowly diffusing while s is fast diffusing. Starting from a small heterogeneity in the concentration distribution of a over the spatial domain, the autocatalytic properties of a allow for the generation of a peak in the a concentration distribution. Due to its slow diffusion, this peak remains locally and during its formation and maintenance it negatively influences the concentration levels of s. Since s is fast diffusing, the concentration of s is not only reduced locally but in a wider area defined by the absolute magnitude of its diffusion constant. Thereby, around the a peak, a field is created where s concentrations are too low to support another a peak. This implementation of the principle of short-range activation and long-range inhibition is sketched in Fig. 3.1 and in result, such a system creates a stable pattern of concentration peaks of a as shown in Fig. 3.2. This so called activator substrate system was proposed by Gierer and Meinhardt [31] and its behavior with respect to time t is described by the following equations [61]:

$$\frac{\partial a}{\partial t} = D_a \Delta a + \rho_a a^2 s - \mu_a, \tag{3.2}$$

$$\frac{\partial s}{\partial t} = D_s \Delta s - \rho_s a^2 s + \sigma_s, \tag{3.3}$$

with the diffusions constants D_a, D_s; a spatial dependency represented by the Laplace operator Δ; the reaction coefficients ρ_a, ρ_s; the degradation

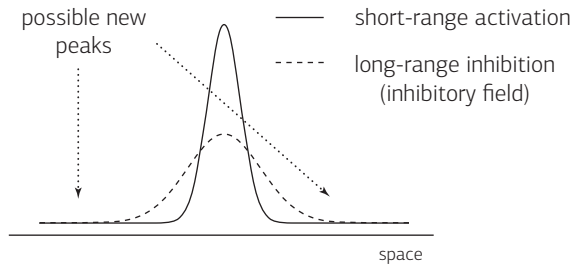

Figure 3.1 Schematic view on the principle of short-range activation and long-range inhibition.

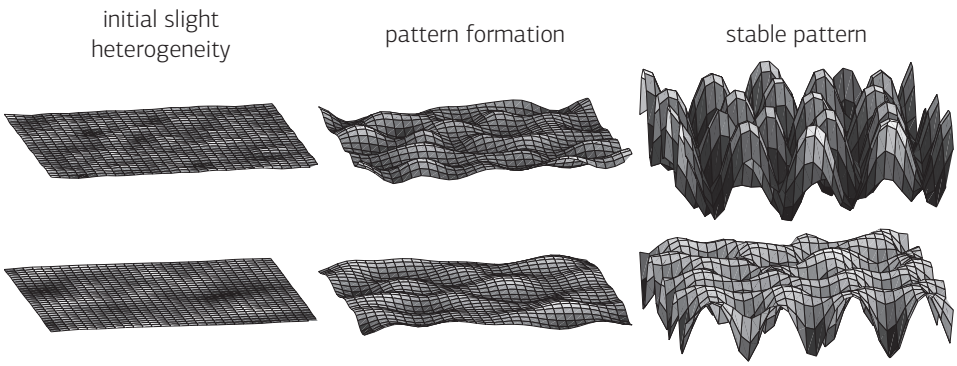

Figure 3.2. Time series documenting the generation of a Turing pattern for an activator substrate system. The top row shows the concentration distribution of a (cf. Eq. 3.2) for three discrete time points and the bottom row shows the corresponding concentration distribution of s (cf. Eq. 3.3). As spatial domain a two-dimensional field with 30×30 inner points and periodic boundary conditions is used.

rate μ_a; a constant production rate σ_s. An analytical solution to a simplified version of this system documenting conditions under which Turing instabilities occur can be found in Koch and Meinhardt [61].

In addition to the presented activator substrate system, there exists a wide range of different RD systems allowing to generate various patterns but at this point only those reaction diffusion systems are going to be presented which are used in this thesis; more complete overviews on different types of systems and their respective biological applications can be found in Koch and Meinhardt [61] and Meinhardt [70].

A system in concept much alike to the activator substrate system is the so called Brusselator [86] given by the following equations:

$$\frac{\partial x}{\partial t} = D_x \Delta x + a - (b + \beta)\, x + c x^2 y, \tag{3.4}$$

$$\frac{\partial y}{\partial t} = D_y \Delta y - c x^2 y + b x. \tag{3.5}$$

Similar to the activator substrate system, x is autocatalytic and consumes or degrades y during its generation. In effect, patterns generated by these two systems therefore show local minima in y and s concentration where peaks in x and a are formed and in-between these peaks, x and a reach local minima while y and s concentrations recover (see Fig. 3.2). In difference to the activator substrate system that incorporates just the two species a and s, in addition to x an y the Brusselator comprises additional species a, b, β, and c which are assumed to remain at constant concentrations throughout the time course of the system.

A conceptually different type of system is the activator inhibitor system [31] described by the following equations:

$$\frac{\partial a}{\partial t} = D_a \Delta a + \rho_a \frac{a^2}{h} - \mu_a,$$

$$\frac{\partial h}{\partial t} = D_h \Delta h - \rho_h a^2 - \mu_h,$$

where D_a, D_h, ρ_a, ρ_h, μ_a, and μ_h have the same meaning as in Eqs. 3.2 and 3.3. In difference to a species that is consumed or degraded, here, the long-range inhibition is implemented via an inhibitor h which is generated together with the autocatalytic species a. Wherever a levels start to rise h levels rise as well. Since h diffuses much faster than a, around local maxima in a levels a region is generated into which h is diffusing and that in result shows high h levels as well. In effect, in this region h levels are high enough to prevent new a peaks and thereby resulting in a stable spatial pattern (cf. Fig. 3.1).

While so far only general pattern formation capabilities have been discussed, considering spatial domain sizes and growth adds new aspects to biological

pattern formation with RD systems. First of all, for a given RD system there exists a critical domain size necessary for initiation of patterns. In consequence, for example in developmental systems patterns emerge once the system reaches a certain size, i.e., after a certain number of cell divisions; a fact fitting observations in embryo development. In addition, patterns generated by RD systems show regenerative properties, e.g., when transplanting regions of the spatial domain, as well as the formed patterns can adapt to growing domains by, e.g., in case of the activator substrate or activator inhibitor system by splitting existing a peaks or introducing new a peaks respectively—a fact that could assist in explaining phyllotaxis, i.e., the arrangement of organs in plants.

3.2 · Simulation By Numerical Integration

Although models for biological systems usually make use of a range of simplifying abstractions in order to allow capturing certain processes of interest, the resulting models tend to remain complex. For example RD models for pattern formation in developmental biology are to complex to be solved analytically, demanding for numerical solutions. Where analytical characterizations of models yield information on, e.g., what types of patterns a certain system is capable of producing for what parameter settings, numerical solutions or simulations follow a different route. Either parameter settings are sought such that system behavior exerted during numerical simulation best resembles experimentally observed behavior or small variations are imposed on such a calibrated model in order to predict what influence this change might have on overall system behavior. In difference to rigorous analytical analyses, numerical solutions therefore highlight certain possible system behaviors but in turn provide a full trajectory of the system behavior under the conditions of the simulation.

Since in RD systems a spatial component needs to be considered in addition to time, the task of numerically solving the system is a boundary value problem where initial values are given for several spatial loci. Due to the spatial component, the systems are described by partial differential

equations (PDEs) which typically are parabolic. In order to solve these PDEs numerically, information on derivatives at a certain point with respect to time and space are used to extrapolate future system states. In effect, the continuous behavioral system description is discretized with respect to both, time and space, to compute stepwise approximations. Depending on the characteristics of the considered PDEs, this process can be more or less demanding but especially the diffusion terms in RD systems tend to be stiff, meaning colloquial speaking that the curvature of the PDEs undergoes drastic changes. These changes in curvature result in stability problems for some of the simpler but fast methods for numerical integration and therefore demand for sophisticated but computationally expensive methods. Since computation time necessary for numerically solving the considered systems reaches orders of magnitude in the range of minutes, the computation time necessary for parameter optimization (see Section 2.3) needs to be considered when choosing a numerical integrator.

In the following we give a brief overview on some of the more common methods for numerical integration that have been used in this thesis, beginning with fast but sometimes unstable explicit methods, followed by stable but computationally demanding implicit schemes, and concluding with a combination of both, so called implicit-explicit schemes. More detailed reviews can be found in Ruuth [91] and Press et al. [85].

3.2.1 · Explicit Schemes

Explicit numerical schemes make use of information of the derivatives up to the current time point to extrapolate the system state at the next time point. The set of explicit integrators can be divided into to different approaches, (i) those which only take into account the current time point and might calculate intermediate steps between the current time point and the next point and (ii) and the so called multi step methods which in addition to the current time point make use of earlier time steps as well. In addition to time discretization, the spatial component in the PDEs needs to be addressed as well; for explicit schemes the Laplace operator Δ for a function $a(x,t)$ at time t is commonly discretized using a spatial grid. For example, a

continuous one-dimensional domain x is divided into n equidistant point x_i with $1 \leq i \leq n$ with mesh spacing δx. For each of these points, the Laplace operator can then be discretized as follows:

$$\Delta\left(x_{i,j}, t\right) = \frac{a\left(x_{i-1}, t\right) + a\left(x_{i+1}, t\right) - 2a\left(x_i, t\right)}{\left(\delta x\right)^2}, \tag{3.6}$$

a strategy that can be extended in a straight forward manner for different grids as well as higher dimensional spaces.

Runge-Kutta Method

One of the most common explicit integrator schemes is the Runge-Kutta fourth order method [85], belonging to the family of Runge-Kutta schemes. It has an accumulated error in the order of h^4 where h is the step size for time discretization and it uses the latest system state and additional three intermediate steps to calculate the next system state. It is given by:

$$\begin{aligned}
y_{n+1} &= y_n + \frac{1}{6}h\left(k_1 + 2k_2 + 2k_3 + k_4\right), \\
t_{n+1} &= t_n + h, \\
k_1 &= f\left(t_n, y_n\right), \\
k_2 &= f\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_1\right), \\
k_3 &= f\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_2\right), \\
k_4 &= f\left(t_n + h, y_n + hk_3\right),
\end{aligned}$$

where $f\left(t, y\right)$ is the DE; y_n and y_{n+1} are the current and the next system state respectively.

Adams-Bashford

The Adams-Bashford integrator [85] is a simple linear multi-step method that computes the next system state using information on a set of previous system states. In this thesis the two step Adams-Bashford scheme is used which is given by the following equation:

$$y_{n+1} = y_n + \frac{3}{2}hf\left(t_n, y_n\right) - \frac{1}{2}hf\left(t_{n-1}, y_{n-1}\right),$$

where $f(t, y)$ is the DE; y_{n+1}, y_n and y_{n-1} are the next system state, the current system state and the previous system state respectively.

3.2.2 · Implicit Schemes

In difference to the already described explicit schemes, implicit integrator make use of an estimate of the next system state by solving the following equation:

$$g(y_n, y_{n+1}) = 0,$$

where y_n is the current system state and y_{n+1} is the next system state. Although solving this equation introduces some further computational overhead, especially for stiff systems implicit methods often yield much better results.

In this thesis a variant of the Crank-Nicolson scheme [38] is used, which is designed to solve PDE systems and we focus on the spatial discretization step used in this method. It is given be the following equation:

$$\frac{y_{n+1}^i - y_n^i}{h} = \frac{\left(y_{n+1}^{i-1} + y_{n+1}^{i+1} - 2y_{n+1}^i\right) + \left(y_n^{i-1} + y_n^{i+1} - 2y_n^i\right)}{2(\delta x)^2},$$

where i is the index for a regular spatial grid on a one dimensional system and $i-1$ and $i+1$ are the neighboring indices; h is the time step; y_n and y_{n+1} are the current system state and the next system state respectively; and δx is the grid spacing of the spatial grid.

3.2.3 · Implicit Explicit Scheme

As mentioned above, numerically solving PDE systems originating from RD models requires a trade-off between computational cost and accuracy. Since in RD systems only the diffusion parts tend to be stiff, it can be advantageous to use implicit-explicit (IMEX) schemes that use a fast explicit scheme for the reaction part of the PDEs while the stiff diffusion terms are solved using more accurate implicit schemes. Thereby one has to take care that the

chosen time-stepping method is at least second order and oscillating error components stemming from the diffusion terms are effectively dampened; otherwise the integration process might converge to a stable system state that is qualitatively wrong [28]. One method fulfilling these requirements is an IMEX scheme based on a two step Adams-Bashford integrator combined with a variant of the Crank-Nicolson scheme [91]. It is given by the following equation:

$$\begin{aligned}\frac{y_{n+1} - y_n}{h} =& \frac{3}{2} f\left(y_n\right) - \frac{1}{2} f\left(y_{n-1}\right) \\ &+ D\left(\frac{9}{16}\Delta y_{n+1} + \frac{3}{8}\Delta y_n + \frac{1}{16}\Delta y_{n-1}\right),\end{aligned}$$

where y_n and y_{n+1} are the current and the next system state respectively; h is the time step; $f\left(y\right)$ is the reaction term from Eq. 3.1; D is the diffusion constant from Eq. 3.1; and the Laplace operators Δ are discretized following Eq. 3.6.

3.2.4 · Discussion

In conclusion, in this section a range of different methods for numerical solution or simulation of DE models have been presented that constitute a trade-off between computational cost and accuracy. While for less complex systems fast explicit schemes are sufficient for numerical integration, for more complex systems like the proposed shoot apical meristem (SAM) model (see Chapter 6) more sophisticated schemes need to be used. Still, since numerical solutions only provide a local view on possible system behavior, especially during parameter optimization or calibration a significant number of simulations is necessary. Therefore, computational costs should be reduced where possible: the use of expensive implicit schemes should be avoided, especially for non-linear terms. In consequence, IMEX schemes are used that provide the necessary accuracy for the stiff diffusion terms applying implicit schemes while the non-linear reaction terms are solved using faster explicit schemes.

4

Methods for Model Calibration: General Considerations

The common modeling technique for developmental systems and their underlying gene regulative networks (GRNs) are differential equation formalisms. They allow for a detailed mapping of bio-chemical reactions constituting gene regulation as well as coarser grained subsumptions of subsystems. Simulations of such models yield information on time-behavior of modeled entities like concentrations of gene products. In order to use such information to validate hypothesis or make predictions concerning the underlying regulation, the parameterized differential equations constituting the model need first to be calibrated using experimental data.

Such calibration processes are difficult for the considered systems. Interactions between components are non-linear and therefore appropriate parameter settings are hard to infer. In addition, the behavior of developmental systems depends on time as well as interactions between different cells of a tissue play a crucial role. This circumstance adds a spatial component to

the already time dependent differential equations. On the other hand, spatially high-resolution quantitative data for tissue or organism-level systems is difficult to obtain; experimental data concerning such systems are mostly of qualitative nature originating from several imaging techniques. For example by using fluorescence microscopy in combination with genes tagged with fluorescent proteins, expression domains of respective genes can be visualized. Such information describe a patterning of the considered tissue with respect to gene expression levels.

To address the challenges posed by the task of model calibration for the given context of developmental systems it makes sense to divide it into a set of different subtask. (i) A modeling framework as well as a general method for minimization of differences between model output and experimental results has to be chosen. The implementations used in this thesis are presented in Section 4.1. (ii) Building on this framework, before model calibration with respect to specific target patterns is pursued, the identification of parameter settings allowing for the emergence of general patterns has to be considered. Using reaction-diffusion (RD) systems as pattern forming mechanism, already this more general task is challenging: the desired spatially heterogeneous and timely stable patterns are only formed for small fractions of the parameter space and the parameter spaces themselves are dominated by plateaus of parameter settings showing qualitatively similar behavior. In consequence, these plateaus obscure information allowing a directed calibration process. This challenge is addressed in Section 4.3 where features of optimization criteria are explored that allow to escape the mentioned plateaus. Building on this work, two further aspects need to be considered: (iii) matching parameter settings not only need to be identified but at the same time system properties like robustness or sensitivity are of interest. In order to assess such characteristics, in Section 4.4 it is investigated in how far optimization criteria can be reformulated to aid an exploration of the considered parameter spaces with respect to such characteristics. Finally, (iv) to allow the identification of parameter settings resulting in a specific patterning rather than an arbitrary patterning, methods for pattern identification in model outputs and quantification of their degree of similarity to target patterns are discussed in Section 4.5.

4.1 · Foundations

In the following a modeling framework for GRNs in multi cell systems and a randomized optimization strategy are described that are used as basis for the presented work on model calibration.

4.1.1 · Framework for Simulations of Developmental Systems

The used modeling framework serves the two goals to allow for (i) hypothesis testing and (ii) hypothesis exploration. On the one hand, it should be testable whether a given system of interacting factors can form certain spatial patterns by finding the necessary parameter settings and simulating the system. On the other hand, based on a calibrated model predictions on a possible patterning resulting from novel intracellular and intercellular interactions shall be possible.

The proposed model is defined by the following core components:

Cells: The model consists of spatially discrete units, the cells. They are used as autonomous units. It is assumed that all cells are similar to each other in design, in particular regarding the underlying GRN, and only differ in their states. Cell volumes as well as shared surface areas are assumed to be of equal magnitude for all cells.

Gene products: The state of a cell v_i is characterized by the concentrations of the gene products produced in the cell. The gene product concentrations are represented by a real valued vector. The term 'gene product' in this case not only refers to the product but has to be understood synonymous for gene products, gene expression levels and all processes on the way from gene to gene product. Since there exists a mapping between expression levels and the resulting amount of gene products, the gene product concentrations are representing the gene expression levels.

Cell structure: The cells are grouped according to a spatial neighborhood defining which cells share common cell surface areas. Internally the cell neighborhood is represented by a graph $G(V, E)$ consisting of a set of cells V. Contacts or interaction pathways between the cells are represented by edges $e = (v_i, v_j) \in E$ between two cells v_i and v_j.

Cell communication: To form spatial heterogeneous patterns, spatial interactions or cell communication between the constituting components are mandatory. In this model diffusive interactions are considered as means of communication and it takes place along the edges between the cells. Therefore implicitly zero flux boundary conditions are used on the boundaries of the considered spatial domain.

Using the framework, system dynamics are explored by simulating the resulting model. During simulation the states of the cells change according to (i) intracellular interaction between genes or gene products and (ii) the intercellular communication via diffusion. In a formal description the state change of a cell v_i follows a transition function $\delta\left(q_i, N\left(v_i\right)\right)$ depending on the current state q_i of the cell and the states of its interaction partners given by the neighboring cells $N\left(v_i\right)$. Each iteration in the simulation corresponds to calculating the transitions made for every cell based on the status quo. As discussed in Chapters 2 and 3, for the considered task of modeling pattern formation and maintenance in developmental systems differential equation (DE) models based on RD systems are used.

4.2 · Method for Model Parameter Optimization

Repeating the formulation of the task of calibrating a model for the emergence and maintenance of a patterning in developmental systems given in Eq. 2.1 in Chapter 2, for model calibration a parameter setting $x \in X$ minimizing the difference between model output g and experimentally determined qualitatively defined patterning p^{tar} is sought using a function h quantifying their degree of dissimilarity:

$$\underset{x \in X}{\operatorname{argmin}} \left\{ h(g\left(x\right), p^{tar}) \right\}, \tag{4.1}$$

with $X \subseteq \mathbb{R}^n$ being the n dimensional model parameter space of usually real parameters.

For this minimization an evolution strategies [25] is used in this thesis. Evolution strategies belong to the class of randomized optimization algorithms

designed for real-valued optimization like the parameter spaces $X \subseteq \mathbb{R}^n$ considered for differential equation models in developmental biology.

In short, evolution strategies map principles of Darwinian evolution to form an optimization cycle (see Fig. 4.1): beginning with an initial guess on an appropriate search distribution p_X on the considered search space X (here the model parameter space) the search space is sampled λ_{SO} times and the samples x_i with $1 \leq \lambda_{SO}$ are evaluated using an objective or fitness function h (here the function quantifying the dissimilarity between model output and experimentally determined tissue patterning). Based on their respective objective or fitness scores, a subset of the λ_{SO} so called offspring vectors is selected to update the search distribution p_X. These steps are then repeated until a certain stopping criterion is met. Stopping criteria that are commonly used are, e.g., a bound on the number of objective function evaluations or a fitness threshold which once exceeded by one of the sample vectors terminates the optimization process. As search distribution, commonly normal distributions $\mathcal{N}(m, s^2C)$ are used, with distribution mean m, a covariance matrix C, and a scaling factor s.

In detail, the CMA-ES designed by Hansen and Ostermeier [44] is used. In each iteration of its optimization cycle it uses information on the selected x_i to infer dependencies between the different dimensions of X. Thereby, it adapts the shape of C. In addition, the scaling factor s is updated. The offspring are drawn according to:

$$x_i \sim m + s\mathcal{N}(0, C).$$

Both, C and s are updated in such a way that the likelihood of sampling a point x^* that is better in terms of the fitness function than the previously drawn samples is increasing.

The CMA-ES shows a set of invariance properties that in effect make it a robust optimization method: (i) since the selection of the λ_{SO} offspring during each cycle or generation is done using a ranking of the sampled points x_i with respect to h, CMA-ES is invariant to order preserving transformations on h; (ii) adaption of C is capable of inferring rotations with respect

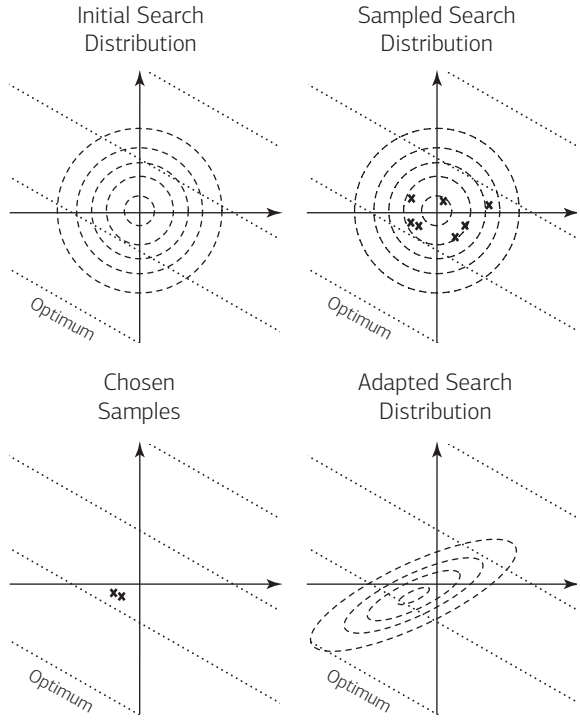

Figure 4.1 Iterative optimization cycle of the CMA-ES. To identify a setting solving the considered optimization task, a search distribution is first sampled; from the samples, the best candidates in terms of the objective are selected; using the selected samples, the search distribution is adapted. Here, the dotted lines are lines representing equal objective values and the optimum of the sketched problem is located in the bottom left corner. The dashed lines represent lines of equal probability density to sample a vector for the considered search distributions.

to search space axes, resulting in an invariance to rotations; (iii) adaption of s allows for some invariance to overall scaling of the parameter space, given that initial C and s are chosen appropriately. In combination, these properties make CMA-ES a search heuristic that has proven its value in a range of real valued optimization tasks [3, 46, 49, 77, 87].

In the described form the CMA-ES is specialized to deal with optimization tasks where only a single objective is considered. On the other hand, in the context of model calibration often a set of objectives has to be considered, e.g., differences between simulated patterning and experimentally determined patterning with respect to a number of genes. Therefore, strategies to deal with multiple objectives need to be considered. In this case, the solution to the problem stated in Eq. 4.1 becomes the set:

$$\left\{x \in X | \nexists\, x' \in X : x' \preceq x \wedge x \npreceq x'\right\}, \tag{4.2}$$

where x, x' are candidate solutions in the parameter space $X \subseteq \mathbb{R}^n$ and '$\preceq$' denotes the weak Pareto-dominance relation with respect to the vector valued function $\mathbf{h}(x)$. For two vectors $x, x' \in \mathbb{R}^n$, x is said to weakly dominate x' ($x \preceq x'$) with respect to $\mathbf{h}(x) = (h_1(x), \ldots, h_m(x))$ iff $h_i(x) \leq h_i(x')\ \forall\ i \in \{1, \ldots, m\}$. Thereby, Eq. 4.2 represents the set of all optimal solutions or Pareto-set of the multiobjective optimization problem; the image of the Pareto-set in the objective space is called the Pareto-front.

Commonly, there are at least two different approaches to identify solutions in this set: (i) methods where all objective functions are aggregated into a single objective, and (ii) methods that consider all objectives simultaneously. With respect to (i) possible aggregations are weighted sum approaches or scalarizations like Chebyshev scalarization [74]. With respect to (ii) multiple objectives are considered simultaneously by comparing different candidate solutions using indicators based on the concept of Pareto dominance [20].

In comparison, approaches explicitly considering all objectives simultaneously have the advantage that they can identify a set of solutions contained in the set defined by Eq. 4.2 while aggregation approaches tend to show less diversity. Although in principle a single fitting parameter setting might be enough to solve the task of model calibration, a set of settings can provide further information with respect to, e.g., robustness of the proposed model. In addition, for a range of different optimization tasks studies have shown that considering multiobjective formulations can have a beneficial restructuring effect on the problem landscape, facilitating the optimization process [14, 41–43, 53, 59, 83].

Nevertheless, singleobjective approaches have the advantage that they impose a total order on the parameter space with respect to the optimization criterion which simplifies selection processes during optimization. Approaches based on the concept of Pareto dominance in turn only impose a partial order, introducing possible ambiguities during optimization. To resolve this matter, state of the art multiobjective approaches use refinements of the Pareto dominance resolving these ambiguities and establishing a total order, e.g., refinements like the hypervolume [110–112]. For hypervolume calculation, a reference point is introduced in the objective space and the

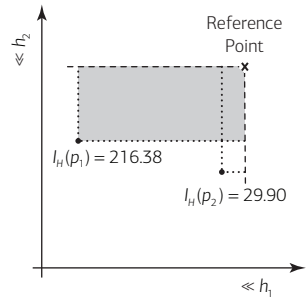

Figure 4.2 Hypervolume indicator for a minimization problem with two objectives. The hypervolume for a point p in the objective space is defined by the volume of the hyper-cube spanned by the point and a reference point. The hypervolume spanned by the point p_1 is shown in gray. During selection, the hypervolume indicator values I_H can be used to build a ranking on different candidate vectors that is a total order and compliant with the concept of Pareto dominance.

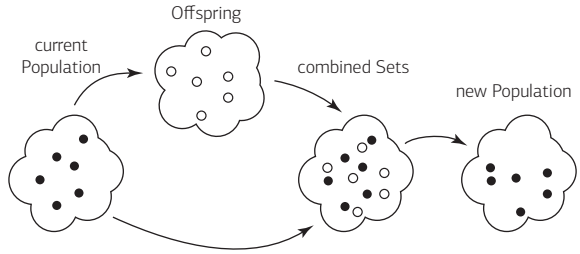

Figure 4.3 During the optimization cycle of the MO-CMA-ES each of the CMA-ES in the population generates a single offspring. Combining the current population with the offspring, the new population is selected.

quality of a given candidate solution is evaluated based on the volume enclosed by the hyper-cube spanned by a candidate solution and the reference point (see Fig. 4.2).

To gain flexibility in terms of numbers of considered objectives, there exists a multiobjective extension of the CMA-ES, the Multiobjective Covariance Matrix Adaption Evolution Strategy (MO-CMA-ES) [52]: in each generation a set or population of CMA-ESs is used, with a population size of λ_{MO}. To generate a new population, each of the λ_{MO} CMA-ESs of the current population generates a single offspring. Instead of a selection just on the offspring, for each CMA-ES the selection is done on the old CMA-ES and its offspring. Thereby λ_{MO} offspring CMA-ES are generated and the population of CMA-ESs for the next generation is selected from the current population and the offspring generated thereof. A single cycle of the MO-CMA-ES is sketched in Fig. 4.3.

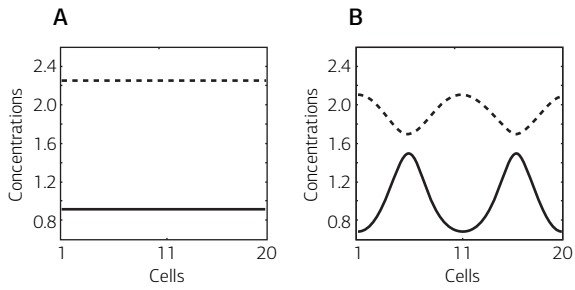

Figure 4.4 Two simulations for a one dimensional Brusselator activator (solid lines) substrate (dashed lines) system (see Eqs. 3.4 and 3.5). Parameter settings are similar in both simulations (a=0.1, b=0.2, β=0.1, c=0.1, D_x=0.1), only one diffusion constant D_y is slightly perturbed. A: D_y=0.69; B: D_y=0.7.

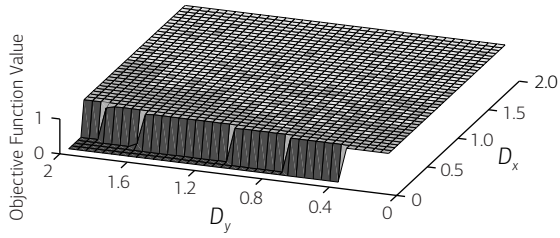

Figure 4.5 Fitness landscape for the parameter space of the Brusselator (see Eqs. 3.4 and 3.5) with respect to the parameters D_x and D_y for the objective function Eq. 4.3. D_x and D_y are sampled in the interval [0.05, 1.95] using steps δ=0.05 while the other parameters are fixed. For the used objective function, values of 0 correspond to spatially heterogeneous patterning while values close to 1 represent spatially homogeneous patterning. These two regions are divided by a ridge-like transition area.

4.3 · Model Parameter Optimization on RD systems: Emergence of Spatial Patterning

Although RD systems can explain a wide range of different patterning in biological systems, calibration of such models with only qualitative data is a difficult task. Due to the non-linear terms in these systems, transitions between qualitative system behavior tend to occur rapidly once certain parameters cross critical values (see Fig. 4.4) while the systems show considerable robustness with respect to parameter perturbations otherwise (see Fig. 4.5). Thereby, already the task of identifying parameter sets for which instead of spatially homogeneous patterns, spatially heterogeneous patterns emerge can be hard to find.

Table 4.1. Initial settings for CMA-ES.

Parameter	Value
Initial Parameters $[a, b, \beta, c, D_x, D_y]$	[0.45, 0.45, 0.45, 0.45, 0.8, 0.8]
Initial Standard deviation for Parameters	[0.25, 0.25, 0.25, 0.25, 0.7, 0.7]
Maximal Number of Objective Function Evaluations	1000

Using the Brusselator RD system (see Eqs. 3.4 and 3.5 in Chapter 3) as an example system, this section investigates what characteristics an objective function for the singleobjective CMA-ES needs to show in order to allow for the identification of parameter settings resulting in the emergence of spatially heterogeneous patterns. Thereby, forming patterns are evaluated with respect to (i) stability of evolving patterning over time and (ii) significance of the heterogeneity of the resulting patterning.

These characteristics and their corresponding objective functions have been evaluated based on 11 runs with each run taking 4 hours on a single CPU of a two chip AMD Opteron $2.6 GHz$ 32-bit machine with $4 GB$ RAM using the initial settings shown in Tab. 4.1.

4.3.1 · Baseline Approach

For the baseline approach, both, spatial heterogeneity of the resulting gene product concentration distribution and convergence of the gene product concentrations over time are scored and aggregated into a single objective $h(x)$ as follows:

$$h(x) = \sum_{i \in gp} (\max(\delta_t - \Delta_{s_i}, 0) + \Delta_{t_i}), \tag{4.3}$$

where gp are all gene products, Δ_{s_i} is the maximal difference in gene product i measured over all cells at the end of the simulation, δ_t is a threshold value which is used to decide if a given spatial heterogeneity is significant, and Δ_{t_i} is the largest change in gene product concentration i in the last integration step. Here, $\delta_t = 0.5$ was used.

In effect, the first term in this objective function can be seen as a penalty term on parameter settings that fail to generate a stable pattern. The second term penalizes settings for which the simulation does not converge within the given number of integration steps.

4.3.2 · Integration of Domain Knowledge

Considering the difficulties in optimizing the parameters without domain knowledge, different scenarios to include domain knowledge into the optimization process are considered. Since sampling the parameter space revealed a fitness landscape wit respect to Eq. 4.3 that is dominated by plateaus (see Fig. 4.5), for the optimization process hardly any search direction is provided. Therefore, domain knowledge is used to generate search direction in the regions where otherwise only random search could be performed. In detail, the proposed to variants are:

1. Introducing a term pointing to a region that it is known to be promising and thereby generating bias towards this region,
2. constraining the parameters considering known dependencies between parameters like the relation between diffusion constants.

For the first variant information on promising regions in parameter space, e.g., stemming from analytical investigations on comparable systems, are introduced in form of an additional term to Eq. 4.3 that penalizes distance to such a known promising region. The resulting objective function reads as follows:

$$f(x) = \begin{cases} \sum_{i \in gp}(\max(\delta_t - \Delta_{s_i}, 0) + \Delta_{t_i}) + \|x - x_t\| & \text{if } \|x - x_t\| > \delta_d, \\ \sum_{i \in gp}(\max(\delta_t - \Delta_{s_i}, 0) + \Delta_{t_i}) & \text{else,} \end{cases}$$

where x_t is a vector representing the target region of the parameter space and δ_d is a minimal length of the difference vector of x and x_t.

The second variant follows a more general idea: it exploits the knowledge about the necessary relation between the two diffusion constants D_x and D_y. Whenever the relation between both constants exceeds a threshold of

0.1, the actual relation is added to the function value. In effect, the search space is soft constrained. The resulting objective function reads as follows:

$$f(x) = \begin{cases} \sum_{i \in gp}(\max(\delta_t - \Delta_{s_i}, 0) + \Delta_{t_i}) + \frac{D_x}{D_y} & \text{if } \frac{D_x}{D_y} > 0.1, \\ \sum_{i \in gp}(\max(\delta_t - \Delta_{s_i}, 0) + \Delta_{t_i}) & \text{else.} \end{cases} \tag{4.4}$$

4.3.3 · Results

The optimization runs undertaken for the baseline variant failed to converge to an optimum within 1000 objective function evaluations. After investigating which parts of the parameter space had been explored during the optimization runs, it turned out that only 3 percent of the tested settings had relations between the diffusion constants D_x and D_y of $D_x/D_y \leq 1/7$. Although it is known from literature that pattern formation using RD systems only takes place if for the relation of the diffusion constants $D_x/D_y \leq 1/7$ holds.

Using the parameter vector (a=0.1, b=0.2, β=0.1, c=0.1, D_x=0.1, D_y=1.5) as a target vector generating search direction towards a good region in parameter space, the obtained results were reasonable but the runs did not converge to a setting with an objective value below 10^{-14}, the convergence threshold used by the CMA-ES. This has to be attributed to the fact that by taking the Euclidean distance between the parameter vector describing the desired parameter region and the actual parameter vector, all parameters equally contribute to the distance between the two vectors. Since D_x is an important parameter that typically takes much smaller absolute values than the other parameters, its contribution to the search direction is overruled by the others and the generated signal is blurred. In fact, emphasizing the contribution of the diffusion constants improves convergence properties. Addressing this problem by directly considering the relations between D_x and D_y in order to guide the optimization process, this variant outperforms the previous two approaches as shown in Fig. 4.6. A representative patterning resulting from this optimization process is shown in Fig. 4.7.

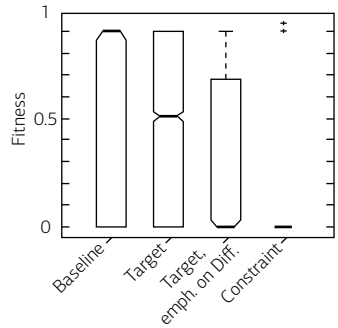

Figure 4.6 Results of conducted runs as boxplots. 'Baseline' refers to the first variant not incorporating any domain knowledge, 'Target' refers to the variant where the domain knowledge was integrated by guiding the search to a promising region, 'Target, emph. on Diff.' refers the runs guided to a promising region with a spatial emphasis on the diffusion constants, and 'Constraint' refers to the variant where knowledge on the relation of the diffusion constants was used.

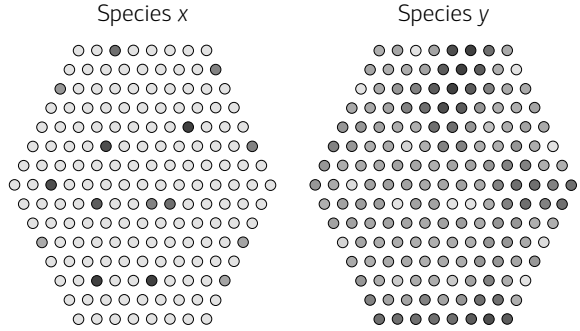

Figure 4.7 Patterning generated by the Brusselator RD system on a two dimensional cell plane with a hexagonal lattice. Each vertex represents a cell. Cells are colored according to their concentrations with respect to the concentrations of the species *x* and *y*. Concentration levels are relative and range from low (light color) to high (dark color).

4.4 · Parameter Space Exploration for RD Models

While for hypothesis validation with respect to GRNs based on modeling one is primarily interested in the identification of a parameter setting for which simulated model output is reasonable close to experimental data, in addition, system properties like robustness of the considered system are of interest. Instead of just a single fitting parameter setting, in this case one is interested in a more rigorous exploration of the parameter space of a given model. In this context, the exploration capabilities of the proposed evolutionary approach for parameter estimation are investigated. The focus thereby is on the design of suitable objective functions for parameter space exploration, based on the segment polarity GRN in *Drosophila melanogaster* as example system.

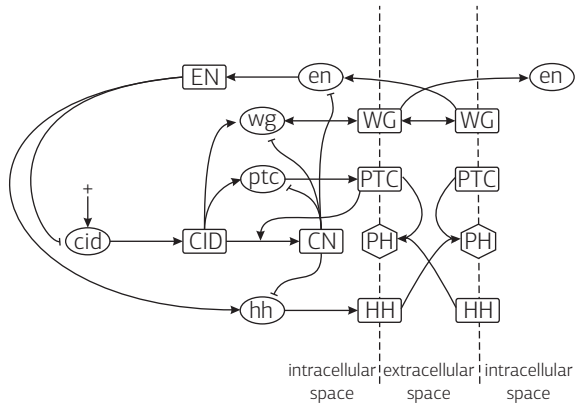

Figure 4.8 A sketch of the segment polarity GRN (adopted from [107]): ellipses represent mRNAss, rectangles represent proteins, the hexagon represents a protein complex and the '+' represents basal expression. The interactions between the different considered species are given by arrows (positive interactions) and arrow stubs (negative interactions). To reflect the spatial component of the GRN, the interface between two neighboring cells is shown.

4.4.1 · Segment Polarity Network

The segment polarity network is a sub-network to the segmentation GRN responsible for body segmentation that plays a crucial role during *Drosophila* development. The formation of the body segments encompasses 3 different groups of genes which are active during different stages. First, information on the location of the anterior/posterior body axis of *Drosophila* is taken up and a first crude division of the body is established by the gap genes. This division is refined to the 14 body segments following the pattern formed by the pair-rule genes. Finally, the polarity of the body segments is defined due to a pattern formed by the segment polarity network.

The segment polarity network uses the pair-rule gene pattern as input stimulus and establishes an asymmetric, stable patterning for the principle outputs, i.e., the gene product distributions of the genes *wingless* (*WG*), *hedgehog* (*HH*), and *engrailed* (*EN*) [107]. The current knowledge on the gene regulative processes between the segment polarity genes is subsumed by a GRN model by von Dassow et al. [107], made up by a set of 13 differential equations and encompassing nearly 50 free parameters such as rate constants and other kinetic constants, cf. supplementary information of [107] for model details. A sketch of the segment polarity GRN is shown in Fig. 4.8.

4.4.2 · Problem Formulation

This section aims at testing the impact of multiobjectivization on parameter space exploration of the segment polarity GRN model of *Drosophila*. The considered GRN is a good example network due to its biological relevance and at the same time being interesting for modeling due to its size while remaining tractable for parameter estimation. In context of parameter space exploration, parameter settings resulting in a patterning of the *Drosophila* body segments as known from experimental studies are to be identified. The target pattern is defined by the spatial distribution of the gene products of 3 segment polarity genes *EN*, *WG*, and *HH* which are supposed to form a timely stable and spatially heterogeneous pattern. Since there exists a set of different parameter settings resulting in the desired pattern [107], exploration aims at identifying a representative set of fitting parameter settings, i.e., a set of parameter settings that result in the desired target pattern while being as diverse as possible with respect to the parameter space. This exploration task is formally described by the following expression:

$$\underset{x \in X}{\operatorname{argmin}}\{h(x)\}.$$

Where x is a parameter setting in the parameter space $X \subseteq \mathbb{R}^n$ and $h : X \mapsto \mathbb{R}$ is a scoring function measuring the deviation between the desired outcome and the outcome resulting from the parameter setting x. An optimal parameter setting x^{opt} therefore has the minimal score of $h(x^{opt}) = 0$.

Instead of considering the single-objective task, it might be advantageous to decompose the function h into a set of functions $\mathbf{h}(x) = (h_1(x), \ldots, h_m(x))$. This multiobjectivization is a straightforward procedure for the given optimization problem: as mentioned before, h already depends on a set of 3 different genes (*EN*, *WG*, *HH*) and for each of these on 2 different characteristics (pattern stability, pattern match) allowing to subdivide h in a set of up to 6 different objectives. The advantage of such a decomposition is anticipated to be twofold: first, the multiobjectivization potentially aids the optimization process [14, 41–43, 53, 59, 83]. Second, in exploration scenarios one is interested in a set of optimal solutions rather than a single solution (as usually provided by singleobjective methods). Therefore reformulating

the task as a multiobjective problem could allow to identify a set of different optimal solutions within a single run and thereby make a more efficient use of the available computation time. The multiobjective problem is formally described by Eq. 4.2:

4.4.3 · Approach

In this section first the used algorithms is described, followed by the proposed objective functions, ending with a description of the implementation of the GRN that is used to generate the data necessary for the evaluation of the objectives.

Algorithms

To contrast the random sampling approach used in [107] and to address the former optimization tasks (see Eqs. 4.1 and 4.2) the MO-CMA-ES is used. A slight variation is introduced to the MO-CMA-ES: instead of the exact hypervolume indicator [110], for the environmental selection step the hypervolume is approximated using a method based on Monte Carlo sampling, HypE [4]. This change is introduced since the computation of the hypervolume is shown to be $\#\mathcal{P}$-hard [13] and the fastest known algorithm has a run time complexity of $\mathcal{O}\left(p^{m/2}\right)$ [5] (with p the number of individuals and m the number of objectives). In consequence, computation of the exact hypervolume indicator is practically infeasible already for settings with 5 objectives and about 50 individuals—a scenario that has to be considered here since more than 5 objectives are used.

Objective Function

As discussed in the supplementary information of [107], a single-objective function can be used for the optimization of model parameters: it encompasses elements scoring pattern match of the gene product distributions of the three genes *EN*, *WG*, *HH* as well as the stability scores for the simulated time of the pattern defined by the 3 genes. In other words, this fitness function is a weighted sum of stability measures and pattern match scores—objective components that are explained in detail in the following.

This composition of different components allows for a straightforward decomposition by considering the single components as separate objectives—a route that is followed for the multiobjectivization approach. In addition, an objective explicitly measuring the diversity of the parameter settings in the parameter space is proposed. This is a plausible addition since the task of parameter estimation of the considered GRN explicitly encompasses maximizing the diversity of the found optimal parameter settings with respect to the parameter space.

Pattern Scoring Following the suggestion of von Dassow et al. [107] (supplementary information) the forming pattern is evaluated by comparing the simulation output to a mask defining the target pattern $p_{ent}^{tar} = \{on, off\}^{|T|}$ (with T being the set of cells forming the considered domain and $|T|$ being its cardinality), indicating which genes are active in which cell. Here the three genes or entities $ent \in \{EN, WG, HH\}$ are considered. To assess the match between simulated pattern and target pattern the simulated data is discretized. For this descretization a threshold of δ_t=0.1 is used: whenever a concentration $[c_i] \geq \delta_t$ $(1 \leq i \leq |T|)$ the corresponding gene is considered to be active in cell i, resulting in another mask $p_{ent}^{sim} = \{on, off\}^{|T|}$. The pattern score h_{ent} then is a discrete function counting the per-cell difference in the simulated pattern and the target pattern for a gene. It is given by the following equation:

$$h_{ent}(p_{ent}^{tar}, p_{ent}^{sim}) = \sum_{p_{ent,i}^{sim} \neq p_{ent,i}^{tar}} 1, \text{ with } 1 \leq i \leq |T| \,. \tag{4.5}$$

Stability Scoring To assess the stability of the resulting patterns, a different approach is chosen than by von Dassow et al. [107]; they used a sliding window on the time-course data of their simulations in order to detect oscillations admitting that the oscillation detection was their main reason for false scoring of patterns (personal communication). Nevertheless, the number of false decisions was rather small. Here, a criterion based on Fourier analysis is used, the spectral envelope [99, 100]. The spectral envelope provides a way to map the variance in a time-series signal to a set of frequency indexed oscillating components (sines and cosines). The spectral envelope

then gives a measure to which proportion the variance in the signal can be attributed to a certain oscillating component. Here, the implementation of the spectral envelope method proposed in [55] is chosen: The time series data $\mathcal{X}_t'$ of all cells with respect to a single gene is considered. It is given as a matrix where the concentrations of the cells are ordered in rows and the columns represent the time course for the time points $t = 1, \ldots, k$. The data is first mean-centered and made unit variant. Using the normalized data $\mathcal{X}$, the frequency discretized periodogram $\hat{I}_k(\omega)$ is approximated by:

$$\hat{I}_k(\omega_j) = \left| \frac{1}{k} \left[\sum_{t=0}^{k-1} x(t) \, \exp\left(-2\pi i \omega_j t\right) \right] \left[\sum_{t=0}^{k-1} x(t) \, \exp\left(-2\pi i \omega_j t\right) \right]^* \right|,$$

where the frequencies $\omega_j = \frac{j}{k}, j = 1, \ldots, \left\lfloor \frac{k}{2} \right\rfloor$ are considered, $x(t)$ denotes the concentrations in all cells at time t, $i^2 = -1$, and '*' denotes the conjugate transpose. The resulting periodogram is smoothed using a sliding window with respect to the frequencies:

$$\hat{P}_{\mathcal{X}}(\omega_j) = \frac{1}{4}\hat{I}_k\left(\omega_{j-1}\right) + \frac{1}{2}\hat{I}_k\left(\omega_j\right) + \frac{1}{4}\hat{I}_k\left(\omega_{j+1}\right).$$

The spectral envelope $\hat{\lambda}(\omega_j)$ can then be estimated by taking the largest eigenvalues of $\hat{P}_{\mathcal{X}}(\omega_j)$ iterating over all frequencies ω_j. For details on the derivation see [55, 100].

Given a noisy signal containing an oscillation, the spectral envelope shows a peak for the frequencies explaining the most variance in the original signal. For a signal that is flat apart from the oscillation the peak should be narrow (cf. Fig. 4 in Jiang et al. [55]). Since the simulation data gathered by numerical integration of the differential equation model is noise-free but the simulations include a transient part where concentrations vary strongly, here a slightly different approach has to be taken in order to identify oscillations in the spectral envelope. In a typical spectral envelope of the simulated data (see Fig. 4.9), one observes a bulge for the lower frequencies that is quasi-monotonically decreasing with growing frequencies. Now, a time course is regarded as oscillating as soon as the envelope $\hat{\lambda}(\omega)$ is getting larger by at least δ_{osc}=10^{-3} for increasing frequencies ω. The oscillation score h_{osc} is

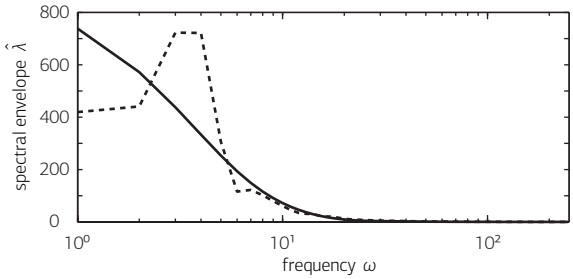

Figure 4.9 Spectral envelopes generated from simulation data for the segment polarity GRN model. The solid line shows the envelope for a typical stable time course of the gene *hh* while the dashed line shows the envelope for an oscillating time course for the same gene.

computed using Eq. 4.6. It has to be minimized and has its optimum at h_{osc}=0. The used δ_{osc} has been determined experimentally and the criterion itself was tested on a set of sample time courses from the segment polarity GRN model.

$$h_{osc}(\hat{\lambda}) = \max_{j=2,\ldots,\lfloor \frac{k}{2} \rfloor} \left\{ \left| \min \left\{ \hat{\lambda}(\omega_{j-1}) - \hat{\lambda}(\omega_j) + \delta_{osc}, 0 \right\} \right| \right\}. \tag{4.6}$$

Diversity Scoring In order to assess decision space diversity the following criterion is used: a ranking on a given set of parameter vectors $\left\{x^1, \ldots, x^k\right\} \in X$ is determined where $X \subseteq \mathbb{R}^n$ is the n dimensional parameter space. The ranking is build using a Monte Carlo procedure: sample parameter vectors $\left\{s^1, \ldots, s^j\right\} \in X$ are uniformly drawn and for each of the sampled vectors s^h $(h = 1, \ldots, j)$ its Euclidean distance to all parameter vectors x^i $(i = 1, \ldots, k)$ is computed. Then, for each of the j sample vectors a ranking on the x^i is generated where the best rank $r = 1$ is assigned to the x^i with the minimal distance to s^h. The worst rank $r = k$ is assigned to the parameter vector x^i with the maximal Euclidean distance to s^h. After all sample rankings have been built the final rankings are determined: for each vector x^i all its j rankings are summed up and the final ranking is given by the ascending order on the summed ranks for every x^i.

For the final ranking j=3810 samples are drawn—following Hoeffding's inequality [45] using 3810 samples results in a probability of 0.90 of resulting in an error in coverage computation of $err \leq 0.125$. For this calculation, the expected distance of two parameter settings in the parameter space $X = [0, 1]^{49}$ is approximated with the worst-case distance of 7.

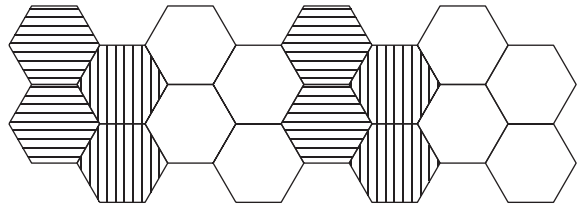

Figure 4.10 Schematic view on the simulated cell domain showing the pre-pattern used as initial condition to the numerical integration: in cells hatched with horizontal lines *WG* levels and those of the encoded protein are set to 1 while in cells hatched with vertical lines *EN* levels and those of the encoded protein are set to 1. For all other genes and cells the initial concentrations are set to 0.

Model Implementation

The model described in the supplementary information of von Dassow et al. [107] is re-implemented in MATLAB®. In agreement with the presented model, as simulated domain 2 rows of 8 hexagonal cells (see Fig. 4.10) are used. The cells are connected with each other via adjacent cell faces allowing for cell-cell communication by exchange of gene products. Like von Dassow et al. [107], periodic boundary conditions are used for the simulated domain. Simulations of the GRN are conducted by numerical integration of the 13 differential equations composing the model. For the numerical integration of the resulting system the implicit-explicit (IMEX) scheme consisting of a modified Crank-Nicolson scheme and an Adam-Bashford scheme [91] is used with a time step $\Delta_t = 0.2$ min and a spatial discretization in cellular resolution. As simulated time interval a window of 1000 min is chosen in compliance with the work by von Dassow et al. [107].

As initial conditions for the numerical simulation the same conditions as proposed by von Dassow et al. [107] are used: all concentrations for all genes are set to 0 except for the RNA and the thereby encoded protein of *EN* in the cells of the 2^{nd} and 6^{th} column as well as those of *WG* in the cells of the 1^{st} and 5^{th} column. In these cells the concentrations are set to 1. Additionally, the basal expression level for the RNA of *CID* in Eq. h (supplementary information of [107]) is set to 0.4. The resulting pre-pattern for *EN* and *WG* is shown in Fig. 4.10.

Table 4.2. Details for the hypervolume calculations: for each of the considered MO-CMA-ES setups (using from 2 to 7 objectives) the reference points and the ordering of the objectives are given. Here 'aggr. score' is an aggregated score for stability and pattern match for a given set of genes, 'stab. score' is the stability score for a single gene, 'pattern score' is the pattern match score for a single gene, and 'div. score' is the diversity score for a set of parameter vectors.

Optimizer	Reference point	Objective order
CMA-ES	–	aggr. score over all genes
MO 2 obj.	(1260, 1.5)	aggr. score over all genes, div. score
MO 3 obj.	(420, 420, 420)	aggr. score separately for *EN*, *WG*, *HH*
MO 4 obj.	(420, 420, 420, 1.5)	aggr. score separately for *EN*, *WG*, *HH*, div. score
MO 6 obj.	(400, 400, 400, 20, 20, 20)	stab. score for *EN*, *WG*, *HH*, pattern score for *EN*, *WG*, *HH*
MO 7 obj.	(400, 400, 400, 20, 20, 20, 1.5)	stab. score for *EN*, *WG*, *HH*, pattern score for *EN*, *WG*, *HH*, div. score

4.4.4 · Simulations and Results

In this section first the performance metrics used for evaluation are described, followed by a description of the simulation setup, and finally presenting results for the simulation runs.

Performance Metrics

The CMA-ES and MO-CMA-ES approaches are compared to random sampling (as reported in von Dassow et al. [107]). The comparison is done with respect to 2 different performance indicators: (i) the mean number of distinct optima found within 3000 evaluations, (ii) the mean inverse coverage of the parameter space achieved within 3000 evaluations. Whereas the first metric is straightforward, a brief explanation how to asses the inverse coverage is given: for a set of optimal parameter vectors $x^i \in \{x^1, \ldots, x^k\}$ the inverse coverage is represented by the expected minimal distance of any of the vectors x^i to any random vector in the parameter space X. This expected value is approximated using Monte Carlo sampling. A set of sample vectors $s^h \in \{s^1, \ldots, s^j\}$ is generated and for each of the s^h the minimal Euclidean distance to the x^i is calculated. The average of these j minimal

distances is used as approximation of the inverse coverage. 381000 samples are drawn for this approximation which using Hoefding's inequality [45] results in a probability of 0.90 of resulting in an approximation error $err \leq 0.0125$. In Hoefding's inequality, the expected distance of vectors in the parameter space $X = [0, 1]^{49}$ is approximated with its worst-case distance of 7.

Simulation Setup

While for random sampling the reported value of in average 1 hit per 200 samples in identifying an optimal parameter setting is taken, the values for the CMA-ES and the different MO-CMA-ES are determined by simulation. For the single-objective CMA-ES the standard parameters except the following are used: maximal number of evaluations of 3000, population size of 30, target fitness of 10^{-8}. Additionally, a restart strategy is used: after termination and as long as there are evaluations left, the CMA-ES is restarted with a new initial vector and the same parameters. For the multiobjective MO-CMA-ES 5 different sets of objective functions are used: (i) the diversity function (see Section 4.4.3) and the objective function for the single-objective case, aggregating stability scores (see Section 4.4.3) and pattern scores (see Section 4.4.3) for all genes; (ii) an aggregation of stability score and pattern score for each gene resulting in 3 objective functions (iii) the 3 afore mentioned objective functions plus the diversity function; (iv) the stability score and pattern score separately for each gene resulting in 6 objective functions; (v) the 6 afore mentioned objective functions plus the diversity function. For the MO-CMA-ES a population size of 100 is used and 3000 evaluations are per run. For the HypE procedure 10000 samples and reference points as shown in Tab. 4.2 are used. For each CMA-ES and MO-CMA-ES setup 21 runs are done with each run taking approximately a day on a two chip dual core AMD Opteron $2.6\,GHz$ 64-bit machine with $8\,GB$ RAM using MATLAB® 7.6 (R2008a).

Results

The reported hit rate for random sampling results on average in 15 optimal parameter settings found in 3000 evaluations, the CMA-ES found on

Table 4.3. Results for sampling, CMA-ES, and MO-CMA-ES runs: for all considered combinations of objectives the mean number of found optima (∅(# opt)) and the mean inverse coverage values (∅(inv cov)) are given, except for the sampling approach: for sampling only the mean number of found optimal parameter settings is given which is calculated based on information given in [107]. In addition, for each considered scenario the number of significantly better (following Conover-Inman) scenarios are given for the coverage values (rank cov) and the numbers of found optima (rank opt) as well as the coverage values (cov).

	Sampling	CMA-ES	MO 2 obj.	MO 3 obj.	MO 4 obj.	MO 6 obj.	MO 7 obj.
∅(# opt)	15	6.572	28.333	25.572	25.381	28.333	26.619
∅(inv cov)	–	0.3634	0.3422	0.3456	0.3442	0.3431	0.3451
rank opt	5	6	0	0	0	0	0
rank cov	–	5	0	0	0	0	0

average about 6 optimal parameter settings in 3000 evaluations, and the MO-CMA-ES found on average more than 25 different optimal parameter settings for any of the considered set of objectives (see Tab. 4.3). The fact that random sampling produces better results than the CMA-ES is surprising on the first glance but has to be relativized: the considered parameter optimization task is comparably easy which is documented by the fact that random sampling is that successful. For the CMA-ES this fact results in a problem since the optimization process would need to converge within 7 generations in order to keep up with random sampling—random sampling identifies an optimum every 200 samples which compares to 7 generation with 30 evaluations per generation for the CMA-ES. To demonstrate its full potential, the CMA-ES would need a more difficult task. Nevertheless, the MO-CMA-ES in average found more than 25 optimal parameter settings clearly improving the results of both, CMA-ES and random sampling. This can be seen as strong indicator for the effectiveness of the multiobjectivization approach and the feasibility of applying evolutionary algorithms like the MO-CMA-ES to the considered optimization task in the first place. The superiority of the MO-CMA-ES is documented by the inverse coverage measure as well: all tested MO-CMA-ES variants improve the inverse coverage compared to the CMA-ES by about 5% (see Tab. 4.3). These results are

tested for statistical significance, first using the Kruskal-Wallis test [18] at a significance level of $\alpha = 0.05$. It was possible to reject the H_0 hypothesis that there is no significant difference in median with respect to number of found optima and coverage between the multiobjective scenarios, the single objective CMA-ES and random sampling. Thereafter, for all pairs of algorithms the difference in the number of found optima and coverage values is compared using the Conover-Inman procedure [18] with the same α level as in the Kruskal-Wallis test. In result the statistical testing shows that the multiobjectivization strategies are significantly better than both, single objective CMA-ES and random sampling, random sampling is significantly better than the single objective CMA-ES, and there is no significant difference between the different multiobjective scenarios. Surprising about this result is that there is no clear difference between the different multiobjectivization setups. At least between those multiobjective variants using the diversity criterion and those only using stability and pattern match a difference would have been expected: while stability and pattern match define the original criteria for the optimization, decision space diversity is completely unrelated and including it in the optimization process should result in a restructuring of the fitness landscape. It would have been expected that this restructuring causes an effect visible in the results. The fact that the results do not reflect this in part might again stem from the fact that the parameter optimization task for the considered network is comparably easy but as well could be an inherent property of the considered GRN or the considered target pattern.

4.5 · Automated Quantification of Similarity between Qualitative Patterns

While in the previous section the pattern recognition was done using masks, this section proposes a more flexible approach for pattern recognition and similarity quantification between simulations and qualitative target pattern.

4.5.1 · Similarity Measure

Based on the simulation framework described in Section 4.1.1 the notion of a target pattern needs to be formalized and pattern components need to be identified in a given picture in order to quantify pattern similarity.

Qualitative spatial gene expression patterns like the one shown in Fig. 4.11A and Fig. 4.11B can be decomposed into a number of basic geometric shapes like rectangles, disks and triangles (Fig. 4.11C). For sake of simplicity, in the following the focus of this section is restricted to a single class of shapes, namely rectangular pattern components. Still, the same approach can be used for other shapes or combinations of shapes as well. To allow for reliable recognition of stripes it is necessary to choose a representation that is invariant against isotropic changes in size, translation and rotation. In this way, e.g., when examining stripe formation in *D*rosophila melanogaster [60], formed stripes aligned with the wrong body axis or deviating slightly in size can be detected; solutions that show promising behavior which otherwise could not be identified. Such an invariant description can be achieved by describing a stripe by a rectangle encompassing the cells the stripe consists of. To allow for invariance against isotropic changes in size, instead of storing absolute values for breadth, length, and depth, the ratio between these three axes is stored; a value staying constant for isotropic transformations. Additionally, since it is not assumed that gene expression domains occur at a specific location in the tissue but rather it is analyzed where in the tissue an expression domain has formed, the description is invariant against translations of domains. Last, since the description of a rectangle by the ratio between its axes ignores the orientation of the rectangle, the description is invariant against rotations. Additionally, by constraining the allowed sizes, recognition of artifacts of small size but correct proportions is avoided.

In order to identify a stripe in the simulation output, a three-step process is used:

1. Identification of connected cells on the same expression level, later on called connected component (CC). Following the intuition that expres-

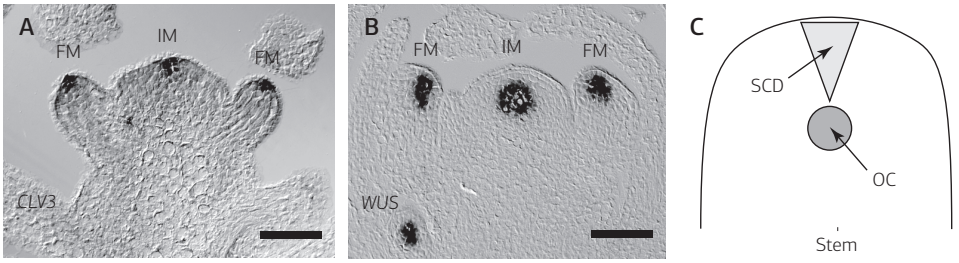

Figure 4.11. Expression patterns of *CLV* and *WUS* in (A) Capsella rubella and (B) *A. thaliana* and a sketch of the resulting SAM patterning derived from these images (C). IM: inflorescence meristem. FM: floral meristem. Scale = 50 μm (RNA in situ hybridizations provided by Ralf Müller, IMPS, The University of Edinburgh).

sion domains formulated in the target patterns are composed of a continuous cell region, CCs are deemed a good representation for domains.
2. Calculation of the principal axis for the identified CCs.
3. Fitting of the target shape onto the CCs.

Following this scheme (see. Fig. 4.12 for a schematic overview on the process), first the cells of the model system are assigned to discrete classes according to their expression profile. Using threshold values, the continuous gene product concentrations (see Fig. 4.12A) obtained from the tissue model are discretized (see second part of Fig. 4.12). Afterwards the sets of connected cells assigned to the same expression class are determined and combined to a CC (see dark cells in Fig. 4.12B). For all CCs assigned to genes and expression levels present in the target pattern, their similarity in shape is measured. First, the CC's principal axes (see black arrows Fig. 4.12B) are determined by taking the first eigenvectors calculated using a principal component analysis. Starting from a rectangle aligned to the principal axes of the CC and encompassing the whole tissue (see outermost rectangle in Fig. 4.12C), a greedy heuristic iteratively shrinking the rectangle is applied. First, slices of the tissue are removed for which all cells are not assigned to the CC. When all these slices are removed, in a second step slices are removed for which the difference between cells not assigned to the CC inside the rectangle and cells assigned to the CC outside the

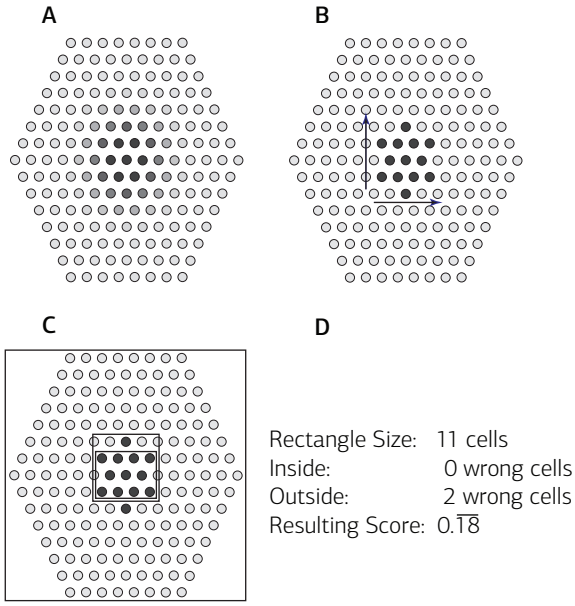

Figure 4.12 A schematic overview on the identification and quantification process of the target pattern formed by basic geometric shapes. From left to right, (A) the distribution of gene product concentrations over the system, (B) the determined CC and its principal axes, (C) rectangles considered during pattern detection, and (D) the resulting score (see Eq. 4.7) is shown. In (C), the outermost rectangle represents the rectangle considered when starting the shrinkage process and the inner two rectangles visualize the last step of the shrinkage process. During this last step, the three cells that are located in the bigger of the two rectangles only, are omitted.

rectangle is getting better, i.e., in the considered slice the number of wrong classified cells is larger than the number of right classified cells. Using this procedure the rectangle is shrunken until no further improvement can be achieved (see Fig. 4.12C). Here, it has to be mentioned that the resulting rectangle not necessarily represents the best rectangle with respect to the mentioned cell ratio but provides a reasonably good approximation. Since the systems considered here only produce stripes parallel to the axis of the rectangular tissue, the alignment step can be omitted.

After identifying the rectangle containing the CC, it is checked if the rectangle adheres to the principal axis ratio and if length and breadth are within the defined intervals. If yes, the scaled difference $g(S_{tis})$ between cells correctly classified by the rectangle and those wrongly classified is used to score

the similarity between the pattern formed for parameter set x and the target pattern. The similarity score $h_{sim}(x)$ is given by:

$$h_{sim}(x) = \begin{cases} 1 - g(S_{tis}) & \text{if } g(S_{tis}) \geq 0, \\ 1 & \text{else.} \end{cases} \tag{4.7}$$

where,

$$g(S_{tis}) = \frac{\sum\limits_{i \in S_{rec}} \mathbf{I}_{S_{CC}}(i) - \left(\sum\limits_{i \in S_{rec}} \mathbf{I}_{S_{CC}^C}(i) + \sum\limits_{i \in S_{rec}^C} \mathbf{I}_{S_{CC}}(i) \right)}{|S_{rec}|}$$

Here, the set S_{tis} denotes all cells in the model, $S_{rec} \subseteq S_{tis}$ refers to all cells within the fitted rectangle (see Fig. 4.12), $S_{CC} \subseteq S_{tis}$ refers to all cells in the considered CC. S_{rec}^C and S_{CC}^C denote the complements of S_{rec} and S_{CC} with respect to S_{tis}. The indicator function $\mathbf{I}_S$ is defined as follows:

$$\mathbf{I}_S(i) = \begin{cases} 1 & \text{if } i \in S, \\ 0 & \text{else.} \end{cases}$$

4.5.2 · Validation of Proposed Approach

In order to validate the proposed pattern identification method and similarity measure, they are integrated into the CMA-ES for model parameter optimization using Eq. 4.4 as objective which is extended by Eq. 4.7. The resulting fitness function is:

$$f(x) = \begin{cases} \sum\limits_{i \in gp} \left(\max(\delta_t - \Delta_{s_i}, 0) + \Delta_{t_i} \right) + f_{sim}(x) + \frac{D_{c1}}{D_{c2}} & \text{if } \frac{D_{c1}}{D_{c2}} < \delta_D, \\ \sum\limits_{i \in gp} \left(\max(\delta_t - \Delta_{s_i}, 0) + \Delta_{t_i} \right) + f_{sim}(x) & \text{else.} \end{cases} \tag{4.8}$$

Test Systems

To evaluate the proposed approach, two RD systems which are known to be able to form stripes [30, 72] are used. During the parameter optimization process, the bold parameters were considered.

1. An activator (a see Eq. 4.9) inhibitor (h see Eq. 4.10) system in a two dimensional plane comprised of 100×10 cells with isotropic diffusion. The plane is initialized with random starting gene product concentration levels $gp_i \in [0.0, 0.1]$:

$$\frac{\partial a}{\partial t} = \boldsymbol{\rho_a} \frac{a^2}{(1 + \boldsymbol{\kappa_a} a^2)h} - \boldsymbol{\mu_a} a + \boldsymbol{\sigma_a} + \mathbf{D_a} \Delta a \tag{4.9}$$

$$\frac{\partial h}{\partial h} = \boldsymbol{\rho_h} a^2 - \boldsymbol{\mu_h} h + \boldsymbol{\sigma_h} + \mathbf{D_h} \Delta h \tag{4.10}$$

   Here, Δ is the Laplace operator, D_i is the diffusion rate of gene product i, σ_i is a base expression of gene i, μ_i is the degradation rate for gene product i, κ is a saturation constant and ρ a cross-reaction coefficient [61].
2. A RD system consisting of two independent activator (a_1, a_2) inhibitor (h_1, h_2) systems like the one described by Eqs. 4.9 and 4.10 in a two dimensional plane comprised of 30×30 cells with anisotropic diffusion—one of the inhibitors diffuses ten times slower in y direction than in x direction and the other inhibitor diffuses ten times slower in x direction than in y direction. Additionally, the system uses an effector gene v [57] depending on both activators (a_1, a_2). It is given by the following Eq. 4.11:

$$\dot{v} = \frac{1}{\tau} u(\sigma v + \rho a_1 + \rho a_2) - \mu_v v \tag{4.11}$$

$$u(x) = \frac{1}{2}\left(1 + \frac{x}{\sqrt{1 + x^2}}\right)$$

   Here, τ denotes the inverse maximal rate of gene v, μ_v is the degradation rate of v and $u(x)$ is a sigmoid function modeling the switch behavior of the response gene. The activator and inhibitor gene product concentrations for this system are initialized with a linear gradient $gp_i \in [0.0, 0.1]$, for the first of the two independent systems the gradient is directed along

Table 4.4. Stripe describing parameters.

Parameter	Value
Breadth Interval	span 5-15 cells
Length Interval	span 90%-100% of one spatial dimension
Considered Expression Gene	activator
Considered Expression Level	expression >80% of the minimum adjusted maximal expression after simulation

the x axis with its maximum on the left border and for the second system in y direction with its maximum on the top border. The initial effector gene product concentrations are set to zero. For Eq. 4.11, the following fixed parameter settings were used: $\tau = 20$, $\sigma = 0$, $\rho = 10$, $\mu_v = 0.1$.

Results

For both test systems a set of twelve separate optimization runs is conducted. Each of these runs tested 1000 parameter vectors and consumed approximately $30h$ of computation time on one core of a two chip dual core AMD Opteron $2.6\,GHz$ 64-bit machine with $8\,GB$ RAM. During the parameter optimization process, the bold parameters in Eqs. 4.9 and 4.10 were considered. All runs managed to identify a stripe pattern matching the representation given in Tab. 4.4 and therefore resulting in a fitness close to zero.

For the first system, the target pattern consisted of one stripe formed by the activator gene whereas for the second system each of the independent activator genes had to form a stripe. Since the second system is designed in such a way that one of the activator inhibitor pairs can form vertical stripes while the other can form horizontal stripes, the effector gene shows a cross-like pattern or if the stripes are located on the borders of the cell plane, an 'L'-shaped pattern. Since the effector gene v depends on both a_1 and a_2, in the region where both concentrations are high the cross or 'L' can degenerate to a diamond or triangular-shaped fraction of a diamond. The resulting patterns for one parameter vector for each test system are shown in Fig. 4.13 and Fig. 4.14. Although the parameter optimization process found appropriate parameter vectors for both systems in almost all runs, none of the runs converged to the theoretical optimum of zero (mean fitness

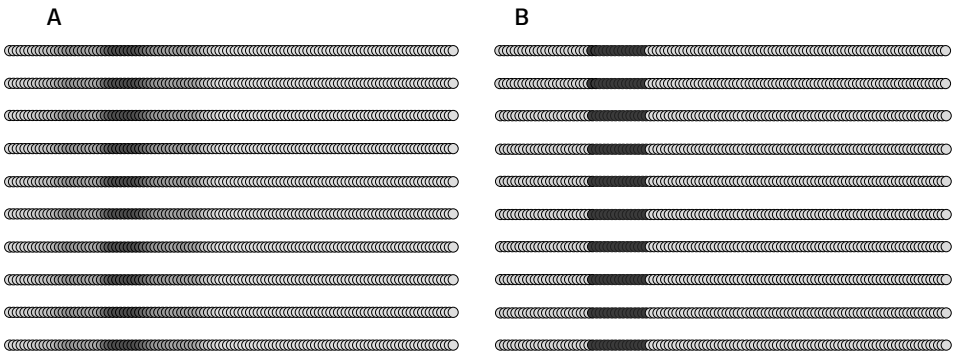

Figure 4.13. For the first test system are shown, A: gene product concentration distribution obtained during simulation; B: the stripe that was identified in (A). In the picture, light colors represent low gene product concentrations and dark colors represent high concentrations.

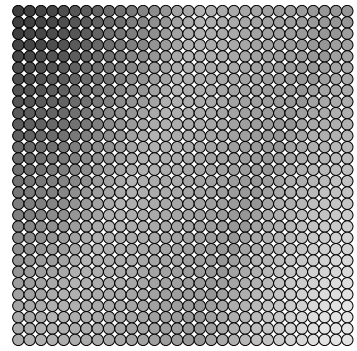

Figure 4.14 Gene product concentration obtained during simulations for the second system. In the picture, light colors represent low gene product concentrations and dark colors represent high concentrations.

over all runs 0.18 for system (1) and 0.367 for system (2) with a standard deviation of 0.23 for system (1) and 0.16 for system (2)). The fact that none of the runs converged to zero results from a penalty for not reaching a significance threshold level of $\delta_t = 0.5$ in Eq. 4.8. Since this threshold was only introduced to make sure that the observed gene product concentration difference is due to the dynamics and does not stem from numerical instabilities in the integration process, the threshold can be lowered in future fitness functions. Therefore in effect, the pattern description and similarity scoring worked reasonably well. An exemplary time course of the evolution

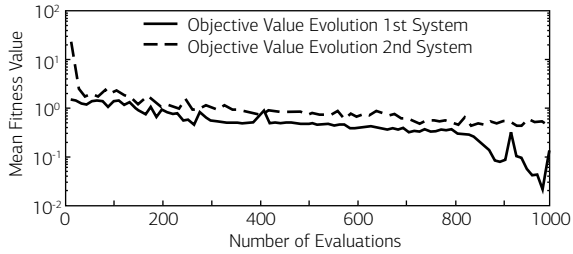

Figure 4.15 Time course data about the optimization process for two exemplary runs, one for system (1) and one for system (2), showing a clear trend towards better rated parameter vectors during the minimization process. Plotted are the mean fitness values for the selected offspring parameter vectors generated during each CMA-ES iteration. Worsening in fitness can occur since no elitist strategy is used.

of the mean fitness of the populations generated during the optimization is showed in Fig. 4.15.

4.6 · Conclusions

This chapter provided an overview on the used modeling and model calibration framework for GRNs in the context of developmental biology that are based on RD systems. Using this framework it has been investigate how to formulate objective functions that are suitable to guide the parameter calibration process for such systems without relying on mostly unavailable quantitative data but only qualitative data. Different aspects of this task have been explored: (i) features that are capable to allow identification of parameter settings resulting in spatially heterogeneous and timely stable patterns, (ii) features that allow assessing properties like sensitivity or robustness and are based on parameter space exploration, and (iii) methods for pattern identification in simulation output as well as suitable functions for quantification of pattern similarities between experimentally determined target patterns and simulated patterns.

While the methods presented in this chapter were conceived as base line approach and therefore tried to avoid inclusion of domain knowledge, it turned out that the problem structure underlying the calibration process makes this

task difficult. Especially ridge-like transitions between qualitative different system behaviors for small changes of parameter values in certain parameter space regions opposed by plateaus that show no recognizable changes for the remaining parts of the parameter space complicate the calibration process. First tests including domain knowledge in turn indicate that such inclusions has the potential to greatly facilitate the process. Therefore, in Chapter 5 two different techniques to include such knowledge are proposed and tested. Still, the results presented in this chapter prove the general feasibility of the chosen approach to model calibration and the presented methods are used during the earlier stages of developing a model for the SAM of *A. thaliana* in Chapter 6.

5

Methods for Model Calibration: Incorporating Domain Knowledge

The previous chapter presented first results concerning the parameter calibration task for models of developmental systems using only qualitative data but it showed limitations as well. These limitations mostly stem from the highly non-linear interactions between modeled components and manifest themselves in form of a problem structure that is dominated by plateaus of indistinguishable parameter settings in terms of respective model fit. In consequence, the optimization method used for calibration hardly receives search feedback towards promising region what results in rather inefficient random sampling behavior on the plateaus.

Here, this limitation is addresses by generating additional feedback for the optimization method by incorporating two different sources of domain knowledge. (i) Already in the previous chapter it was tried to include knowledge concerning certain conditions that are necessary for reaction-diffusion (RD) systems in order to generate a desired patterning: using knowledge on the relation between the diffusion constants of the different

species, the optimization process could clearly be improved. Following this idea of including analytical results concerning the pattern generation mechanism RD system, in a more rigorous approach machine learning techniques are used to extract general features characterizing RD systems. These features reveal structure of the parameter space for an unknown RD system and thereby assist in guiding the optimization process towards promising regions. (ii) In the second approach, observations concerning the developmental trajectories are incorporated into the calibration process instead of knowledge concerning the pattern generation mechanism. While setting up the model for the shoot apical meristem (SAM) of *Arabidopsis thaliana* (*A. thaliana*) described in the next chapter, it occurred that during development of such systems a series of transient states are visited before the final steady state is reached. This fact was exploited in form of a hierarchical decomposition of the developmental trajectory that was used during the manual parameter calibration process at the time the model was set up. Taking up this idea, an automated approach for parameter calibration incorporating information on the developmental trajectory is presented.

5.1 · Exploitation of General Analytical Results Concerning Pattern Formation Capabilities of RD Systems

Encouraged by the fact that inclusion of knowledge on analytical results concerning the used model types can considerably facilitate the parameter calibration task (see Section 4.3), here an approach generalizing this idea is presented. Where in the previous chapter system specific information have been used that are only available for small systems, these analytical results are used in order to train a predictor revealing information concerning the structure of the considered parameter spaces of new systems. In detail, the predictor is used to estimate the distance of a given parameter setting from a boundary in parameter space that discriminates between qualitatively different system behaviors. This predictor then guides an optimization method to identify a well distributed set of parameter settings constituting such boundaries. Knowledge on such boundaries in turn can be

used to concentrate the search process on specific promising regions. Since the considered models for pattern formation in biological systems are based on the same general principles of local self-activation and long range inhibition [31, 69] underlying RD systems, the predictor trained on data for a simple RD system is supposed to generalize well for other RD systems.

5.1.1 · Approach

Summarizing the concept underlying the approach, analytical results concerning a stability analysis of a simple RD system are used to annotate corresponding numerical data with respect to qualitative system behavior. Possible qualitative types of system behavior are shown in Fig. fig:concept. This annotated data set is then used to train a predictor to estimate the closest distance in parameter space of a given parameter setting to a boundary that delimits regions in parameter space resulting in qualitatively different system behavior. For this purpose the numerical data is processed in order to reduce it to some meta characteristics that capture important features to determine to which qualitative region a given parameter setting belongs—a necessary step to allow that the considered characteristics become invariant to the exact specification of the considered RD system and therefore allow for generalization. Using this predictor, in the following it is used to guide the optimization method on its task to explore the parameter space to locate these boundaries. In effect, this approach is divided into two phases, (i) data generation and training of the predictor, and (ii) parameter space exploration. Here, as predictor artificial neural networks (ANNs) are chosen and the Multiobjective Covariance Matrix Adaption Evolution Strategy (MO-CMA-ES) (see Section 4.2) is used for parameter space exploration. The resulting scheme is sketched in Fig. 5.2.

Training Data Generation

To generate the training data for the ANN numerical data has to be annotated with analytical results for an analytically tractable and therefore simple RD system. Here, a simplified variant of the activator-inhibitor system

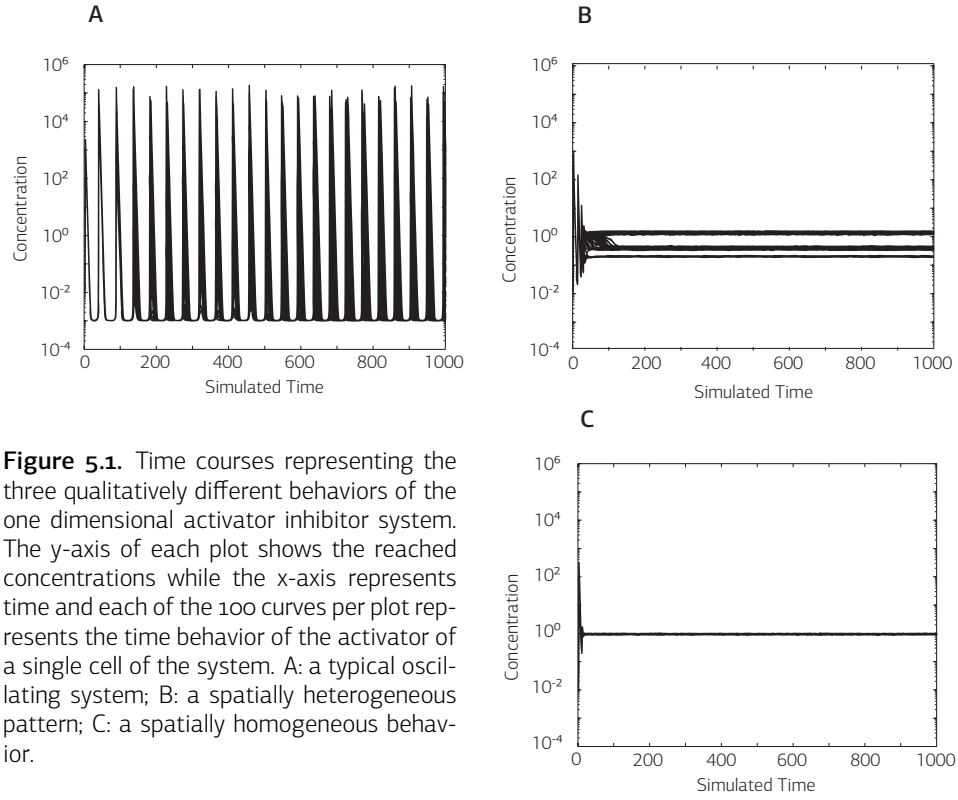

Figure 5.1. Time courses representing the three qualitatively different behaviors of the one dimensional activator inhibitor system. The y-axis of each plot shows the reached concentrations while the x-axis represents time and each of the 100 curves per plot represents the time behavior of the activator of a single cell of the system. A: a typical oscillating system; B: a spatially heterogeneous pattern; C: a spatially homogeneous behavior.

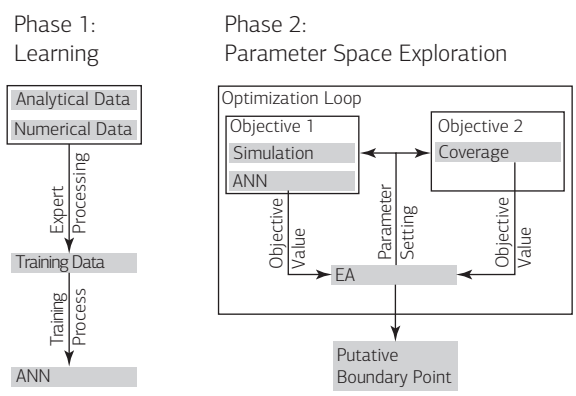

Figure 5.2 Sketch of the training phase of the artificial neural network as a predictor for the distance between a parameter setting and the closest boundary in parameter space delimiting regions resulting in qualitatively different behavior (left) and the information flow in the CMA-ES that is devised to identify points on such a boundary for an unknown reaction diffusion system, building on the artificial neural network (right).

is used that was already analytically investigated by Koch and Meinhardt [61]. It is given by the following equations:

$$\frac{\partial a}{\partial t} = D\Delta a + \xi\frac{a^2}{h} - a + \sigma,$$
$$\frac{\partial h}{\partial t} = \Delta h + \xi\mu(a^2 - h).$$

This system consists of two interlinked species, an activator a and an inhibitor h. Their respective time behavior is described by partial differential equations (PDEs) that in addition to time depend on spatial information: a diffusion term represented by the Laplace operator Δ and a diffusion constant D. Both species encompass a reaction term, perturbed by a uniformly random value $\xi \in [0.95, 1.05]$, and a decay term. The reaction- and decay term of the inhibitor are quantified with a constant μ. In addition, the activator contains the term σ that represents basal expression. This system depends on three constants: D, σ, and μ.

To generate numerical data for this system, an implementation of this system in a one-dimensional spatial domain consisting of 100 cells with periodic boundary conditions is considered. The parameter space is sampled using an equidistant grid of 5000 parameter settings on the parameter sub-space spanned by (D, μ). The basal expression term is set constant to σ=0.001 and as initial conditions $(a_i, h_i) = (0.01, 0.01)$ is used for all cells $1 \leq i \leq 100$ for both species a and h. The grid spans $(D, \mu) = [0.006, 0.3] \times [0.04, 4]$ with respective step sizes of $(0.006, 0.04)$. For numerical integration the interval $[0, 1000]$ of dimensionless time is considered and the implicit-explicit (IMEX) scheme consisting of a modified Crank-Nicolson integrator and an Adams-Bashford integrator [91] is used. Time is discretized based on an equidistant grid with step size δ_t=0.125 and for space discretization a spatial grid in cellular resolution is applied.

After identifying the analytically determined boundaries in (D, μ) parameter space in the numerical data, the shortest Euclidean distance for each simulated parameter setting to these boundaries is computed and normalized: the $(D, \mu) = [0, 0.3] \times [0, 4]$ parameter space is rescaled to $[0, 1] \times [0, 1]$.

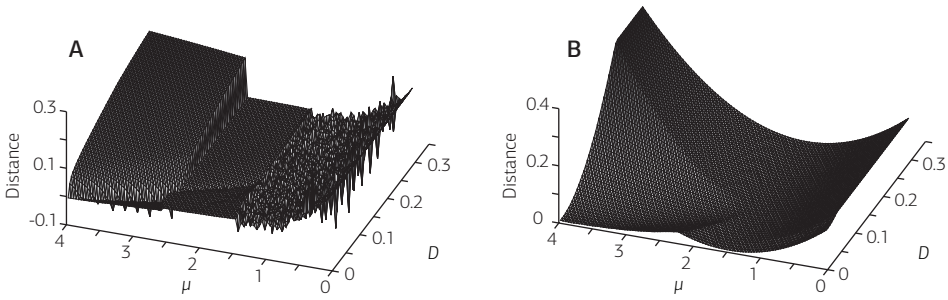

Figure 5.3. Normalized distances of each parameter setting in the $(D, \mu) = [0, 0.3] \times [0, 4]$ parameter space as determined from the simulation data (A) and predicted by the ANN (B).

The resulting distances are shown in Fig. 5.3A. Following the generation of the outputs used for training the ANN, in a last step the integration data (per parameter setting an $n \times m$ matrix X with n being the number of cells and m being the number of considered time points) is reduced to a set of meta characteristics that capture system invariant features. They allow the ANN to learn the mapping between parameter setting (represented by the features) and the shortest distance to a boundary in parameter space. Analyzing the available time course data it turned out that in principle two characteristics should be sufficient to characterize the different parameter settings: (i) the spatial difference occurring between all cells during a late integration step and (ii) the dominating oscillation period estimated from the data (for non-oscillating time courses it can be set to a very small positive value, here $10e^{-14}$). These two characteristics have the advantage that they are invariant with respect to variations of the simulated domain, both in numbers of cells and changes of dimensionality. Still, these two characteristics allow to capture features discriminating between oscillatory and stable system behavior and spatially homogeneous or heterogeneous states. When in addition considering these two characteristics only for the activator, invariance with respect to possible other realizations of an inhibition is gained, e.g., instead of a direct inhibition by an inhibitor, inhibition could be realized by depleting a substrate.

While the computation of the spatial difference is a straight forward procedure, it is briefly explained how the dominant oscillation period is estimated. In a first step, the existing time course data X is reduced to a consensus time course X_{max} by taking the maximum over all cells for each time point. This has the advantage of generating a more regular signal since due to the stochastic ξ terms the considered time course might show some irregularities in single cells. In a second step the data is discretized using a threshold $\delta_{thresh} = 1/2\text{mean}\,(X_{max})$. For all time points where $X_{max} \geq \delta_{thresh}$, the discretized time course data X_{disc} is set to 1 and 0 otherwise. Then, the periods between '1' peaks are determined and gathered in a histogram with buckets encompassing 5 time steps. Using a sliding window covering 5 consecutive buckets, the period with the most occurrences is determined where in case of equal occurrences we have a preference for shorter periods. The process of determining the dominant period is sketched in Fig. 5.4.

Artificial Neural Networks

To learn the mapping between inputs extracted from the time course data of the numerical integration to the distance of the respective parameter setting to a boundary in parameter space delimiting partitions of qualitatively different system behavior, ANNs are used [6]. Since learning the described mapping is a regression problem, feed forward multi-layer perceptrons with two layers of adaptive weights and in total three layers are used: an input layer with two neurons, a hidden layer of 50 neurons with hyperbolic tangent functions as activation functions, and an output layer with a single neuron and a linear activation function. To further enhance the predictive capabilities of the ANN, instead of a single ANN an ensemble of ANNs [96] is trained. In detail the W-SECA procedure proposed by Granitto et al. [36] is used to construct the ANN ensemble where the ensemble prediction is the weighted mean of all ANNs in the ensemble. It is computed using the following weighting function w_i for each ensemble member i [36]:

$$w_i = \frac{e_i^{-2}}{\sum_j e_j^{-2}}.$$

Here, e is the prediction error of an ensemble member with respect to the data set containing all 5000 data points and j iterates over all ensemble

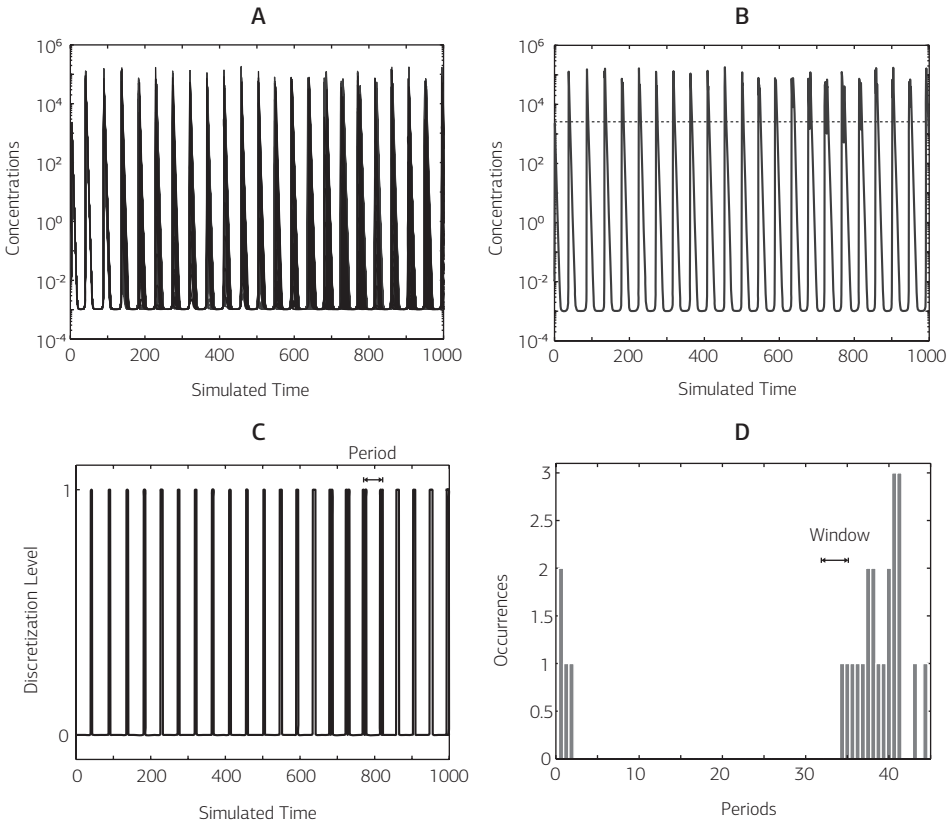

Figure 5.4. A sketch of the process used to estimate the dominant oscillation period in time course data. A: time course data of the activator for a 100 cell activator inhibitor system; B: time courses are reduced to a single time course that represents the maximum for each time point over the 100 cells (solid curve) and the threshold δ_{thresh} used for discretization (dashed line); C: discretized time course data for which the periods between '1' peaks are computed; D: histogram of the calculated periods and the sliding window used to determine the dominant period in terms of occurrences.

members. Since the single input values can become rather large, to facilitate training they are transform by taking their logarithm. Each ensemble member is trained using the scaled conjugate gradients algorithm [6]. For training, the available data is divided in a training set and a validation set using bootstrapping: the training set consists of 5000 bootstrap samples while the not-sampled points are used as validation set. The ANNs are trained using the training set until the prediction error for the validation set in successive training epochs gets worse. Ensemble predictions of the ensemble used in the following are shown in Fig. 5.3.

Optimization Algorithm

Aiming at the identification of a representative set of parameter settings delimiting regions in parameter space resulting in qualitatively different system behavior, the optimization criterion of minimizing the distance to a boundary is complemented by a second objective: the coverage of the parameter space by identified parameter settings. The coverage of the parameter space is assessed by ranking the parameter settings x^i in a population G of the MO-CMA-ES using their distance to uniformly random parameter settings $x^j \in S$ as described already in Section 4.4.3. In total$|S|$=29958 random parameter settings are drawn; following Hoeffding's inequality [45] this results in a probability of 0.95 of producing in an error in coverage computation $err \leq 0.01$ for the considered two dimensional normed parameter space $[0, 1]^2$. In order to identify trade-offs between these two objectives the MO-CMA-ES [52] is used.

5.1.2 · Simulations and Results

In the following results for this approach tested on three systems are shown. As a proof of principle the activator-inhibitor system for which the training data was generated is used. To test the generalizability two additional, conceptually different systems are considered; both realize the long-range

inhibition by some sort of depleting substrate. Equations 5.1 and 5.2 thereby constitute a simplified variant of the activator-substrate system [31, 61, 69].

$$\frac{\partial a}{\partial t} = D\Delta a + \xi a^2 s - a + \sigma \tag{5.1}$$

$$\frac{\partial s}{\partial t} = \Delta s + \xi\mu(1 - sa^2) \tag{5.2}$$

Similar to the activator inhibitor system used for training, the basal expression term is set constant (σ=0.001) and the thereby reduced (D, μ) parameter space is explored. The remaining two Eqs. 5.3 and 5.4 constitute a simplified variant of the Brusselator [86] for which the three-dimensional parameter space (D, x, y) is considered.

$$\frac{\partial x}{\partial t} = D\Delta x + a - (b+1)x + \xi x^2 y + \sigma \tag{5.3}$$

$$\frac{\partial y}{\partial t} = \Delta y + bx - \xi x^2 y \tag{5.4}$$

In all four equations ξ represents a random perturbation uniformly drawn from the interval [0.95, 1.05]. Again, Δ denotes the Laplace operator representing the spatial component, D is a diffusion constant, σ represents basal expression, μ is a degradation rate, and a and b are additional species, the concentrations of which are assumed to remain constant.

Before results concerning these three systems are presented, the used experimental setup as well as the means of validation for the found settings are briefly described.

Methodology

For the optimization process, on each system the same ANN ensemble is used and the MO-CMA-ES is run 10 times. For each run, a maximum of 2500 function evaluations is used with a population size of 50. The hypervolume is calculated for the reference point (1, 100) for the two objectives (i) distance to boundary and (ii) coverage of the parameter space. Each of the runs took approximately 2 days on a two chip dual core AMD Opteron

2.6GHz 64-bit machine with 8GB RAM using MATLAB® 7.6 (R2008a) and the NETLAB [81] implementation for ANNs and related algorithms.

For the evaluation of the runs two different factors are considered: (i) the reproducibility of the identified sets of parameter settings over all runs for each system, and (ii) the goodness of the identified boundaries. Although the reproducibility of the found sets of parameter settings is difficult to asses quantitatively, nevertheless, visual inspection of the sets clearly showed that certain sub spaces contained no identified parameter settings while others were well-populated for all runs; an observation that is deemed to be sufficient to document reproducibility. In order to validate the identified boundaries, two different approaches are used: since for the activator-inhibitor system the boundaries are known, the identified parameter settings are compared visually to the known boundaries. For the remaining two systems the putative boundaries inferred from the identified parameter settings are validated by probing the behavior around the putative boundaries: parameter settings residing on vectors orthogonal to the assumed boundaries are simulated in order to test if a qualitative change of system behavior occurs in the vicinity of the putative boundaries. In addition, using the same probing technique it is tested if parameter settings located in regions for which no boundary is detected exhibit qualitatively similar behavior.

Proof of Principle

The proposed method is used to identify boundaries partitioning the parameter space in regions resulting in qualitative different system behavior for the activator-inhibitor system (Eqs. 5.1.1–5.1.1) that was used for training the ANN ensemble. During preliminary runs it was observed that the coverage of the parameter space became worse during the optimizations process corresponding with a reduction in number of distinct parameter settings constituting the estimated boundaries. Eventually, the algorithm converged ending up with only one or two parameter settings. When analyzing the landscape of distances predicted by the ANN ensemble, it turned out that although the general distance landscape is in good agreement with the calculated distances (see Fig 5.3) not all boundary constituting parameter settings are mapped to the same globally optimal value: for example in

the region with small D-values and large μ-values the predicted distances become negative and in terms of minimization better than those for other boundary points. Thereby, the approach traded off coverage for concentrating on the regions containing negative values. In order to prevent these false global optima from dominating the optimization process, a minimum value of 0 for the predicted distances is introduced. Thereby it is achieved that most boundary constituting parameter settings are mapped to the globally optimal value of 0 but at the same time some false positive boundary points are introduced, e.g., again in the region with small D-values and large μ-values. Using this modification the boundary determined by the approach is in good agreement with the known boundary (see Fig. 5.5). Still, the $(D, \mu) \in [0, 0.3] \times [0, 1.56]$ regime corresponding to oscillating system behavior contains a considerable number of false positive settings. When again checking the predicted distances (see Fig. 5.3B) it can be seen that these false positive settings correspond to narrow spikes in the predicted landscape—a fact that could be addressed either by considering the robustness of the predicted distance to a boundary with respect to some sort of neighborhood around the considered parameter setting or by further refining the training process of the ANNs, e.g., by including regularization terms to smoothen ANN outputs by preventing possible over fitting. Nevertheless, although a number of parameter settings corresponds to false positive boundary points, the approach in its current form already clearly shows that large parts of the parameter space belong to qualitatively similar regions and therefore can be neglected.

Test of Generalizability

After this proof of principle, the generalization capabilities of the approach have been tested by applying it to the remaining two test systems. When checking the data for a simplified variant of an activator-substrate system (Eqs. 5.1–5.2), the identified parameter settings clearly outline a boundary from small D-values and large μ-values towards large D-values and small μ-values. To validate if these settings constitute a true boundary between qualitatively differently behaving parameter space regions the behavior in a neighborhood around the putative boundary was probed using the vec-

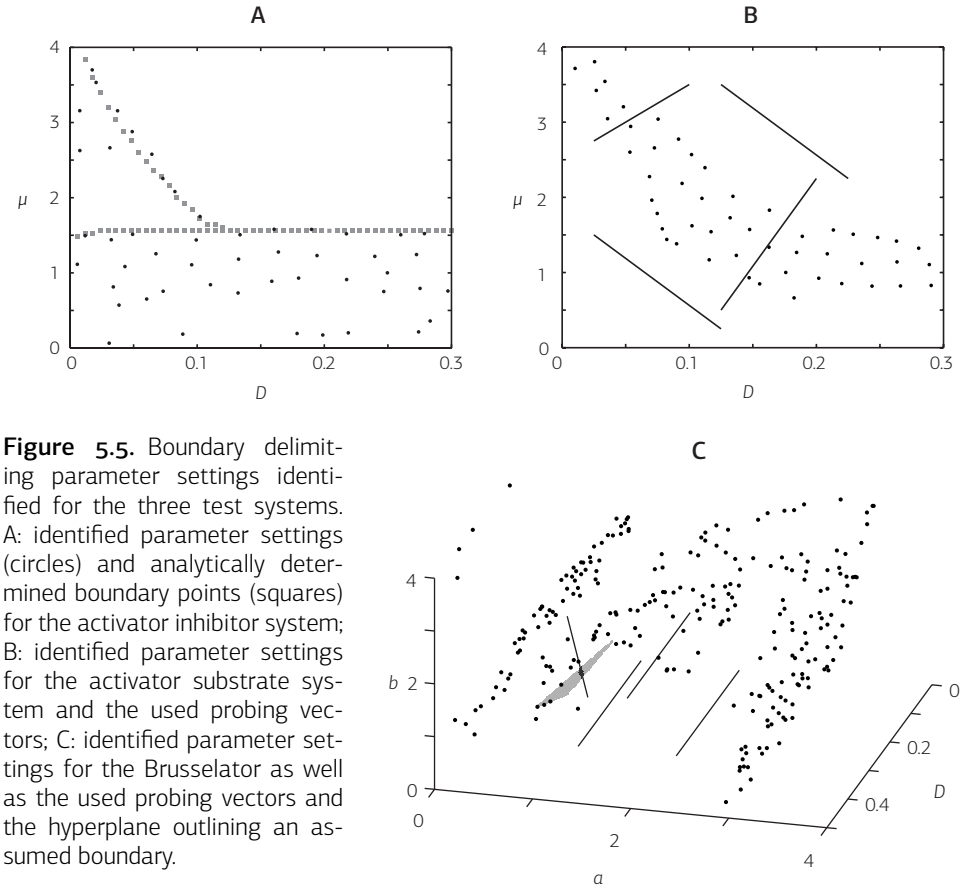

Figure 5.5. Boundary delimiting parameter settings identified for the three test systems. A: identified parameter settings (circles) and analytically determined boundary points (squares) for the activator inhibitor system; B: identified parameter settings for the activator substrate system and the used probing vectors; C: identified parameter settings for the Brusselator as well as the used probing vectors and the hyperplane outlining an assumed boundary.

tors shown in Fig. 5.5B. Evaluating the corresponding simulations it was confirmed that on the lower boarder of the identified boundary the system shows a change in behavior from a spatially heterogeneous pattern (lower region in Fig. 5.5B) to a spatial homogeneous pattern (upper region in Fig. 5.5B). In addition, along the probing vectors located in regions for which no boundary was predicted, indeed no qualitative change in system behavior was observed.

When looking at the putative boundary constituting parameter settings identified for the Brusselator (see Fig. 5.5C), one recognizes that identifying boundaries becomes increasingly more difficult when dealing with higher dimensional search spaces especially when the boundaries stem from non-linear relations between parameters. Still, it was possible to identify a hyperplane outlined by found parameter settings. Using the same probing approach (see Fig. 5.5C for exact location of hyper plane and probing vector) to validate this putative boundary, a change from spatially homogeneous timely stable solutions to timely oscillations was observed when following the probing vector in direction of increasing b values. Again, probing regions that according to the approach were not supposed to contain boundaries showed no qualitative change in system behavior.

5.2 · Hierarchical Approach to Model Calibration

For the calibration of models for developmental systems for which only qualitative data is available, commonly information on a finally emerging stable qualitative patterning of the considered tissue is used to determine the degree of fit between data and model. Here, an approach is proposed that exploits the fact that developmental systems run through a series of intermediate states before the system reaches its final stable state. Thereby, each of the transient stages involves only parts or loosely connected subsystems of the complete system under investigation, e.g., gap genes, pair rule genes, or segment polarity genes in *Drosophila* development [107]. In detail, the approach includes knowledge of these different stages to decompose the optimization task into a set of optimization tasks for the known intermediate states that are concerned with parameters for subsystems only. By considering the intermediate states as well, more information becomes available for the original optimization task. In addition, the parameter spaces that have to be considered for the new optimization subtasks constitute only subspaces of the original parameter spaces what potentially further facilitates the optimization process.

In this context, two hypotheses are investigated: such a decomposition (i) improves the convergence of heuristics for parameter optimization and (ii) results in savings in necessary computational cost due to the fact that for early stages only subsystems need to be considered which are cheaper in terms of computational cost. As test system, the differential equation (DE) model for development and maintenance of the patterning in the SAM of *A. thaliana* presented in Chapter 6 is used.

5.2.1 · Approach

Investigating developmental processes like growth of a plant, one observes that they take place in continuous time and involve changes in expression of a multitude of genes, ultimately responsible for the observed phenotype. Thereby, expression is pinpointed based on concentrations of transcribed RNA or gene products, again living in a continuous scale. Still, states of such system are commonly described involving discretizations: instead of the developmental trajectory a certain time point is considered, often marking an equilibrium state, and concentrations are reduced to levels like high and low. Thereby, especially the restriction to a single point in time results in a loss of information concerning the development. Therefore, the approach proposed here is designed to allow to incorporate information on the developmental trajectory in form of a set of states that are considered for model calibration.

Entity Discretization

With respect to time, commonly the experimentally determined last stable system state is used as target pattern p^{tar} for model calibration. Since the focus of this section is on the incorporation of additional information with respect to time, here expression patterns are represented in a simple form using qualitative masks. These masks are generated for the different model entities for certain points in time using thresholds; in case of the considered SAM model, for each of the five model entities $ent \in \{[\mathrm{WUS}], [\mathrm{FacX}], [\mathrm{WUS_{ext}}], [\mathrm{st}], [\mathrm{CLV3}]\}$ this results in a pattern $p_{ent}^{sim} = \{on, off\}^{|T|}$ (with T being the set of modeled cells of cardinality $|T|$). Masks

deduced from model simulations can then get compared to qualitative expression patterns by counting differences in gene expression states between the two on a per entity and cell basis. In case of the SAM, p^{tar} consists of three masks, one for each of the entities [WUS], [st], and [CLV3]. For model calibration these masks are used to define three difference functions h^3_{WUS}, h^3_{st}, and h^3_{CLV3} that are instantiations of Eq. 4.5 that is repeated here:

$$h_{ent}\left(p^{sim}_{ent}, p^{tar}_{ent}\right) = \sum_{p^{sim}_{ent,i} \neq p^{tar}_{ent,i}} 1, \text{ with } 1 \leq i \leq |T| \,.$$

Time Discretization

Adding further time information might provide useful information for the quantification of differences between simulations and experiments. Here, the fact that developmental processes are often organized hierarchical and in effect run through a series of known transient states can be exploited. For the example system SAM in *A. thaliana* such a decomposition into transient states looks as follows: the structure of a wild type SAM can be described as pattern consisting of two spatially confined domains, stem cell domain (SCD) and organizing center (OC) (see Chapter 6). During SAM de novo assembly, first cells expressing *WUSCHEL* (*WUS*) become visible, thereby indicating formation of an OC. Initiated by WUSCHEL (WUS), stem cell identity is initiated documented by the onset of *CLAVATA3* (*CLV3*) expression [35, 98]. It is decomposed into the states (i) emergence of an OC and (ii) formation of a corresponding SCD. From a modeling perspective the formation of the SCD can be further decomposed: although hierarchically structured, SAM regulation includes feedback of CLAVATA3 (CLV3) on *WUS* expression, therefore introducing a circular component to SAM development. In order to identify suitable model parameter settings, it is helpful to further disrupt this cycle by introducing an artificial intermediate step during which (iii) feedback of CLV3 on *WUS* is ignored.

These three different stages result in the corresponding target patterns with respect to the model entities [WUS], [st], and [CLV3] shown in Fig. 5.6. For the first stage the target pattern captures the formation of a spatially

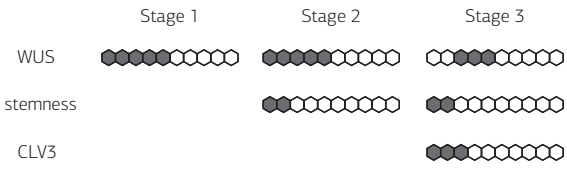

Figure 5.6 Patterning of the SAM used during parameter calibration for different stages with respect to the species [WUS], [st], and [CLV3] for a one-dimensional system.

confined OC at a random location in the modeled domain. Emergence of this domain only depends on a subsystem of the used model encompassing the model entities $\{[\mathrm{WUS}], [\mathrm{FacX}]\}$. The patterning for the second stage, on top of an OC, captures the emergence of a spatially confined SCD. Since for the second stage only $\{[\mathrm{WUS}], [\mathrm{FacX}], [\mathrm{WUS_{sig}}], [\mathrm{st}]\}$ are considered and the negative feedback between SCD and OC via $\mathrm{CLV3}$ signaling is neglected, there are no constraints imposed on their relative arrangement making it most likely that both domains overlap: the negative feedback repelling both domains is missing and the promoting influence of the OC on SCD is strongest at the location of the OC. In turn, the steady state patterning of a wildtype meristem is used for stage three. In this pattern, both functional domains from stage two are still present, this time focusing on their relative arrangement: juxtaposition of both domains and a stable feedback between both with respect to promoting (WUS signaling) and inhibiting ($\mathrm{CLV3}$ signaling) influences. For this stage, the full model is considered.

Complementing the difference functions h^3_{WUS}, h^3_{st}, and h^3_{CLV3} for the third and final stage, for the transient stages and corresponding stage models (see Appendix B) new difference functions are added: (i) for the first stage, h^1_{WUS} evaluating the fit of an emerging OC, and (ii) for the second stage, h^2_{WUS} and h^2_{st} measuring the fit of developing OC and SCD.

The simplest way to incorporate these stages is to always simulate all stages simultaneously. While such an approach in principle should have a maximum of information available in each stage, simulating all stages during every evaluation of a parameter setting introduces considerable overhead in terms of computation time. To avoid this overhead, it makes sense to consider the different stages in a sequential manner, an approach that is in-

spired by strategies for multilevel optimization [50, 75]. Thereby, for every evaluation of a parameter stetting only one modeled stage needs to be simulated. In addition the fact that stages are ordered in a hierarchical manner can be exploited: later stages contain the subsystems and thereby already calibrated parameters from early stages and it could be possible to transfer further information between the optimization processes for consecutive stages, e.g., dependencies between involved parameters or general information on problem structure. In consequence, the problem of calibrating the model with respect to the final stage is decomposed in smaller subproblems, the sum of which could be easier to optimize than the original problem like in dynamic programming approaches.

Implementation

Here, the MO-CMA-ES (see Section 4.2) is used for parameter calibration. With such a strategy for real-valued multiobjective optimization at hand, for the proposed sequential optimization approach the question arises what criterion can be used to determine when the switches between the different stages should be done. Since pretests showed that for a successful sequential optimization it is necessary that the stage change only occurs when a suitable diversity of sufficiently good parameter settings is accumulated for a given stage, stage change occurs once at least 50% of the λ_{MO} CMA-ES in the population of the MO-CMA-ES have identified a unique parameter setting with an optimal fitness of 0 for all considered criteria in the respective stage.

As discussed above, it could be beneficial for the optimization process to transfer information on problem structure or dependencies of parameters between consecutive stages and the design of the MO-CMA-ES naturally supports such transfers: on top of a parameter setting representing the best guess of the heuristic for the considered optimization task, the user is provided with further information on the last search distribution $\mathcal{N}$ from which the final parameter setting was sampled. Thereby, the user gets a hand on information concerning conditioning (C) and overall scaling (s) of the section of the parameter space containing promising solutions. While the CMA-ES typically are started using the identity matrix as starting

covariance matrix C_{new}, in the described scenario it is possible to replace sub-matrices of C_{new} by entries taken from a covariance matrix generated for an earlier stage C_{prev}. In addition, information on scaling can be incorporated as well when comparing the scaling s_{prev} resulting from optimizing an earlier stage to the initial scaling s_{new}: using the quotient $(s_{prev}/s_{new})^2$ as a factor to scale entries taken from C_{prev}.

In addition to these details on optimization methodology, the question how the search space $X \subseteq \mathbb{R}^n$ is best represented in the optimization process should be addressed: while the MO-CMA-ES is designed to handle real-valued spaces, considered model parameters like reaction rates, basal expression terms, degradation rates, kinetic constants as well as diffusion constants tend to live in a logarithmic scale within certain intervals marking a feasible region. To facilitate handling parameters in the evolution strategy, the real-valued parameter space X is mapped to log-scale and normalized to $[0, 1]$ before it is handed over to optimization. In case of the considered SAM model, the parameter space considered during optimization is $X = [0, 1]^{15}$ and only during the third stage the full dimensional parameter space needs to be considered: two of these fifteen parameters are exclusively used for the third stage. In the second stage, thirteen parameters are considered, with five of these newly introduced in comparison to the first stage; resulting in eight parameters used during the first stage.

5.2.2 · Results and Discussion

It is investigate how experimental data can be used in the process of model calibration or model parameter optimization. In this regard hypotheses are tested that inclusion of information on intermediate system states can facilitate parameter optimization for gene regulative network (GRN) models in developmental biology for which mostly qualitative data is available. As example system the model for emergence and maintenance of the SAM in *A. thaliana* presented in Chapter 6 is used.

Test System

The developmental trajectory of the SAM can be hierarchically decomposed into three different stages with respect to time:

1. Formation of a spatially confined domain of cells expressing the *WUS*, the so called OC.
2. Formation of a spatially confined domain of cells taking up stem cell identity, the SCD.
3. Relative spatially arrangement of the two functional domains.

For each stage there exists a target patterning with respect to the considered biological entities (see Fig. 5.6), the emergence of which is to be reproduced with the model. During the three stages the following difference functions are taken into account:

Stage 1: h^1_{WUS} measuring the fit of a developing OC,
Stage 2: h^2_{WUS} and h^2_{st} measuring the fit of developing OC and SCD,
Stage 3: h^3_{WUS}, h^3_{st}, and h^3_{CLV3} measuring the match of developing OC, SCD, and area under $\mathrm{CLV3}$ influence.

Test Setup

As already described, there is a range of possibilities to consider information on the developmental trajectory during model calibration: (i) a baseline approach using only the final system state during model calibration, (ii) an approach that simultaneously considers all three stages, (iii) approaches that consider the stages in a sequential manner. Here, the sequential approaches are differentiated by the amount of information that is transferred between the calibration process of subsequent stages:

- only those parameters newly introduced to the model in the respective stage are optimized and all previously calibrated parameters remain fixed (*fix*),
- approaches where in each stage all respective parameters are optimized but no information transfer takes place (*new*),
- approaches where between stages covariance matrices C of respective search distributions are transferred (*adapt-I*),

- approaches where between stages covariance matrices C and scaling factors s of respective search distributions are transferred (*adapt-II*).

Since in all these approaches a set of difference functions needs to be considered simultaneously at least during certain phases, two variants of each approach are considered: (i) a singleobjective variant were all difference functions are summed up to form an aggregated difference function, and (ii) a variant that considers all objectives simultaneously using a hypervolume based approach (see Section 4.2). A list of all approaches is given in Tab. 5.1.

The different strategies to include these information are tested using the MO-CMA-ES as method for parameter optimization. Possible gains are evaluated with respect to convergence of the optimization process and savings in necessary overall computation time for model simulations. Thereby, for the latter criterion one has to bear in mind that the computational cost during parameter estimation stems from simulation time while the optimization technique itself introduces only negligible computational overhead. Therefore, when for some model evaluations instead of the complete modeled system it is sufficient to simulate only a subsystem, this can result in considerable computational savings. In case of the considered SAM model, computational cost scales linearly with the number of simulated species (see Tab. 5.2)—thereby potentially providing significant savings when considering that parameter optimization runs on state of the art computers still run in the order of days.

Results for the different strategies are tested for statistical significance, first using the Kruskal-Wallis test [18] at a significance level of α=0.05. In these pairwise tests as H_0 hypothesis is tested if there is no significant difference in median with respect to:

convergence capability to identify a parameter setting showing no difference between simulated pattern and target pattern,

quality aggregated score reached at the end of an optimization run,

Table 5.1. Overview on computational approaches (top part) that are considered in two variants (bottom part).

base	Only the last stage of the developmental process is considered for model calibration.
para	All stages of the developmental process are considered in parallel for model calibration.
fix	All stages of the developmental process are considered for model calibration but in a sequential manner. Here, no information is transferred between the different phases of the calibration process and parameters already optimized during earlier stages are fixed.
new	All stages of the developmental process are considered for model calibration but in a sequential manner. Here, no information is transferred between the different phases of the calibration process and all considered parameters remain free.
adapt-I	Another sequential approach that transfers information on the last used covariance matrix C between phases with all considered parameters remaining free.
adapt-II	Another sequential approach that transfers information on the last used covariance matrix C and the scalings s between phases with all considered parameters remaining free.
SO	Singleobjective variant where all difference functions h are aggregated to form a single objective value.
MO	Multiobjective variant that in case of the baseline approach *base* and the sequential approaches considers all difference functions h simultaneously. In case of the parallel approach *para*, the difference functions of the stages are aggregated into one objective for each stage, resulting in three simultaneously considered objectives.

runtime used computation time for a run where a simulation for the first stage produces computational costs c=0.4, the second stage of c=0.8, and the third stage of c=1.

Thereafter, for all pairs of algorithms the differences are compared using the Conover-Inman procedure [18] with the same α level as in the Kruskal-Wallis test (see Tab. 5.3), resulting in a ranking of the strategies (see Tab. 5.4).

Table 5.2. Overview on computational costs for simulations of different stages. For each stage 50 simulations on the one dimensional 10 cell system for representative parameter settings have been done and their respective runtime in seconds have been recorded. Presented are the mean computation times and their respective standard deviations (*std*) per stage as well as per stage runtime normalized with the number of considered species.

	1st stage (2 genes)	2nd stage (4 genes)	3rd stage (5 genes)	1st stage, normalized	2nd stage, normalized	3rd stage, normalized
mean in sec	18.2266	36.5133	45.4424	9.1133	9.1283	9.0885
std in sec	0.1539	0.5203	0.4352	0.0770	0.01301	0.0870

Table 5.3. Pairwise comparison results between different strategies to include information on transient stages into the parameter optimization process. For each tested strategy, the pairwise significance test results are shown with respect to convergence (*c*), quality (*q*), and runtime (*r*). Thereby, the rows show entries in a column whenever the method depicted in a row shows significantly better performance with respect to one of the performance indicators. Significance testing is done using Kruskal-Wallis tests combined with the Conover-Inman procedure.

	$base_{SO}$	$base_{MO}$	$para_{SO}$	$para_{MO}$	fix_{SO}	fix_{MO}	$adapt\text{-}ll_{SO}$	$adapt\text{-}ll_{MO}$	$adapt\text{-}l_{SO}$	$adapt\text{-}l_{MO}$	new_{SO}	new_{MO}
$base_{SO}$		---	---	---	---	---	---	---	---	---	---	---
$base_{MO}$	---		---	---	-q-	---	---	---	---	---	---	---
$para_{SO}$	cq-	c--		--r	cq-	cq-	---	---	c--	---	---	---
$para_{MO}$	---	---	---		-q-	---	---	---	---	---	---	---
fix_{SO}	--r	--r	---	-r-		---	---	---	---	---	---	---
fix_{MO}	--r	--r	---	-r-	---		---	---	---	---	---	---
$adapt\text{-}ll_{SO}$	cqr	-qr	--r	-qr	cqr	cqr		---	c-r	---	---	---
$adapt\text{-}ll_{MO}$	cqr	cqr	--r	cqr	cqr	cqr	---		cqr	---	---	---
$adapt\text{-}l_{SO}$ C	--r	--r	--r	--r	-q-	---	---	---		---	---	---
$adapt\text{-}l_{MO}$	cqr	cqr	cqr	cqr	cqr	cqr	---	---	cqr		---	-q-
new_{SO} C	cqr	--r	--r	--r	cq-	cq-	---	---	---	---		---
new_{MO}	cqr	--r	--r	--r	cq-	c-r	---	---	---	---	---	

Table 5.4. Ranking on significance results of the different strategies to include information on transient stages into the parameter optimization process. For each of the tested strategies and all three performance indicators, convergence (*c*), quality (*q*), and runtime (*r*), the number of strategies significantly besting the respective strategy (-) as well as the number of strategies significantly bested by the respective strategy (+) are given as pair '-/+'.

	$base_{SO}$	$base_{MO}$	$para_{SO}$	$para_{MO}$	fix_{SO}	fix_{MO}	$adapt\text{-}II_{SO}$	$adapt\text{-}II_{MO}$	$adapt\text{-}I_{SO}$	$adapt\text{-}I_{MO}$	new_{SO}	new_{MO}
c	6/0	3/0	0/5	2/0	6/0	6/0	0/4	0/6	4/0	0/8	1/3	1/3
q	6/0	3/1	0/3	3/1	9/0	5/0	0/5	0/6	2/1	0/7	0/5	1/2
r	8/0	8/0	6/1	9/0	3/3	4/0	0/7	0/7	3/4	0/7	0/4	0/5

The testing is done on basis of 12 runs for each strategy, allowing up to 3000 simulations or 300 generations of the MO-CMA-ES using a population size of λ_{mo}=10.

In addition, for this study we decided to circumvent some of the computation time necessary for simulations: instead of considering the two dimensional artificial longitudinal section through the SAM as spatial domain, it is restricted to a one dimensional system containing 10 cells (see Fig. 5.6).

Comparison

With the comparison of the proposed approaches, the two the hypothesis are validated that (i) inclusion of information on transient tissue patterning improves the overall convergence of model calibration processes and (ii) that exploiting such information is capable of reducing the necessary computation time.

Results: Convergence Properties In a first go and addressing the first hypothesis, the baseline approaches $base_{SO}$ and $base_{MO}$ are compared to strategies where for every parameter setting that is considered during the calibration process all three stages are simulated, namely $para_{SO}$ and $para_{MO}$. Indeed, inclusion of these information improves the convergence of the parameter optimization process significantly (see Tab. 5.3): $para_{SO}$ shows bet-

ter convergence properties than both, $base_{SO}$ and $base_{MO}$. At the same time, $para_{MO}$, $base_{SO}$, and $base_{MO}$ are indifferent with respect to convergence.

Including the sequential approaches into the comparison, convergence and quality properties are further improved: $adapt\text{-}II_{SO}$, $adapt\text{-}II_{MO}$, and $adapt\text{-}I_{MO}$ show significant improvements compared to $base_{SO}$, $base_{MO}$, and $para_{MO}$. This finding probably has to be attributed to the fact that the dependence between the different stages during development can be exploited when information on successful search distributions for the parameter subspaces from previous stages is transferred to later stages. An idea that is supported by the fact that the strategies that do not transfer such information but always start of with a new search distribution (new_{SO} and new_{MO}) are not significantly better than either $para_{SO}$ or $para_{MO}$ (see Tab. 5.3). Although it is beneficial to transfer information on search distributions between stages, the early stages seem to be insufficient to already identify definitive settings for the involved parameters that remain fixed during later stages. Respective strategies fix_{SO} and fix_{MO} are both significantly worse than most of the approaches where parameters remain free.

Thereby, the results support the first hypothesis that inclusion of information on transient tissue patterning improves the overall convergence of the model calibration processes.

Results: Computational Cost With respect to the second hypothesis that exploitation of information on transient tissue patterning is capable of reducing the necessary computation time, again first the baseline approaches $base_{SO}$ and $base_{MO}$ are compared to the parallel approaches $para_{SO}$ and $para_{MO}$. While the parallel approaches have access to a maximum of information, simulating always every stage introduces considerable additional computational cost. In case of the SAM model where for the first stage two species, for the second stage four species, and for the final stage 5 species are considered, this amount to 2.2 times the computational effort.

Nevertheless, in terms of runtime the four strategies are indistinguishable despite the extra in computational cost (see Tab. 5.3). This can be explained by the slightly better convergence properties of the latter two approaches

(even when not significant in case of $para_{MO}$): some of the runs converge already quite fast using much less computation time than the baseline approaches.

Considering the sequential approaches as well, all of these provide a significant improvement in necessary computation time when compared to $base_{SO}$, $base_{MO}$, $para_{SO}$, and $para_{MO}$ (see Tab. 5.3). Thereby, the second hypothesis is validated.

Only the influence of using multiobjective variants compared to singleobjective variants is less clear: apart from the variants where all stages are considered simultaneously, the multiobjective approaches are superior to the singleobjective ones (see Tab. 5.4). Since a pre-study has shown that diversity in the found parameter setting is crucial for a successful optimization process, this can be explained with the fact that multiobjective approaches in general support diversity while even set based singleobjective approaches tend to show less diversity. Still, in case of *para* and *new*, the singleobjective variants seem to be superior. A possible explanation for this finding might be that in these cases the multiobjective approach in combination with discrete fitness functions introduce ambiguities that cannot be resolved using the hypervolume and in consequence obscure the optimization process.

5.3 · Conclusions

This chapter presented two different strategies to include domain knowledge in the process of model calibration. The two approaches followed different routes to achieve this goal: (i) information concerning analytical results with respect to the used pattern formation mechanism RD systems have been used and (ii) information concerning the course of the developmental trajectory of the systems under consideration have been used.

In the first approach, the MO-CMA-ES was used to identify parameter settings that constitute boundaries that partition the parameter space of RD systems in regions showing qualitatively different system behavior. This search process was guided by an ANN ensemble which was trained using

numerical data generated for a simple reaction diffusion system and annotated with analytical results. This setup was tested on three different RD systems: the activator-inhibitor system that was used for training data generation, and two conceptually different reaction diffusion systems: an activator substrate system and the Brusselator.

Although this approach is not suited to outline the boundaries separating parameter space regions showing qualitatively different system behavior for high-dimensional parameter spaces due to an exponential increase of settings that would be necessary to outline such boundaries, the method reliably identified parameter settings residing on such boundaries. It therefore can be used to guide the search process towards a region showing a certain desired qualitative system behavior. Thereby, plateaus containing parameter settings that are indistinguishable in terms of resulting qualitative behavior can be avoided and the overall calibration process gets more efficient.

The second approach showed on the example of a model for SAM development and maintenance in *A. thaliana* to aid the process of model calibration or parameter estimation exploiting the fact that developmental trajectories of such systems are hierarchically structured and run through a range of known transient states before they reach their steady state. Using these information it was possible to decompose the respective model into a set of models for the different transient states. The proposed scheme provided additional information for the overall optimization process. At the same time it resulted in a reduction of computation time necessary for model calibration—it allowed to simulated less time consuming models for sub systems instead of the full system for many evaluated parameter settings and thereby saved computation time. Since developmental systems share the hierarchical structure exploited here, this approach is considered to be transferable to further systems.

In conclusion, the proposed approaches documented that the process of model calibration can considerably benefit from the inclusion of domain knowledge in the calibration process. In addition, it provided schemes for

such inclusions for both, mechanism specific or technical knowledge as well as biological knowledge.

6

Modeling Stem Cell Homeostasis and Patterning in the Shoot Apical Meristem of *Arabidopsis thaliana*

Arabidopsis thaliana belongs to the model organisms considered in developmental biology. It is used to study various aspects of regulative processes that often have archetype character for other plants. In addition, some results can even be transferred to other types of organisms like mammals. In this regard, especially aspects of plant development related to stem cells are of interest since for mammals such studies are much more complicated from an ethical perspective.

This chapter is concerned with studying the gene regulative network (GRN) underlying shoot apical meristem (SAM) development and maintenance in *Arabidopsis thaliana* (*A. thaliana*), a tissue that is responsible for aerial growth of the plant and for this purpose harbors a pool of stem cells. As mentioned above, studying the homeostasis of this stem cell domain is of interest due to its possible model character for, e.g., mammalian organisms,

as well as more thorough understanding of might enable researchers to take influence on plant growth related aspects like crop sizes.

With respect to modeling SAM regulation, Jönsson et al. [57] had previously described a model capturing part of the interactions between the two domains mainly responsible for shoot apical meristem regulation and function, the *WUSCHEL* expressing organizing center and the *CLAVATA3* expressing stem cell domain. That model concentrated on the generation of the stem cell domain by a signal derived from from the regulating organizing center [56]. It did not yet include the negative feedback regulation of *CLAVATA3* signaling upon *WUSCHEL* expression, and the creation and maintenance of the *WUSCHEL* domain was not simulated. To confine the *CLAVATA3* domain to the meristem tip, the authors proposed that an (unidentified) factor diffusing from the outermost meristem layer, the L1, together with the *WUSCHEL*-dependent signal, induced *CLAVATA3* expression. They later used a reaction-diffusion (RD) model combined with two repressive signals, derived from the L1 and stem tissue, to activate *WUSCHEL* expression in a deeper meristem region [57]. Both models were successful at reproducing either the organizing center or the stem cell domain, but did not incorporate the mutual interdependence between the factors that shape the two domains, and were less parsimonious with system components than the model described here.

Another recently published model by Geier et al. [29] did not describe the spatial arrangement of the SAM regulating domains, but addressed the observation that stem cell domain and organizing center sizes vary strongly under changing environmental conditions. The model describes the stem cell domain and organizing center as cell pools that are connected via differentiation rates and expand due to cell proliferation, which is regarded as an externally controlled parameter. Variation in the relative sizes of stem cell domain and organizing center can be explained by assuming that a differentiation signal X is produced by organizing center or stem cell domain, which can buffer the response of the cell pools against changes in proliferation rates. Although this model did not allow reproducing all mutant and overexpression experiments that are simulated here, it combined modeling

approaches with quantitative data, and highlighted the enormous developmental plasticity of the meristem.

The model proposed here is concerned with simulating the origin and maintenance of a defined stem cell domain at the tip of *A. thaliana* SAMs, based on the assumption that meristems are self-organizing systems. This model comprises two coupled feedback regulated genetic systems that control stem cell behavior. Based on a minimal set of spatial parameters, it allows to predict autonomous generation, shape and size of the stem cell domain, and the underlying organizing center. It thereby extends the previously published SAM models and is used to explore the parameter space that allows stem cell maintenance, and to simulate the consequences of mutations, gene mis-expression and cell ablations. here, the exploration of the model parameter space has been done using techniques that were developed in the concurrently pursued work on model calibration. In turn the manual, hierarchical model calibration process used for this model and described in Section 6.2.2 inspired the approach for automated hierarchical model calibration presented in the previous chapter.

6.1 · Biological Background on SAM Regulation

Growth of the aerial parts of higher plants relies on a life-long supply with cells by the SAM. The SAM contains a small population of non-differentiating stem cells in the central zone at the meristem tip [98]. After cell divisions in the stem cell domain (SCD), daughter cells are shifted towards the surrounding peripheral zone, where organ primordia are initiated and cells can enter a differentiation pathway. The architectural makeup of flower primordia, which gives rise to the plant¥s reproductive organs, resembles that of the SAM with the main difference that stem cell activity is switched off in flowers after generation of a species-specific number of organs. It is evident that land plants such as trees, which can grow in size and produce new organs for hundreds of years, must have developed robust regulatory systems that enable them to maintain active stem cell

populations also under changing or adverse environmental conditions. Disturbing stem cell regulation can arrest the growth of a plant¥s shoot tip, or may result in gross tissue over-proliferation and failure to reproduce. More subtle alterations in stem cell proliferation can affect overall size of a seed-producing inflorescence structures, such as a maize cob, the size of a fruit, or the number of petals in a horticultural flower. We are only just beginning to understand how the fate of the stem cell population is regulated in higher plants.

Maintenance of the undifferentiating stem cell population depends on signals from cells of the organizing center (OC), which reside underneath the SCD in a deeper region of the meristem. Several gene products have been identified that enable these adjacent cell groups to communicate with each other. The stem cells of *A. thaliana* secrete the CLAVATA3 (CLV3) peptide, consisting of 12 amino acids [9, 24, 63]. *CLAVATA3* (*CLV3*) was shown to interact with the LRR-receptor kinase *CLAVATA1* (*CLV1*) that is expressed in and surrounding the OC [17, 84]. A second receptor system composed of the LRR-protein *CLAVATA2* (*CLV2*) and the membrane associated kinase *CORYNE* (*CRN*) is more widely expressed in the meristem and vasculature, and also contributes to signal perception [54, 79]. *CLV3* dependent activation of the two receptor systems represses the expression of *WUSCHEL* (*WUS*), a homeodomain transcription factor that is normally produced from OC cells, and which is required for the maintenance of stem cells [66, 68]. *WUS* itself acts non-cellautonomously to promote stem cell fate at the meristem tip. The *WUS* protein does not seem to move, and it could control the expression of other genes that generate a diffusible signal which ultimately promotes stem cell identity [37]. Searches for target genes showed that several *ARABIDOPSIS RESPONSE REGULATOR* (*ARR*) genes, which are negative regulators of cytokinin signaling, are repressed by *WUS*, thus involving cytokinin in meristem maintenance [67]. However, *WUS* induces stem cell fate only at the meristem tip, and not in the (*WUS* expressing) OC cells or other surrounding cells, indicating that a spatially restricted cofactor, or a competent cellular state is required to respond to *WUS* activity [104].

Because stem cells signal back to the OC via *CLV3* and its receptors to restrict *WUS* expression, a feedback circuitry is established that maintains a stable stem cell population. Support for this model of stem cell homeostasis comes from a number of experimental observations: (i) loss-of-function mutants of *WUS* cannot maintain stem cells [66]; (ii) loss-of-function mutants of *CLV3* (or *CLV1*, *CLV2* or *CRN*) allow for less restricted *WUS* expression and production of excessive stem cells [9, 24, 78, 79]; (iii) constitutive high level expression of *CLV3* represses *WUS*, causing stem cell loss [9]; (iv) when *WUS* expression is uncoupled from repression by *CLV3*, e.g., when controlled from a heterologous promoter, the stem cell domain expands [10, 94]; (v) the *CLV3*/*WUS* circuitry is capable of self-organization. This was revealed by laser ablation experiments in tomato, showing that after elimination of both SCD and OC, new domains of *WUS* expression are generated at peripheral sites that then initiate new SCDs, which support further growth of the SAM [90].

However, all previous studies performed on various mutants or constitutive mis-expression lines did not allow studying the immediate consequences of system perturbations. Cell ablation experiments are further complicated by wounding effects, and ectopic cell divisions in the SAM which are required for regeneration. Analyzing the dynamics of the *CLV3*/*WUS* circuitry at a shorter timescale required rapid and transient perturbations of gene expression. In a first study of this type, *CLV3* expression was silenced by Dexamethason-induced expression of a foldback *CLV3* RNA [89]. Live imaging of the SAM before and after *CLV3* silencing showed that expression of a *CLV3*:GFP transgene, acting as a reporter for stem cell identity, extended into cells adjacent to the central zone within 24 hours after induction. Importantly, this re-specification of peripheral cells to stem cell identity was not preceded by cell divisions. In a similarly designed experiment, induction of high level *CLV3* expression down-regulated both *WUS* expression, and the stem cell marker *CLV3*, within 3 hours [78]. Together, these experiments showed that the *CLV3*/*WUS* circuitry is acting throughout development, and that the output, stem cell number, can be continuously readjusted in response to changing amounts of the signaling components. In line with

this, fluctuations of central zone size were observed, indicating continuous activity of the circuitry [89].

However, the *CLV3*/*WUS* circuitry was also found to be surprisingly robust and to tolerate changes in *CLV3* expression levels over a tenfold range [78], indicating that stem cells do not directly communicate their number via the amount of released *CLV3* signal. Furthermore, while strong *CLV3* signaling rapidly repressed *WUS* expression, a slowly acting compensation mechanism appeared to upregulate *WUS* with time. The components of this compensatory circuitry are unknown, but may be found among the gene set that controls *WUS* expression. *SPLAYED* (*SYD*) encodes a *SNF2*-type chromatin-remodeling ATPase that is required for *WUS* transcription [65]. BARD1, carrying BRCT and RING domains, interacts with and antagonizes *SYD* to restrict *WUS* expression to the OC [40]. *HANABA TARANU* (*HAN*), coding for a GATA-transcription factor, represses *WUS* post-embryonically from the developing vasculature [109]. The interplay between these components is not understood, and they may act exclusively to establish a discrete *WUS* expression domain when meristems are generated. During development, a second feedback mechanism could operate via the cytokinin signaling pathway. *WUS* represses the expression of several *ARR*s in the meristem, which restrict cytokinin signaling [67]. In turn, continuous activation of *ARR*s arrests meristem activity and *WUS* expression, suggesting that *WUS* and *ARR*s mutually repress each other.

Here, a computational model of stem cell fate regulation by the *CLV3*/*WUS* circuitry in the SAM is proposed. The model incorporates two feedback regulatory systems that merge upon *WUS* regulation. The driving force for modeling was to better understand the forces that shape the *CLV3* and *WUS* expression domains, while making the minimal number of necessary assumptions about the factors to be involved. the model is used to study the effects of targeted system perturbations, and to explore the parameter space that allows for stem cell homeostasis under fluctuating conditions.

6.2 · SAM Model

In this section the proposed model is presented. This presentation encompasses the derivation of the model equations and descriptions of the the occurring model parameters, details on model calibration, and information concerning the numerical simulation process used to generate model predictions.

6.2.1 · Modeling Details

Based on the gene interaction diagram shown in Fig. 6.1, a system of coupled differential equations (DEs) is set up that describes the temporal evolution of concentrations of the factors [WUS], [FacX], [$\mathrm{WUS_{sig}}$], [st], and [$\mathrm{CLV3}$] inside the cells of the SAM. The interplay between [WUS] and [*FacX*] is implemented by a variant of the activator substrate RD system that is known to produce circular domains that remain mobile [61]. Thereby, an OC can form in the meristem tip and move down after initiation of the SCD. The kinetics for the remaining species [$\mathrm{WUS_{sig}}$], [st], and [$\mathrm{CLV3}$] are inspired by the law of mass action. The three-dimensional dome of cells constituting the SAM is represented in the model by a two-dimensional artificial longitudinal section (see Fig. 6.1). This section was generated by positioning cell centers and using a Voronoi decomposition in order to generate possible cell walls and thereby defining the cells. For the simulations zero flux boundaries confining the simulated domain are assumed. The model equations are given in the following:

$$\begin{aligned}\frac{\partial\,[\mathrm{WUS}]}{\partial t} =& D_{\mathrm{WUS}}\Delta\,[\mathrm{WUS}] + w_{ko}\xi\rho_{anc}\frac{[\mathrm{WUS}]^2\,[\mathrm{FacX}]}{1+([\mathrm{CLV3}]+[\mathrm{CLV3_{ext}}])^3}\\ &- \mu_{\mathrm{WUS}}\,[\mathrm{WUS}] + \sigma_{\mathrm{WUS}}, \qquad (6.1)\\ \frac{\partial\,[\mathrm{FacX}]}{\partial t} =& D_{\mathrm{FacX}}\Delta\,[\mathrm{FacX}] + w_{ko}\xi\rho_{anc}\frac{[\mathrm{WUS}]^2\,[\mathrm{FacX}]}{1+([\mathrm{CLV3}]+[\mathrm{CLV3_{ext}}])^3}\\ &+ \frac{\sigma_{\mathrm{FacX}}}{1+\frac{[\mathrm{FacX}]}{K_{\mathrm{FacX}}}}, \qquad (6.2)\end{aligned}$$

$$\frac{\partial [\mathrm{WUS_{sig}}]}{\partial t} = D_{\mathrm{WUS_{sig}}} \Delta [\mathrm{WUS_{sig}}] + \rho_{\mathrm{WUS_{sig}}} [\mathrm{WUS}] - \mu_{\mathrm{WUS_{sig}}} [\mathrm{WUS_{sig}}], \tag{6.3}$$

$$\frac{\partial [\mathrm{st}]}{\partial t} = D_{\mathrm{st}} \Delta [\mathrm{st}] + \mathbf{1}_{id}(i)\, \rho_{\mathrm{st}} \frac{\left(\frac{[\mathrm{WUS_{sig}}]}{K_{\mathrm{st}}}\right)^5}{1 + \left(\frac{[\mathrm{WUS_{sig}}]}{K_{\mathrm{st}}}\right)^5} - \mu_{\mathrm{st}} [\mathrm{st}], \tag{6.4}$$

$$\frac{\partial [\mathrm{CLV3}]}{\partial t} = D_{\mathrm{CLV3}} \Delta [\mathrm{CLV3}] + c_{ko} \rho_{\mathrm{CLV3}} [\mathrm{st}] - \mu_{\mathrm{CLV3}} [\mathrm{CLV3}]. \tag{6.5}$$

The model depends on the following parameters: reaction rates ρ, basal expression rates σ, degradation rates μ, and kinetic constants K. While most parameters settings are similar for all cells of the considered simulated domain, there are two exceptions: (i) the reaction rate ρ_{anc} (Eqs. 6.1 and 6.2) is given by a distribution with its maximum in the meristem tip (see Fig. 6.1), an artificial spatial component necessary for correct location of developing SCD and OC within the meristem. In addition ρ_{anc} is perturbed by a small uniformly random value $\xi \in$ [-0.025, 0.025] which is a random influence necessary for this subsystem to produce patterns. (ii) The reaction term guiding [st] depends on the indicator function $\mathbf{1}_{id}$. Since only cells in the outer cell layers are assumed to be competent to acquire stem cell identity, for competent cells i the indicator function returns a value $\mathbf{1}_{id}$=1 and $\mathbf{1}_{id}$=0 otherwise. While the former parameters refer to processes taking place within the cells, the model includes interactions between neighboring cells as well. This interaction is modeled by diffusion terms $D\Delta$, where Δ is the Laplace operator in two dimensions and D is a diffusion rate. Note that although [WUS] and [st] are considered to be only locally active, their model equations (Eqs. 6.1 and 6.4) contain diffusion terms since for these two factors a weak leakage diffusion is assumed.

On top of the parameters that are already described, the model contains a set of parameters that is used to accommodate the different modeled mutants and experiments. In this context, c_{ko} is used to simulate regulation of endogenous CLV3; c_{ko} modifies the reaction term guiding *CLV3* expression where c_{ko}=1 simulates a wildtype situation while c_{ko}=0 represents a

knockout and values c_{ko} >1 represent overexpression of endogenous CLV3. Choosing c_{ko} values $c_{ko} \in [0, \infty)$, graded scenarios can be simulated. Using a comparable parameter $w_{ko} \in [0, 1]$, it is possible to simulate graded *WUS* knockouts (see Appendix A). Again a value of w_{ko}=1 represents a wildtype scenario while w_{ko}=0 represents a knockout. The parameter $[CLV3_{ext}]$ is used to simulate a CLAVATA background in all cells. In a wildtype scenario, $[CLV3_{ext}]$ is set to 0 and with $[CLV3_{ext}] \in [0, \infty)$ a range of graded CLV3 background strengths can be simulated. In addition, for simulations testing conditions with respect to *factor X* (*FacX*) under which OC and SCD are generated the already introduced parameter K_{FacX} is varied.

6.2.2 · Model Calibration

The model parameter setting for the wildtype simulations as well as for the simulations of mutants are given in Tabs. 6.1 and 6.2. As mentioned already in the introduction, at the time where the model was generated only prototypes of the later developed parameter calibration method have been available. Therefore, at this stage model parameters have been tuned by hand using a type of hierarchical decomposition of the system: from a developmental perspective firstly the OC is formed which then induces the formation of a SCD. For the parameter tuning process therefore the system has been divided into two parts. (i) Equations 6.1 and 6.2 which are responsible for the formation of the OC. (ii) Equations 6.3, 6.4, and 6.5 that constitute the SCD generating part of the system. During a three-stage tuning process first only the OC generating system is considered in order to identify parameters resulting in a single and spatially confined OC domain. For this subsystem, initial parameter settings as documented previously [61] were used, and it was possible to identify fitting parameters reasonably close to the initial settings. Using the resulting OC as input, the parameters for the SCD generating part of the system have been tuned. Here, the feedback of CLV3 on *WUS* was neglected and instead the goal was to identify parameter settings which resulted in an SCD of appropriate size and an area under CLV3 influence that extends the SCD but remains spatially confined as well. In the last step, the full system was considered,

Table 6.1. Summary of constant model parameters. ρ_{anc} follows a constant distribution shown in Fig. 6.1.

Parameter	D_{WUS}	D_{FacX}	$D_{WUS_{sig}}$	D_{st}	D_{CLV3}	ρ_{anc}	$\rho_{WUS_{sig}}$	ρ_{st}	ρ_{CLV3}
Value	0.002	0.02	0.02	0.002	0.02	dist.	0.5	0.6	0.03
Parameter	μ_{WUS}	$\mu_{WUS_{sig}}$	μ_{st}	μ_{CLV3}	σ_{WUS}	σ_{FacX}	K_{FacX}	K_{st}	
Value	0.004	0.05	0.05	0.01	0.0002	0.004	0.2	1	

Table 6.2 Scenario dependent model parameters with their respective values. WT describes the wild type setting. CLV3↑: over-expression of CLV3 in all cells (stem cells and non-stem cells). CLV3↓: CLV3 loss of function mutant. *CLV3*≫WUS: Expression of WUS in the stem cell domain, controlled by the *CLV3* promoter.

	$[CLV3_{ext}]$	c_{ko}
WT	0	1
Laser Ablation	0	1
CLV3↑, medium	1	1
CLV3↑, strong	1.5	1
CLV3↓	0	0
CLV3↓, gradual	0	0.2
CLV3 ≫ WUS	0	1

adjusting the parameters responsible for the feedback process between the two subsystems. This hierarchical approach to model calibration gave rise to the automated hierarchical calibration method described in Section 5.2.

6.2.3 · Simulation Details

For the numerical simulation of the model partial differential equations (PDEs), zero flux conditions on the boundaries of the cell plane are assumed. In order to simulate the time evolution of the model, the model equations are numerically integrated. The equations have to be discretized with respect to time and space, here using a constant time step δ_t=0.25 and applying a finite difference scheme in cellular resolution to the space-dependent diffusion terms. Each cell is thereby represented by its center and for the sake of simplicity, cell volumes are considered to show the gradients that would be assumed between the concentrations simulated for the cell centers. In addition, free diffusion is assumed as the only means of communication between cells. Because the precise communication underlying WUS dependent signaling is still unknown, free diffusion represents a sort

of 'maximum entropy choice'. In terms of modeling complexity we benefit from this fact: with free diffusion, extracellular spaces, cell walls and membranes can be neglected during the simulation process. In addition, due to the Voronoi decomposition used to generate the considered section through the meristems, cell volumes and surfaces tend to even out. Since the model is not supposed to generate quantitative data, but rather to investigate qualitative behavior, the influence of cell surfaces and volumes during simulations is neglected.

The diffusion terms in the considered system tend to be stiff, and therefore a variant of the second order implicit Crank-Nicolson integrator is applied to these terms, while to the only-time-dependent terms a faster explicit Adams-Bashford scheme is applied. This implicit-explicit (IMEX) scheme (see Section 3.2.3 in Chapter 3) is chosen in order to reach an appropriate trade-off between necessary computational effort and simulation accuracy: due to their stiffness, the diffusion terms are difficult to simulate. Therefore, the computationally demanding but more accurate Crank-Nicolson method is used. The reaction terms in contrast are much less demanding and can therefore be simulated using a much faster but less accurate explicit scheme.

The numerical simulations for the considered scenarios are done in a two-stage process. The first stage is used to equilibrate the system starting from the initial conditions. Here, under 'equilibrium state' a system state in which all derivatives are reasonably close to zero is understood. During the second stage the parameters are adapted in order to accommodate the considered scenarios. As initial condition for the first stage, the WUS level of all cells is homogeneously initialized with a starting concentration of [WUS]=0.01. All other species are initialized with a value of 0. In the first stage, the system is simulated for 30000 time units. For the second stage the equilibrium concentrations obtained in the first stage are used as initial conditions, the parameters are adapted and in case of the laser ablation scenario the tissue topology is adapted. Afterwards the system is simulated for further 15000 time steps.

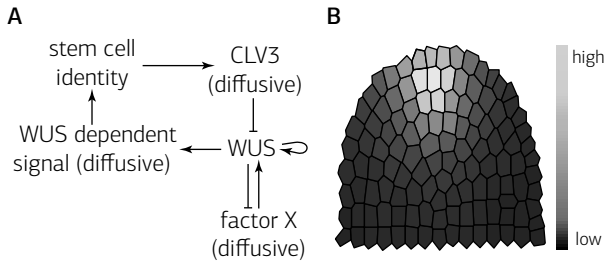

Figure 6.1 Modeling components and cellular framework. A: Components and their interactions regulating stem cell homeostasis in *A. thaliana*. B: Two dimensional meristem frame with assumed anchoring distribution indicated by coloring.

6.3 · Simulation Results

A PDE model following the dynamics of gene regulation across the SAM (see Fig. 6.1) is proposed. Conceptually, at the center of the model lies the regulation of *WUS* via two separable feedback operated RD systems, a commonly used type of DE models for developmental processes in biology [103]. Instead of representing the entire meristem structure, the spatial component is restricted to two dimensions using an artificial longitudinal section through the SAM (see Fig. 6.1). Cells within this meristem section are modeled as discrete entities. Growth and cell divisions are neglected for two reasons: firstly, this study focuses on meristem homeostasis, i.e., meristem size remains unaltered, and secondly, the gene regulation that is considered in this study is faster than the cell cycle. The cellular or tissue framework thus remains static. The regulative processes within cells are mapped to a set of PDEs constituting a gene regulative program, which is executed in each cell. The components and underlying assumptions of the model are summarized as follows (see Fig. 6.1):

Stemness: Cells of the meristem can acquire stem cell identity, reflected in their level of stemness, which is controlled by a WUS-dependent signal (*WUS*-signal, see below). An artificial and static cut-off concentration for *WUS*-signal is avoided, above which cells switch to the stem cell status, and instead a dynamic, but sigmoidal response to *WUS*-signal is established, resulting in variable levels of stemness to represent a cell¥s state.

Experimental evidence from *WUS* misexpression shows that only outer cell layers acquire stem cell identity. The underlying factors responsible for this are not known. Bow not another signal is postulated, but this observation is taken into account by allowing only cells in outer cell layers to acquire stem cell identity. Therefore, only cells in the outer layers of the meristem are competent to react to *WUS*-signal and stemness is restricted to the outer layers. Stem cells express the signaling molecule CLV3, proportional to their stemness level. The stemness levels are expressed by the model variable [st].

CLV3: CLV3 freely diffuses to neighboring cells. To avoid flooding the entire model with CLV3, the CLV3 peptide is regarded to decay with time. The need for receptor proteins or other signaling components is eliminated because insufficient quantitative data are available to assess their contribution. Furthermore, *CLV3*-signaling appears to be largely controlled by the amount of available CLV3 peptide [78]. Thus, the local CLV3 concentration is computed to directly restrict *WUS* expression. The CLV3 levels are expressed by the model variable [CLV3].

WUS: It is assumed that WUS protein is not mobile and therefore remains mostly in the cells where the *WUS* gene is expressed [37], except for weak leakage diffusion. While all cells of the model meristem are in principal able to express WUS, a spatial parameter is added which makes cells that reside closer to the meristem tip more competent to activate WUS expression (Fig. 6.1). This anchoring was found to be necessary in the presented model to ensure correct positioning of the two functional domains (SCD and OC) within the dome. Without the spatial component, immediately neighboring SCD and OC are still formed, but at more random locations (see Fig. 6.2). The requirement for a spatial component reflects the fact that this virtual meristem is not structured, i.e., all cells are intrinsically equal and carry no positional information. In plant meristems, such spatial information will be provided by signals within and between cell layers, or from the vasculature. A positive feedback loop for *WUS* via auto-activation is introduced. Although not experimentally proven, it is supported by the observation of a rapid upregulation of *WUS* expression in regenerating callus [35]. WUS promotes the expression of

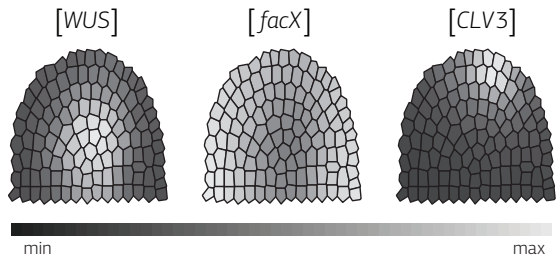

Figure 6.2 System without anchoring distribution. Equilibrium state of a simulation where the anchoring distribution guiding WUS expression is exchanged by a constant *WUS* reaction rate. As a result of the missing positional information, SCD and OC keep their positions relative to each other, but the SCD is now formed at a random location in the outer meristem layers.

WUS-signal, which is mobile and can diffuse to neighboring cells. The WUS levels are expressed by the model variable [WUS].

WUS-signal: WUS-signal is generated by all *WUS* expressing cells, and the amount produced depends on the levels of WUS expression. Similar to CLV3, WUS-signal is mobile and degraded at a constant rate. Cells react to the amount of WUS-signal they receive with stemness. Only outer cell layers of the meristem are competent to respond to the WUS-signal. The WUS-signal levels are expressed by the model variable [$\mathrm{WUS_{sig}}$].

factor X: To account for *CLAVATA* (*CLV*) independent regulation of *WUS* expression, *FacX* is incorporated. At the start of the simulation, *FacX* is expressed homogeneously in all cells and is freely diffusing. FacX induces *WUS* expression, but is itself under negative feedback regulation by WUS [78]. This is implemented through active degradation or consumption of FacX by WUS. The interactions between *WUS* and *FacX* are thus based on an activator-substrate-like mechanism that will generate a discrete WUS domain. FacX levels are expressed by the model variable [FacX].

The described entities are compiled into a PDE representation of the intracellular gene regulative program given by Eqs. 6.1-6.5 (see Section 6.2). An overview of the interactions is given in Fig. 6.1.

The resulting model depends on a set of parameters like kinetic constants; validation of the model therefore required first to identify a parameter setting that allow reproducing the two functional meristem domains, i.e., the *CLV3*-expressing SCD and the *WUS*-expressing OC, at approximately those

locations which are experimentally observed in wildtype meristems (see Fig. 6.3).

6.3.1 · Simulation of Wildtype

Starting from almost zero concentrations of all considered components, the system was simulated until an equilibrium state was reached (see Fig. 6.3); since system behavior is investigated by means of numerical integration of the model equations, under 'equilibrium state' a state is understood where all derivates are zero or reasonably close to zero. In the wild-type scenario, a given meristem showed *WUS* expression first at the meristem tip, triggered by a sufficiently high level of FacX. WUS then increased, thereby repressing FacX at the same location. The center of the OC is shifted downwards. The distribution of WUS-signal overlaps with that of WUS, but since WUS-signal is diffusible, it is located in a wider domain and extended always to the meristem tip. Together with an increase of stemness at the meristem tip, CLV3 becomes expressed, pushing the OC downwards from its initial position. This time course reproduces the dynamic changes in gene expression patterns that are observed during embryonic development of the SAM.

6.3.2 · Simulations of mutants and system perturbations

In the following, further simulations for known mutant experiments are shown and compared to the respective experimental data.

Reducing *CLV3* expression

First the consequences of reducing *CLV3* expression is tested. When *CLV3* is downregulated during plant development, both the SCD and the OC expand laterally due to unrestricted *WUS* expression [9, 89, 94]. Furthermore, *WUS* expression is then no longer excluded from the meristem tip.

The in silico analysis started from an equilibrated wildtype meristem, thus simulating a conditional knock-out of *CLV3*. When *CLV3* expression was stopped, the SCD rapidly enlarged due to recruitment of lateral cells. At the same time, the OC expanded and shifted towards the meristem tip (see

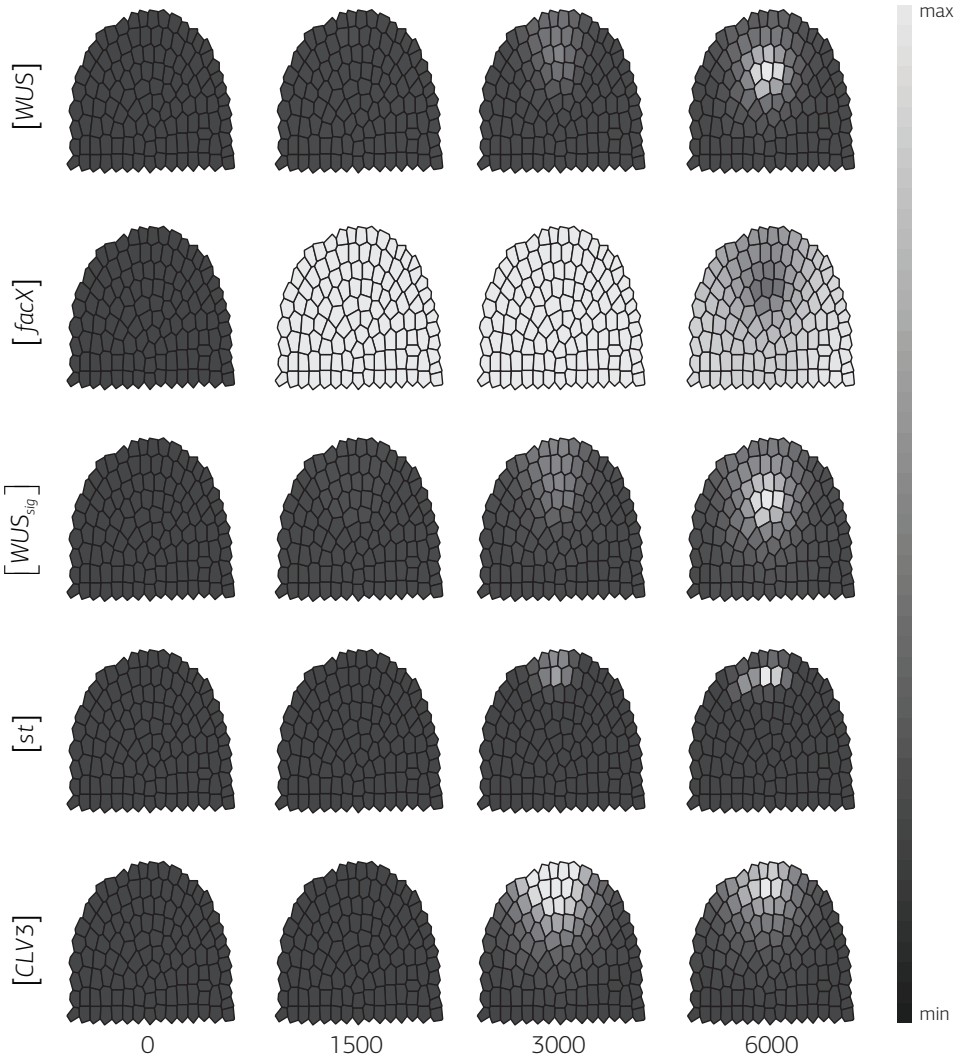

Figure 6.3. Time course simulation of a wild-type meristem. Grading from dark (minimum) to light (maximum) illustrates the relative concentration of the indicated components. From left to right, the simulations were developed from close to zero concentrations for number of steps shown below, until the (stable) equilibrium state was reached.

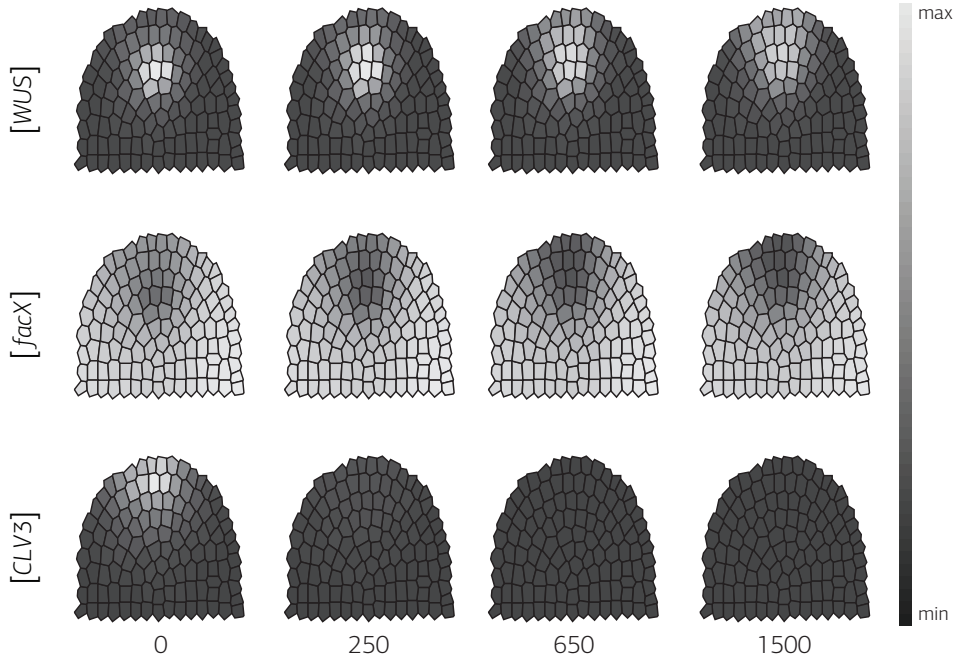

Figure 6.4. Time course simulation for a *clv3* loss-of-function mutant. Only the concentrations of WUS, FacX and CLV3 are shown, starting from an equilibrated wildtype. At the beginning of the simulation, CLV3 concentration is set to 0, and simulations proceed until equilibrium. Note that the WUS domain expands and shifts upwards.

Fig. 6.4). A similar, but less pronounced effect was seen when *CLV3* was still expressed, but at reduced levels (see Fig. A.1 in Appendix A).

Increasing *CLV3* Expression

Plants that continuously express *CLV3* fail to maintain a shoot meristem due to an early arrest of *WUS* expression and stem cell differentiation [9]. However, inducible overexpression during development was found to be compensated in some flower meristems, resulting in a recovery of *WUS* expression at later stages [78]. In simulations, high level expression of *CLV3* in all cells caused a rapid shrinkage of the OC, downregulation of *WUS*, and a reduction in stemness, concomitant with a reduction in *CLV3* expression

levels from the SCD (see Fig. 6.5). Overexpression of *CLV3* set at an intermediate level resulted in *WUS* repression and a rapid loss of stemness, which recovered with time (see Fig. 6.5). A similar behavior was observed in plant floral meristems, in response to induced overexpression of *CLV3* [78]. Low level overexpression of *CLV3* allowed the system to reach a new stable equilibrium state, with a smaller OC and SCD (see Fig. A.3 in Appendix A).

Further, the robustness of the system against perturbations was tested by analyzing the response to altered endogenous *CLV3* expression in small, discrete steps: the effectiveness of stemness-dependent *CLV3* expression is tested in a range from 10% to 620% in 10% steps. Varying *CLV3* levels from 90% to 620% compared to wildtype affected the size of the SCD, while OC cell number remained constant (see Fig. A.3 in Appendix A). This indicates that OC and SCD sizes are not strictly coupled, which has been also noted experimentally when analyzing the sizes of OC and SCD in plants grown under diverse environmental conditions [29].

Altering *WUS* expression

Lowering *WUS* expression levels reduces the sizes of both SCD and OC to a similar extent, and will cause a loss of both domains when *WUS* is fully repressed (see Fig. A.4 in Appendix A). Ectopic *WUS* expression was tested by changing the effect of *CLV3* signaling on *WUS* activity from repressive to activating. In plants, this has been achieved by expressing *WUS* from the *CLV3* promoter [10], which caused the coalescence of OC and SCD at the meristem tip, together with lateral expansion of this joint domain. This cell behavior is also observed in the presented simulations (Fig. 6.6).

Regeneration and De-Novo Generation of OC and SCD

After ablating SCD and OC from the meristem by pointing a laser beam at the meristem tip, *WUS* becomes expressed at the periphery, and the OC and SCD are regenerated with time [90], highlighting that cells at the periphery are capable of, but normally inhibited from the acquisition of OC identity. The laser ablation experiment was simulated starting from an equilibrated wildtype meristem, where all *WUS* or *CLV3* expressing cells are eliminated

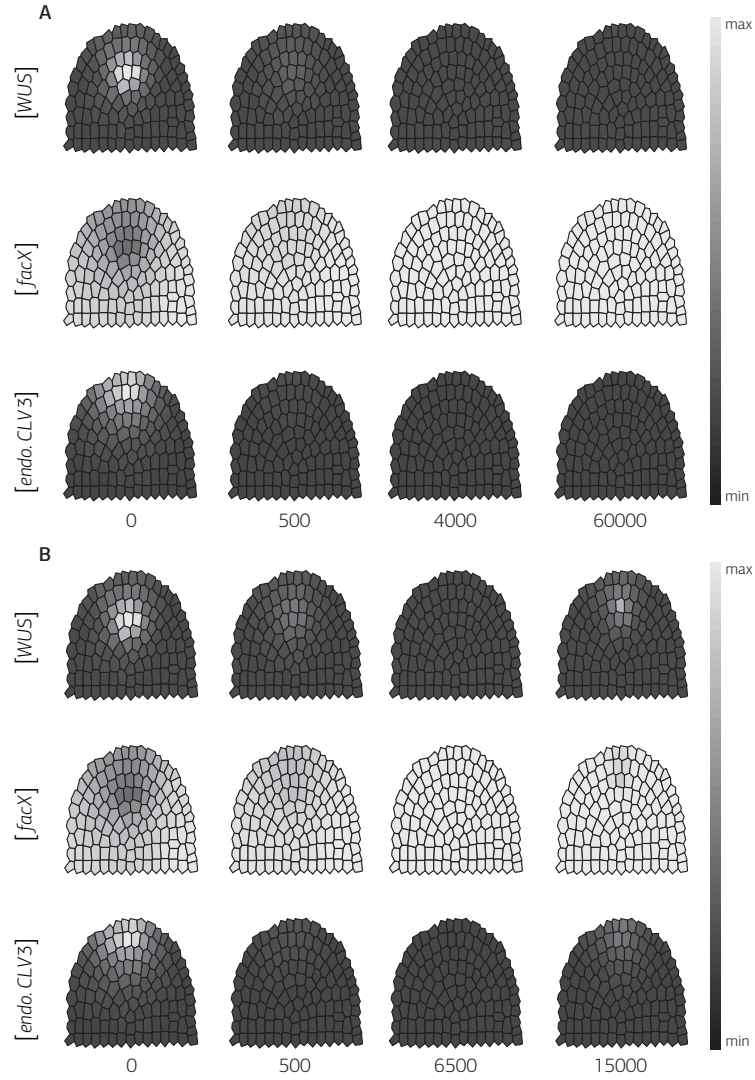

Figure 6.5. Response to *CLV3* overexpression in the entire meristem. Time-evolution of *WUS*, *FacX* and *CLV3* concentrations upon strong (A) or intermediate (B) level overexpression of *CLV3* is shown, starting from equilibrated wild-type meristem until the simulations reach a new equilibrium state. Here, [*endo. CLV3*] corresponds to *CLV3* expression from the endogenous promoter, taken as reporter for stem cell identity. Note that both, a *WUS* expression domain and stem cells are re-initiated in (B), but not in (A).

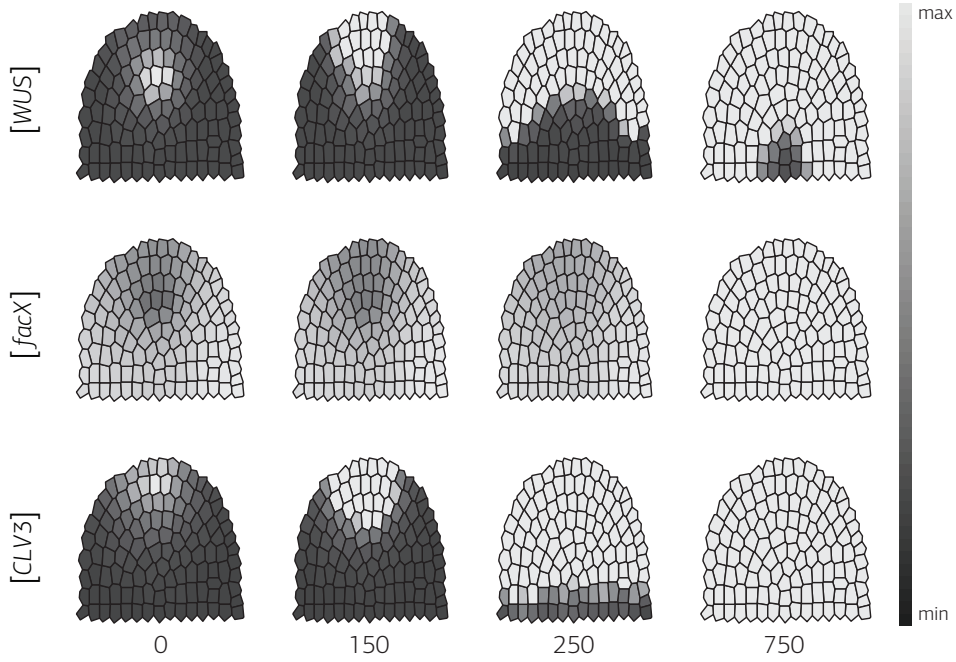

Figure 6.6. Simulation for *WUS* misexpression in the stem cell domain. Starting from an equilibrated wildtype meristem, misexpression of *WUS* from the *CLV3* promoter (*CLV3*≫*WUS*) is simulated until a new equilibrium state is reached. Cells in the meristem now acquire mixed identities and express both the OC and SCD marker genes.

from the meristem. It turned out that new OC was generated which induced a SCD nearby (see Fig. 6.7). This shows the self-generative capacity of the meristem model. During normal plant development, new meristems are generated during embryogenesis, flowering, and when axillary meristems are initiated. By simply altering *FacX* expression levels in a given cellular framework, it was possible to simulate the generation of a new OC and SCD, which coordinated increase in size when *FacX* is further upregulated (see Fig. A.5 in Appendix A).

Role of *FacX* and *WUS* Feedback Regulation

In addition to the already described scenarios that are all inspired by previ-

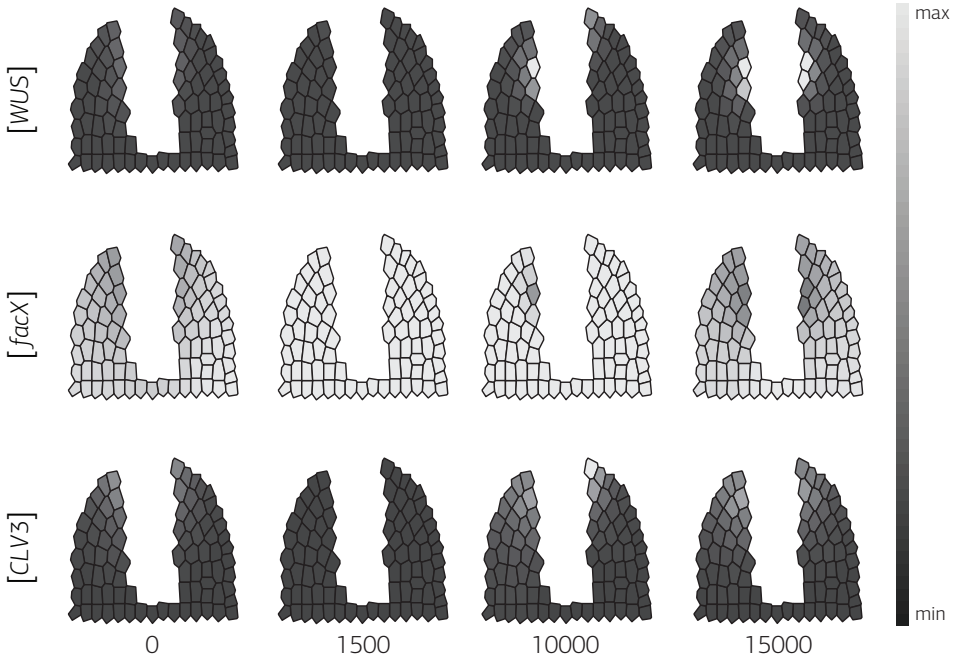

Figure 6.7. Simulation for a cell ablation scenario. The central region of the meristem was eliminated by virtual cell ablation. Note that even in the absence of an SCD and underlying OC, both domains can be partially restored.

ously conducted experiments, the role of *FacX* and the interaction of *WUS* with *FacX* is analyzed. Without feedback of *WUS* to *FacX*, *FacX* could be exchanged for a constant expression of *WUS*. This idea was tested by eliminating the feedback term of *WUS* on *FacX* (see Eq. 6.2). However, using an evolution strategy to search the parameter space of the resulting model (see Chapter 4), no parameter settings resulting in the desired system behavior could be found, showing separate but adjacent SCD and OC. In conclusion, *FacX* and the feedback between *WUS* and *FacX* are vital for the model to initiate stable Turing patterning. Candidates to realize the role of *FacX* are genes and functions that control *WUS* expression. Because eliminating *FacX* destabilizes the OC, mutations in genes contributing to *FacX*

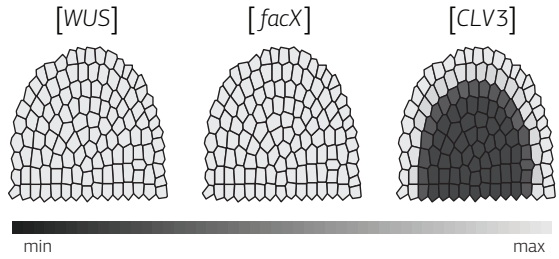

Figure 6.8 System neglecting feedback of *WUS* on *FacX*. Archetype system behavior for simulations without feedback of *WUS* on *FacX*: the meristem overproliferates, *WUS* is expressed in all meristem cells and all cells in the outer layers acquire stem cell identity indicated by their expression of the stem cell marker *CLV3*.

function should result in either unrestricted *WUS* expression and meristem overproliferation (see Fig. 6.8), or meristem arrest.

Influence of Anchoring Factor Distribution on System Behavior

Artificial positional information in form of an anchoring distribution (see Fig. 6.1) was introduced to stimulate SCD and OC positioning at the meristem tip. To test the possible influence of this anchoring distribution on the pattern formation capabilities of the model, the distribution is exchanged for a constant value. Although positioning of SCD and OC became more variable now, still parameter settings that result in spatially confined and adjacent SCD and OC (see Fig. 6.2) have been identify. This indicates that the anchoring distribution has no significant influence on the qualitative behavior of the proposed model.

6.4 · Discussion

The initial attempt in modeling the interplay between the functional domains SCD and OC was solely based on the interaction between two activator-inhibitor type systems ($WUS/CLV3$, and $WUS/FacX$) which were linked via WUS as the common node. Conceptually, the underlying assumption was that SCD and OC could originate independently of each other, but that their maintenance and relative position are controlled by mutual feedback

regulation. However, such a model failed to reproduce the domain arrangement observed in actual meristems for the model's parameter space which was explored using model calibration techniques presented in the previous two chapters. This indicates that an essential component was missing from this model. The most common outcome was not juxtaposition, but an overlap of the SCD with the OC. To improve the spatial separation of the two domains, it was additionally considered that cells within an actual meristem differ from each other by their position. Several misexpression experiments using *WUS* had previously shown that the cellular response to *WUS*-derived signals depends on a cells¥ relative position within the meristem, corresponding to its developmental trajectory. Only by adding spatial components to the model it was possible to achieve a realistic sizing and arrangement of the two domains within the meristem; removal of this spatial component causes extensive spatial overlapping of the SCD and OC.

The model was challenged by altering central system parameters to simulate mutant phenotypes and published transgenic experiments. In all experiments, the presented model proved to be robust against small-scale perturbations (see Section A.6 in Appendix A). This stability probably results from the combination of two feedback operated system, whereby one of them, the *WUS*/*FacX* system, acts as a buffer that dampens fluctuations in *WUS* levels. Furthermore, the reaction rates that influence *WUS* expression in the presented simulations are one order of magnitude smaller than those controlling $CLV3$ levels. Increased stability against signaling noise was uncovered in the analysis of coupled positive feedback systems if the two linked regulatory loops operated at different speeds [11, 12]. The model has revealed that combined feedback systems are sufficient to allow the generation and robust maintenance of two distinct cellular domains in a meristem, requiring only minimal assumptions about spatial restrictions of the system. The challenges ahead are now to extend the cell model into the third dimension, and incorporate cell divisions, but also to add other regulatory networks that control organ initiation and cell differentiation, approaching the goal of a virtual meristem.

7

Conclusions

This thesis is concerned with the modeling of GRNs that underly pattern formation process with respect to gene expression that control developmental processes. It addressed two different areas: (i) Building a model for the GRN underlying development and maintenance of the SAM in *A. thaliana*, and (ii) the development of methods for calibration of such models for which mostly qualitative data is available. With respect to these two topics it provides the following key contributions:

- A PDE model for the GRN responsible for SAM development and maintenance is set up. The model builds on RD systems as mechanism for pattern generation. It covers interactions between the two involved functional domains, SCD and OC, and modeling these signaling loops is capable of reproducing reported experiments with respect to perturbations of interactions between these two domains. The model is used to explore the parameter space underlying this regulative processes.

- A framework for parameter calibration of spatial DE models using only qualitative data is set up. It constitutes the model representation, choice of a suitable method for model simulation using numerical integration, choice of a heuristic for model calibration, and necessary interfaces between these components. In this regard, the major result is an investigation on necessary characteristics for a criterion to guide the calibration process using only qualitative data. Alongside with this study, methods for quantification of similarities between simulated and experimentally determined patterns are proposed.
- A method to incorporate analytical results concerning the pattern formation mechanism 'RD systems' is proposed that is capable of avoiding the problem of regions containing parameter settings that are indistinguishable in terms of qualitative model fit—plateaus which are resulting from typically non-linear interactions and otherwise complicate a directed calibration process and result in undesirable, rather random search behavior.
- A strategy is proposed to make use of transient states during developmental trajectories instead of only considering the final equilibrium system state. It is based on a hierarchical decomposition of respective processes with respect to the involved modules in the considered GRNs. This method is demonstrated on the example of the proposed SAM model. Combined with the developed scheme of when to consider certain stages during the calibration process, it is capable of considerably improving the convergence characteristics of the respective model calibration processes.

Thereby, this thesis provided a model for the regulative feedback loops that are central for development and maintenance of SAMs in *A. thaliana*. Together with the presented tools for model calibration, it forms a basis that can be extended to include subsystems responsible for further aspects of plant development. In consequence, it can be seen as a first step towards the ultimate goal of building a virtual meristem or even plant—a structure that would be a comprehensive knowledge-base for plants. At the same time, it implicitly provides insights into comparable systems for other species, e.g., stem cell regulation in insects like *Drosophila* or even mammalians. In addition, the proposed framework and calibration techniques themselves can be used for modeling other developmental systems.

7.1 · Future Perspectives

The results presented here form a sound basis for future research:

- The SAM model is based on a two-dimensional spatial domain representing an artificial longitudinal section through the three-dimensional dome forming the meristem. Although this two-dimensional domain is sufficient to provide insights concerning the pattern formation capabilities of a putative GRN, the model gains in authenticity when replacing the artificial two-dimensional domain with experimentally determined three-dimensional structures.
- The regulative processes considered in the proposed SAM model all happen in time-scales smaller than the cell cycle. Therefore it was sufficient to consider static tissue representations for modeling purposes. Following the aim of creating full virtual meristems, on the other hand it is necessary to include growth mechanics as well. Therefore, combining the proposed GRN model with a frame work simulating growth constitutes a promising research direction.
- The modeled $WUS/CLV3$ feedback loop is central for homeostasis of the stem cell pool in the SAM. Nevertheless, other regulative systems have some influence on this process as well as the existence of the stem cells impacts a range of structures related to it. It therefore is promising to add such additional regulative systems like those forming the boundary region of the meristem or AUXIN related systems. Here, the latter might have an influence on positioning of the SCD while the former probably depends on the existence of the SCD or at least depend on a common signal. Inclusion of such interactions would deepen the understanding of meristem regulation and would be a further step toward a virtual meristem.
- The reported studies concerning the development of methods for model calibration up to now focus on individual aspects of the overall task. A baseline calibration strategy has been designed and extensions like methods to incorporate different sources of domain knowledge or to for pattern recognition have been proposed. Although all single aspects have their respective use, combining the different ideas and techniques

into one integrated approach would provide a valuable addition for the growing field of modeling of developmental system.

- In addition, the proposed framework and used modeling techniques are suitable for other types of organisms. With a comparable long-term goal of creating virtual organisms, transferring the developed concepts provides the possibility to create comprehensive knowledge bases that can assist biologists in visualizing current knowledge and to design new experiments to deepen the understanding of these organisms.

A

Complementary Simulations on Shoot Apical Meristem Maintenance in *Arabidopsis thaliana*

In addition to the simulation results presented in Chapter 6: *Modeling Patterning in the SAM of* A. thaliana, here a set of complementary simulations are shown that further explore the respective model behavior.

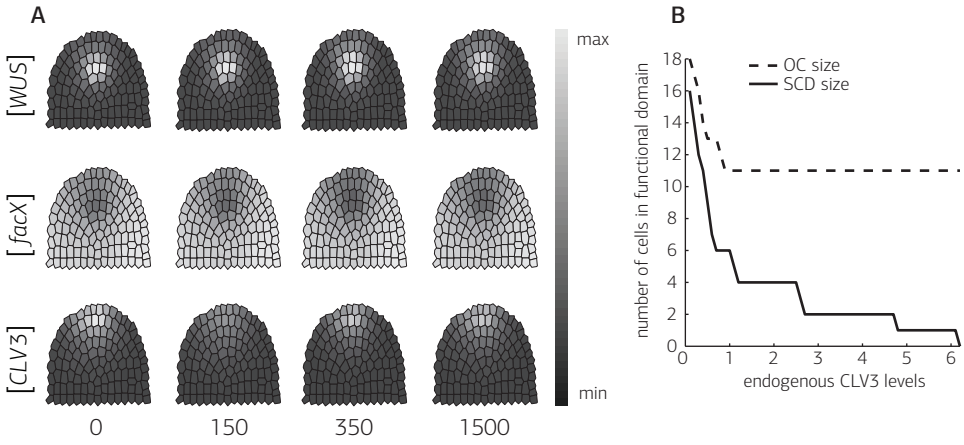

Figure A.1. Effects of reduced CLV3 levels on OC and SCD. A: Time course simulation for a conditionally reduced *CLV3* expression. B: Impact of *CLV3* expression levels on the sizes of OC and SCD. To assess the number of cells in the respective domains, the concentrations were discretized using thresholds relative to the wild type concentrations: For [st]: δ_{st}=0.21, for [WUS]: δ_{WUS}=0.31.

A.1 · Reducing *CLV3* Expression

Conditional loss-of-function of *CLV3* results in an expansion of the OC and the SCD. In addition, the OC is shifted towards the tip due to a loss of negative feedback by *CLV3*. When *CLV3* expression levels are only mildly reduced, a graded response was observed: SCD and OC enlarge, but the location of the OC is not affected due to active CLV3 signaling. A typical time course of a simulation for a conditional reduction in *CLV3* expression is shown in Fig. A.1. Interestingly, OC size remained unaffected for a wide range of endogenous *CLV3* expression levels (see Fig. A.1), indicating a partial uncoupling of OC and SCD sizes.

A.2 · Simulating Loss-Of-Function Mutants in *CLV1*

The phenotype of *clv1* loss-of-function mutants is comparable to a downregulation of *CLV3* expression. Mathematically, a *clv1* loss-of-function mutant

can be incorporated by extending Eqs. 6.1 and 6.2 with a factor $c1_{ko}$ that modulates the response of *WUS* to *CLV3* expression. The corresponding equations are given here:

$$\begin{aligned}\frac{\partial\,[\mathrm{WUS}]}{\partial t} =& D_{\mathrm{WUS}}\Delta\,[\mathrm{WUS}] \\ &+\xi\rho_{anc}\frac{[\mathrm{WUS}]^2\,[\mathrm{FacX}]}{1+(c1_{ko}\,[\mathrm{CLV3}]+[\mathrm{CLV3_{ext}}])} - \mu_{\mathrm{WUS}} + \sigma_{\mathrm{WUS}}, \\ \frac{\partial\,[\mathrm{FacX}]}{\partial t} =& D_{\mathrm{FacX}}\Delta\,[\mathrm{FacX}] \\ &+\xi\rho_{anc}\frac{[\mathrm{WUS}]^2\,[\mathrm{FacX}]}{1+(c1_{ko}\,[\mathrm{CLV3}]+[\mathrm{CLV3_{ext}}])} + \frac{\sigma_{\mathrm{FacX}}}{1+\frac{[\mathrm{FacX}]}{K_{\mathrm{FacX}}}}.\end{aligned}$$

By setting $c1_{ko}$ to values $c1_{ko} \in [0, 1]$, different residual activity levels of *CLV1* can be simulated, where $c1_{ko}$=0 represents a loss-of-function, and $c1_{ko}$=1 represents the wildtype scenario. Typical simulated time courses for graded *clv1* loss-of-function mutants are shown in Fig. A.2.

A.3 · Gradual Increase of Exogenous *CLV3* Expression Levels

In Section 6.3.2 simulations for meristems with strongly increased *CLV3* expression levels have been shown. Increased levels of *CLV3* resulted in meristem arrest, and intermediate overexpression levels caused an only transient loss of the SCD, which recovered with time. By studying a range of *CLV3* overexpression levels, it was uncovered that both domains are gradually reduced in size (see Fig. A.3). However, the SCD (solid line in Fig. A.3) responds more readily, indicating no strict interdependence of domain sizes.

A.4 · Altering *WUS* Expression

The consequences of reducing *WUS* expression to different levels have been explored: minor reductions of WUS levels cause a corresponding small re-

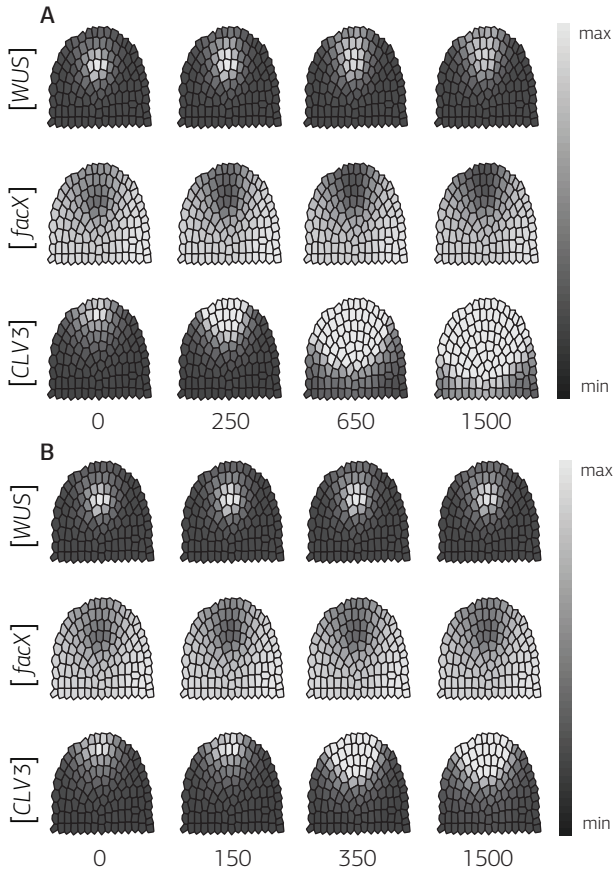

Figure A.2 Time courses for graded *clv1* loss-of-function mutant simulations; A: *CLV1* loss-of-function scenario ($c1_{ko}$=0), B: simulation where *CLV1* retains some activity ($c1_{ko}$=0.2).

duction in OC and SCD size (see Fig. A.4). A medium reduction causes a transient loss of SCD and OC until FacX reaches a level that allows to re-initiate a new meristem (see Fig. A.4). In case of a *WUS* knockout, the meristem is arrested without recovery (see Fig. A.4).

A.5 · Regeneration and De Novo Generation of OC and SCD

New meristems can be formed during emryogenesis or flowering. The proposed model allows simulating the de novo generation of meristems by vary-

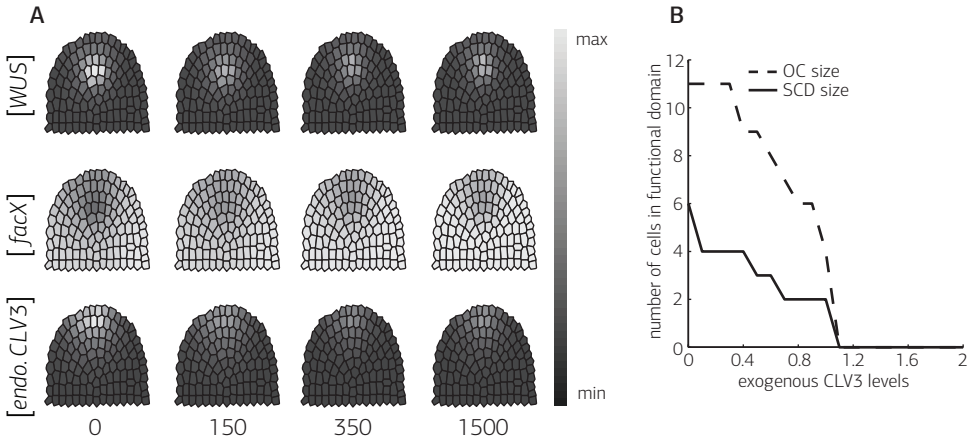

Figure A.3. Gradual response to exogenous *CLV3* expression. A: Time course simulation for a low level *CLV3* overexpression ($[CLV3_{ext}]$=0.2). B: Graded system response to different exogenous *CLV3* expression levels. The $[CLV3_{ext}]$ level is varied in $[CLV3_{ext}] \in [0, 2]$. To assess the number of cells in the respective domains, the concentrations of stemness and *WUS* were discretized using thresholds relative to the wildtype concentrations. For [st]: δ_{st}=0.21; for [WUS]: δ_{WUS}=0.31.

ing *FacX* regulation. Figure A.5 shows the system response to different levels of the kinetic constant K_{FacX} and the resulting sizes of OC and SCD. K_{FacX} is varied over a range $K_{\mathrm{FacX}} \in [0.1, 1]$, while K_{FacX}=0.4 is used for wild type simulations.

A.6 · Sensitivity Analysis

The sensitivity of the proposed model was investigated with respect to perturbations of single parameters of the wildtype parameter setting. The parameters have been perturbed using the following changes in absolute values: {-50%, -25%, -10%, 25%, 50%, 100%, 150%, 200%}. The corresponding simulations document a considerable robustness with respect to these perturbations: apart from varying the five parameters (ρ_{WUS}, μ_{WUS}, $\rho_{\mathrm{WUS_{sig}}}$, $\mu_{\mathrm{WUS_{sig}}}$, K_{st}), the simulated system response was qualitatively unchanged;

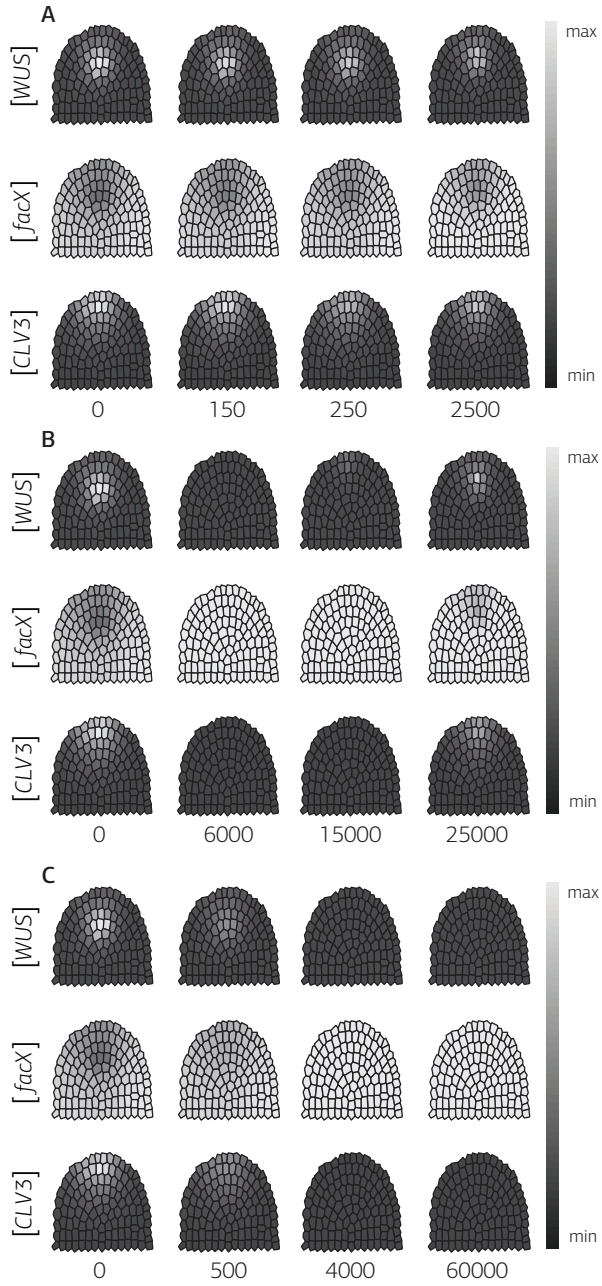

Figure A.4 Time courses for graded *WUS* knockout mutant simulations. A: Simulation for slightly reduced *WUS* expression level (w_{ko}=0.6); B: simulation for intermediate *WUS* expression (w_{ko}=0.4); C: simulation for a *WUS* loss-of-function scenario (w_{ko}=0).

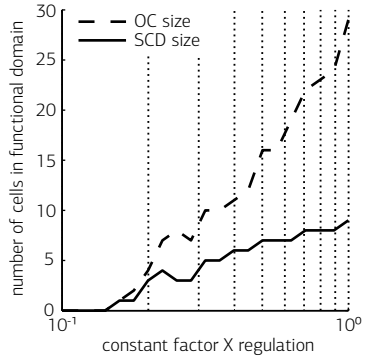

Figure A.5 Consequences of *FacX* expression levels for the sizes of OC and SCD. To assess the number of cells in the respective domains, concentrations were discretized using thresholds relative to the wild-type concentrations. For stemness: δ_{st}=0.21; for WUS: δ_{WUS}=0.31.

only slight variations in domain size and degree of separation between OC and SCD occurred.

ρ_{WUS} Variation

For a perturbation in ρ_{WUS} of -50%, OC and SCD are no longer spatially separated but both are located in the tip. For variations of +150% or +200%, the functional domains are enlarged but separated, still during simulations small numerical instabilities occurred indicating that ρ_{WUS} is reaching level that becomes intractable by the current simulation setup. For the remaining perturbations the system remained qualitatively stable.

μ_{WUS} Variation

For a perturbations of +150% and +200% the formed domains are no longer spatially separated and located in the meristem tip. For the remaining perturbations the system remained qualitatively stable.

$\rho_{WUS_{sig}}$ Variation

With growing $\rho_{\mathrm{WUS}_{\mathrm{sig}}}$ levels, domain sizes start to shrink and for perturbations of +100% to +200% simulations show small numerical instabilities. For the remaining perturbations the system remained qualitatively stable.

$\mu_{WUS_{sig}}$ Variation

For a perturbation of -50% domains are no longer spatially separated and

both are located in the tip. This simulation shows small numerical instabilities as well. For perturbation levels of +150% and +200% simulations again no longer show spatial separation between functional domains, this time the simulations have been numerically stable. For the remaining perturbations the system remained qualitatively stable.

K_{st} Variation

Only for a perturbation of -50%, the system shows small numerical instability, but for all tested perturbation levels the system remains qualitatively stable.

B

Models for Different Stages of Decomposed Shoot Apical Meristem Development

For the approach to model calibration based including information on the developmental trajectory of the respective system, the model for the test system stem cell homeostasis in the SAM of *A. thaliana* presented in Chapter 6 is decomposed into models for different stages of its development.

Before the newly introduced models for earlier developmental stages are presented, first the full model is repeated:

$$\begin{aligned}
\frac{\partial\,[\mathrm{WUS}]}{\partial t} =& D_s \Delta\,[\mathrm{WUS}] + w_{ko}\xi\rho_{anc}\frac{[\mathrm{WUS}]^2\,[\mathrm{FacX}]}{1+[\mathrm{CLV3}]^3} \\
&- \mu_{\mathrm{WUS}}\,[\mathrm{WUS}] + \sigma_{\mathrm{WUS}},] \\
\frac{\partial\,[\mathrm{FacX}]}{\partial t} =& D_q \Delta\,[\mathrm{FacX}] + w_{ko}\xi\rho_{anc}\frac{[\mathrm{WUS}]^2\,[\mathrm{FacX}]}{1+[\mathrm{CLV3}]^3} + \frac{\sigma_{\mathrm{FacX}}}{1+\frac{[\mathrm{FacX}]}{K_{\mathrm{FacX}}}}, \\
\frac{\partial\,[\mathrm{WUS_{sig}}]}{\partial t} =& D_q \Delta\,[\mathrm{WUS_{sig}}] + \rho_{\mathrm{WUS_{sig}}}\,[\mathrm{WUS}] - \mu_{\mathrm{WUS_{sig}}}\,[\mathrm{WUS_{sig}}]\,, \\
\frac{\partial\,[\mathrm{st}]}{\partial t} =& D_s \Delta\,[\mathrm{st}] + \mathbf{1}_{id}\,(i)\,\rho_{\mathrm{st}}\frac{\left(\frac{[\mathrm{WUS_{sig}}]}{K_{\mathrm{st}}}\right)^5}{1+\left(\frac{[\mathrm{WUS_{sig}}]}{K_{\mathrm{st}}}\right)^5} - \mu_{\mathrm{st}}\,[\mathrm{st}]\,, \\
\frac{\partial\,[\mathrm{CLV3}]}{\partial t} =& D_q \Delta\,[\mathrm{CLV3}] + c_{ko}\rho_{\mathrm{CLV3}}\,[\mathrm{st}] - \mu_{\mathrm{CLV3}}\,[\mathrm{CLV3}]\,.
\end{aligned}$$

Here, the D_q and D_s are diffusion constants which are assumed to be similar for the quickly diffusing components ([FacX], [$\mathrm{WUS_{sig}}$], and [CLV3]) and for the local components ([WUS] and [st]) that still underly some weak leakage diffusion. The ρ_i with $i \in \{anc, \mathrm{WUS}, \mathrm{FacX}, \mathrm{WUS_{sig}}, \mathrm{st}, \mathrm{CLV3}\}$ are reaction rates that are equal in all cells with one exception being ρ_{anc}, the so called anchoring distribution that is necessary in the model to compensate for missing structural information guiding exact positioning of the functional domains active in the meristem. μ_i with $i \in \{anc$, WUS, FacX, $\mathrm{WUS_{sig}}$, st,CLV3$\}$ are degradation rates, σ_{WUS} and σ_{FacX} are terms representing basal expression, a characteristic function $\mathbf{1}_{id}$ determining stem cell competent layers, and a random perturbation $\xi \in$ [-0.025, 0.025] necessary to initiate pattern formation. An overview on feasibility intervals used for the different types of parameters is given in Tab. B.1.

$$\mathbf{1}_{id}(i) = \begin{cases} 1, & \text{if cell } i \in \{L1, L2\} \\ 0, & \text{else} \end{cases}$$

Table B.1. Feasibility ranges used for different model parameters during optimization.

Parameter	D	ρ	μ	σ	K
Range	[0.0001, 0.5]	[0.01, 1.5]	[0.0001, 0.1]	[0.0001, 0.01]	[0.001, 1]

During the first stage of SAM development, only the so called OC is formed. To initiate this domain of *WUS* expressing cells only the model entities [WUS] and [FacX] need to be considered resulting in a reduced model considered in this stage given by the following equations:

$$\begin{aligned}\frac{\partial [\mathrm{WUS}]}{\partial t} =& D_s \Delta [\mathrm{WUS}] + \xi \rho_{anc} \rho_{\mathrm{WUS}} [\mathrm{WUS}]^2 [\mathrm{FacX}] \\ & - \mu_{\mathrm{WUS}} [\mathrm{WUS}] + \sigma_{\mathrm{WUS}}, \\ \frac{\partial [\mathrm{FacX}]}{\partial t} =& D_q \Delta [\mathrm{FacX}] - \xi \rho_{anc} \rho_{\mathrm{FacX}} [\mathrm{WUS}]^2 [\mathrm{FacX}] + \frac{\sigma_{\mathrm{FacX}}}{1 + \frac{[\mathrm{FacX}]}{K_{\mathrm{FacX}}}}.\end{aligned}$$

Moving on to the second developmental stage, the model of stage one is extended components realizing the formation of a SCD, yet leaving out the feedback from SCD to the OC. The model for stage two is given be the following equations:

$$\begin{aligned}\frac{\partial [\mathrm{WUS}]}{\partial t} =& D_s \Delta [\mathrm{WUS}] + w_{ko} \xi \rho_{anc} [\mathrm{WUS}]^2 [\mathrm{FacX}] \\ & - \mu_{\mathrm{WUS}} [\mathrm{WUS}] + \sigma_{\mathrm{WUS}}, \\ \frac{\partial [\mathrm{FacX}]}{\partial t} =& D_q \Delta [\mathrm{FacX}] + w_{ko} \xi \rho_{anc} [\mathrm{WUS}]^2 [\mathrm{FacX}] + \frac{\sigma_{\mathrm{FacX}}}{1 + \frac{[\mathrm{FacX}]}{K_{\mathrm{FacX}}}}, \\ \frac{\partial [\mathrm{WUS_{sig}}]}{\partial t} =& D_q \Delta [\mathrm{WUS_{sig}}] + \rho_{\mathrm{WUS_{sig}}} [\mathrm{WUS}] - \mu_{\mathrm{WUS_{sig}}} [\mathrm{WUS_{sig}}], \\ \frac{\partial [\mathrm{st}]}{\partial t} =& D_s \Delta [\mathrm{st}] + \mathbf{1}_{id}(i) \rho_{\mathrm{st}} \frac{\left(\frac{[\mathrm{WUS_{sig}}]}{K_{\mathrm{st}}}\right)^5}{1 + \left(\frac{[\mathrm{WUS_{sig}}]}{K_{\mathrm{st}}}\right)^5} - \mu_{\mathrm{st}} [\mathrm{st}].\end{aligned}$$

The SAM simulations are done using numerical integration of the respective systems. For the integration the IMEX scheme based on a modified Crank-

Nicolson integrator combined with an Adams-Bashford scheme [91] is used. As spatial discretization for the Crank-Nicolson scheme a grid of cellular resolution is used and time is discretized using a constant time steps δ_t=0.5 of dimensionless time. Each simulation lasts for 120000 time steps.

Bibliography

[1] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. *Molecular Biology Of The Cell*. Garland, New York, NY, USA, 3rd edition, 1994.

[2] K. Amonlirdviman, N. A. Khare, D. R. Tree, W. S. Chen, J. D. Axelrod, and C. J. Tomlin. Mathematical Modeling of Planar Cell Polarity to Understand Domineering Nonautonomy. *Science*, 307(5708):423–426, 2005.

[3] A. Auger and N. Hansen. Performance evaluation of an advanced local search evolutionary algorithm. In *Congress on Evolutionary Computation (CEC 2005)*, volume 2, pages 1777–1784, Piscataway, NJ, USA, 2005. IEEE Press.

[4] J. Bader and E. Zitzler. HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization. TIK Report 286, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, November 2008.

[5] N. Beume and G. Rudolph. Faster S-Metric Calculation by Considering Dominated Hypervolume as Klee's Measure Problem. Technical Report CI-216/06, Sonderforschungsbereich 531 Computational Intelligence, Universität Dortmund, 2006. shorter version published at IASTED International Conference on Computational Intelligence (CI 2006).

[6] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, UK, 1995.

[7] S. Bornholdt. Boolean network models of cellular regulation: prospects and limitations. *J R Soc Interface*, 5(Suppl. 1):S85–S94, 2008.

[8] D. Bouyer, F. Geier, F. Kragler, A. Schnittger, M. Pesch, K. Wester, R. Balkunde, J. Timmer, C. Fleck, and M. Hülskamp. Two-Dimensional Patterning by a Trapping/Depletion Mechanism: The Role of TTG1 and GL3 in Arabidopsis Trichome Formation. *PLoS Biol*, 6(6): e141, 2008.

[9] U. Brand, J. C. Feltscher, M. Hobe, E. M. Meyerowitz, and R. Simon. Dependendence of Stem Cell Fate in Arabidopsis on a Feedback Loop Regulated by CLV3 Activity. *Science*, 289(5479):617–619, 2000.

[10] U. Brand, M. Grünewald, M. Hobe, and R. Simon. Regulation of CLV3 Expression by Two Homeobox Genes in Arabidopsis. *Plant Physiol*, 129(2):565–575, 2002.

[11] O. Brandman and T. Meyer. Feedback Loops Shape Cellular Signals in Space and Time. *Science*, 322(5900):390–395, 2008.

[12] O. Brandman, J. E. Ferrell, Jr., R. Li, and T. Meyer. Interlinked Fast and Slow Positive Feedback Loops Drive Reliable Cell Decisions. *Science*, 310(5747):496–498, 2005.

[13] K. Bringmann and T. Friedrich. Approximating the Volume of Unions and Intersections of High-Dimensional Geometric Objects. In S. H. Hong, H. Nagamochi, and T. Fukunaga, editors, *International Symposium on Algorithms and Computation (ISAAC 2008)*, volume 5369 of *LNCS*, pages 436–447, Berlin, Germany, 2008. Springer.

[14] D. Brockhoff, T. Friedrich, N. Hebbinghaus, C. Klein, F. Neumann, and E. Zitzler. Do Additional Objectives Make a Problem Harder? In D. Thierens et al., editors, *Genetic and Evolutionary Computation Conference (GECCO 2007)*, pages 765–772, New York, NY, USA, 2007. ACM Press.

[15] T. A. Brown. *Genomes*, chapter 6. Wiley-Liss, 2nd edition, 2002.

[16] M. J. Buck and J. D. Lieb. ChIP-chip: considerations for the design, analysis and application of genome-wide chromatin immunoprecipitation experiments. *Genomics*, 83(3): 349–360, 2004.

[17] S. E. Clark, R. W. Williams, and E. M. Meyerowitz. The CLAVATA1 Gene Encodes a Putative Receptor Kinase That Controls Shoot and Floral Meristem Size in Arabidopsis. *Cell*, 89 (4):575–585, 1997.

[18] W. J. Conover. *Practical Nonparametric Statistics*. John Wiley, 3rd edition, 1999.

[19] H. de Jong. Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *J Comput Biol*, 9(1):67–103, 2002.

[20] K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley, Chichester, UK, 2001.

[21] B. Domon and R. Aebersold. Mass Spectrometry and Protein Analysis. *Science*, 312 (5571):212–217, 2006.

[22] D. Endy and R. Brent. Modelling cellular behaviour. *Nature*, 409(6818):391–395, 2001.

[23] O. Fiehn. Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol*, 48(1–2):155–171, 2002.

[24] J. C. Fletcher, U. Brand, M. P. Running, R. Simon, and E. M. Meyerowitz. Signaling of Cell Fate Decisions by CLAVATA3 in Arabidopsis Shoot Meristems. *Science*, 283(5409):

1911–1914, 1999.

[25] J. A. Foster. Evolutionary Computation. *Nat Rev Genet*, 2(6):428–436, 2001.

[26] N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using Bayesian Networks to Analyze Expression Data. *J Comput Biol*, 7(3–4):601–620, 2000.

[27] A. Garg, K. Mohanram, A. Di Cara, G. De Micheli, and I. Xenarios. Modeling stochasticity and robustness in gene regulatory networks. *Bioinformatics*, 25(12):i101–109, 2009.

[28] M. R. Garvie. Finite-Difference Schemes for Reaction-Diffusion Equations Modeling Predator-Prey Interactions in MATLAB. *B Math Biol*, 69(3):931–956, 2007.

[29] F. Geier, J. U. Lohmann, M. Gerstung, A. T. Maier, J. Timmer, and C. Fleck. A Quantitative and Dynamic Model for Plant Cell Regulation. *PLoS ONE*, 3(10):e3553, 2008.

[30] A. Gierer. Generation of Biological Patterns and Form: Some Physical, Mathematical, and Logical Aspects. *Prog Biophys Mol Biol*, 37(1):1–47, 1982.

[31] A. Gierer and H. Meinhardt. A theory of biological pattern formation. *Kybernetik*, 12(1): 30–39, 1972.

[32] A. Gierer, S. Berking, H. Bode, C. N. David, K. Flick, G. Hansmann, H. Schaller, and E. Trekner. Regeneration of hydra from reaggregated cells. *Nat New Bio*, 239(91):98–101, 1972.

[33] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem-US*, 81(25):2340–2361, 1977.

[34] D. T. Gillespie. The chemical Langevin equation. *J Phys Chem-US*, 113(1):297–306, 2000.

[35] S. P. Gordon, M. G. Heisler, G. V. Reddy, C. Ohno, and E. M. Meyerowitz. Pattern formation during de novo assembly of the Arabidopsis shoot meristem. *Development*, 134(19): 3539–3548, 2007.

[36] P. M. Granitto, P. F. Verdes, and H. A. Cecatto. Neural network ensembles: evaluation of aggregation algorithms. *Artif Intell*, 163(2):139–162, 2005.

[37] R. Groß-Hardt, M. Lenhard, and T. Laux. WUSCHEL signaling functions in interregional communication during Arabidopsis ovule development. *Gene Dev*, 16(21):1129–1138, 2002.

[38] R. Haberman. *Elementary Applied Differential Equations with Fourier Series and Boundary Value Problems*. Prentice Hall, Englewood Cliffs, NJ, USA, 2nd edition, 1987.

[39] P. Haffter and C. Nüsslein-Volhard. Large scale genetics in a small vertebrate, the zebrafish. *Int J Dev Biol*, 40(1):221–227, 1996.

[40] P. Han, Q. Li, and Y.-X. Zhu. Mutation of Arabidopsis BARD1 Causes Meristem Defects by Failing to Confine WUSCHEL Expression to the Organizing Center. *Plant Cell*, 20(6): 1482–1493, 2008.

[41] J. Handl and J. Knowles. An Evolutionary Approach to Multiobjective Clustering. *IEEE T Evolut Comput*, 11(1):56–76, 2007.

[42] J. Handl, S. Lovell, and J. Knowles. Multiobjectivization by Decomposition of Scalar Cost Functions. In G. Rudolph et al., editors, *Conference on Parallel Problem Solving From Nature (PPSN X)*, volume 5199 of *LNCS*, pages 31–40. Springer, 2008.

[43] J. Handl, S. C. Lovell, and J. Knowles. Investigations into the Effect of Multiobjectivization in Protein Structure Prediction. In G. Rudolph et al., editors, *Conference on Parallel Problem Solving From Nature (PPSN X)*, volume 5199 of *LNCS*, pages 702–711. Springer, 2008.

[44] N. Hansen and A. Ostermeier. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evol Comput*, 9(2):159–195, 2001.

[45] W. Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *J Am Stat Assoc*, 58(301):13–30, 1963.

[46] T. Hohm and E. Zitzler. Modeling the Shoot Apical Meristem in A. thaliana: Parameter Estimation for Spatial Pattern Formation. In Elena Marchiori, Jason H. Moore, and Jagath C. Rajapakse, editors, *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics (evoBIO 2007)*, volume 4447 of *LNCS*, pages 102–113. Springer, 2007.

[47] T. Hohm and E. Zitzler. Multiobjectivization for parameter estimation: a case-study on the segment polarity network of drosophila. In F. Rothlauf et al., editors, *GECCO '09: Genetic and Evolutionary Computation Conference (GECCO 2009)*, pages 209–216, New York, NY, USA, 2009. ACM.

[48] T. Hohm and E. Zitzler. A Multiobjective Evolutionary Algorithm for Numerical Parameter Space Characterization of Reaction Diffusion Systems. In V. Kadirkamanathan et al., editors, *International Conference on Pattern Recognition in Bioinformatics (PRIB 2009)*, volume 5780 of *LNBI*, pages 162–174, Heidelberg, Germany, 2009. Springer.

[49] T. Hohm and E. Zitzler. Multicellular Pattern Formation: Parameter Estimation for ODE-based Gene Regulatory Network Models. *IEEE Eng Med Biol*, 28(4):52–57, 2009.

[50] R. Horst, P. M. Pardalos, and N. Van Thoai. *Introduction to Global Optimization*. Springer, Berlin, Germany, 1st edition, 1995.

[51] A. Hunding, S.A. Kauffman, and B.C. Goodwin. Drosophila Segmentation: Supercomputer Simulation of Prepattern Hierarchy. *J Theor Biol*, 145(3):369–384, 1990.

[52] C. Igel, N. Hansen, and S. Roth. Covariance Matrix Adaptation for Multi-objective Optimization. *Evol Comput*, 15(1):1–28, 2007.

[53] M. T. Jensen. Helper-Objectives: Using Multi-Objective Evolutionary Algorithms for Single-Objective Optimisation. *Journal of Mathematical Modelling and Algorithms*, 3(4): 323–347, 2004.

[54] S. Jeong, A. E. Trotochaud, and S. E. Clark. The Arabidopsis CLAVATA2 Gene Encodes a Receptor-like Protein Required for the Stability of the CLAVATA1 Receptor-like Kinase. *Plant Cell*, 11(10):1925–1934, 1999.

[55] H. J. Jiang, M. A. A. S. Choudhury, and S. L. Shah. Detection and diagnosis of plant-wide oscillations from industrial data using the spectral envelope method. *J Process Contr*, 17 (2):143–155, 2007.

[56] H. Jönsson, B. E. Shapiro, E. M. Meyerowitz, and E. Mjolsness. Signalling in multicellular models of plant development. In S. Kumar and J. Bentley, editors, *On growth, Form and Computers*, pages 156–161. Elsevier, Amsterdam, NL, 2003.

[57] H. Jönsson, M. Heisler, G. V. Reddy, V. Agrawal, V. Gor, B. E. Shapiro, E. Mjolsness, and E. M. Meyerowitz. Modeling the organization of the WUSCHEL expression domain in the shoot apical meristem. *Bioinformatics*, 21(Suppl. 1):i232–i240, 2005.

[58] S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol*, 22(3):437–467, 1969.

[59] J. Knowles, R. Watson, and D. Corne. Reducing Local Optima in Single-Objective Problems by Multi-objectivization. In E. Zitzler et al., editors, *Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, volume 1993 of *LNCS*, pages 269–283, Berlin, 2001. Springer.

[60] E. Knust. Drosophila morphogenesis: Movement behind the edge. *Curr Biol*, 7(9): R558–R561, 1997.

[61] A. J. Koch and H. Meinhardt. Biological pattern formation: from basic mechanisms to complex structures. *Rev Mod Phys*, 66(4):1481–1510, 1994.

[62] I. S. Kohane, A. T. Kho, and A. J. Butte. *Microarrays for an Integrative Genomics*. MIT Press, 2003.

[63] T. Kondo, S. Sawa, A. Kinoshita, S. Mizuno, T. Kakimoto, H. Fukuda, and Y. Sakagami. A Plant Peptide Encoded by CLV3 Identified by in Situ MALDI-TOF MS Analysis. *Science*, 313(5788):845–848, 2006.

[64] B. A. Krizek and J. C. Fletcher. Molecular mechanisms of flower development: an armchair guide. *Nat Rev Genet*, 6(9):688–698, 2005.

[65] C. S. Kwon, C. Chen, and D. Wagner. WUSCHEL is a primary target for transcriptional regulation by SPLAYED in dynamic control of stem cell fate in Arabidopsis. *Genes Dev*, 19(8):992–1003, 2005.

[66] T. Laux, K. F. X. Mayer, J. Berger, and G. Jürgens. The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis. *Development*, 122(1):87–96, 1996.

[67] A. Leibfried, J. P. C. To, W. Busch, S. Stehling, A. Kehle, M. Demar, J. J. Kieber, and J. U. Lohmann. WUSCHEL controls meristem function by direct regulation of cytokinin-inducible response regulators. *Nature*, 438(7071):1172–1175, 2005.

[68] K. F. X. Mayer, H. Schoof, A. Haecker, M. Lenhard, G. Jürgens, and T. Laux. Role of WUSCHEL in Regulating Stem Cell Fate in the Arabidopsis Shoot Meristem. *Cell*, 95(6): 805–815, 1998.

[69] H. Meinhardt. *Models Of Biological Pattern Formation*. Academic Press, London, UK, 1982.

[70] H. Meinhardt. A model for pattern-formation of hypostome, tentacles and foot in Hydra: how to form structures close to each other, how to form them at distance. *Dev Biol*, 157 (2):321–333, 1993.

[71] H. Meinhardt. *The algorithmic beauty of sea shells*. Springer, New York, NY, USA, 1995.

[72] H. Meinhardt and A. Gierer. Generation and Regeneration of Sequence and Structure During Morphogenesis. *J Theor Biol*, 85(3):429–450, 1980.

[73] P. Mendes and D. B. Kell. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14(10):869–883, 1998.

[74] K. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer, Boston, MA, USA, 1999.

[75] A. Migdalas, P. M. Pardalos, and P. Värbrand. *Multilevel Optimization: Algorithms and Applications*. Springer, Berlin, Germany, 1st edition, 1997.

[76] E. Mjolsness, D. H. Sharp, and J. Reinitz. Connectionist Model of Development. *J Theor Biol*, 152(4):429–453, 1991.

[77] C. G. Moles, P. Mendes, and J. R. Banga. Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods. *Genome Res*, 13(11):2467–2474, 2003.

[78] R. Müller, L. Borghi, D. Kwiatkowska, P. Laufs, and R. Simon. Dynamic and Compensatory Responses of Arabidopsis Shoot and Floral Meristems to CLV3 Signaling. *Plant Cell*, 18 (5):1188–1198, 2006.

[79] R. Müller, A. Bleckmann, and R. Simon. The Receptor Kinase CORYNE of Arabidopsis Transmits the Stem Cell-Limiting Signal CLAVATA3 Independently of CLAVATA1. *Plant Cell*, 20(4):934–946, 2008.

[80] J. D. Murray. *Mathematical Biology*. Springer, New York, NY, USA, 2003.

[81] I. T. Nabney. *NETLAB: Algorithms for Pattern Recognition*. Advances in Pattern Recognition. Springer, Oxford, UK, 2nd edition, 2003.

[82] A. Nakamasu, G. Takahashi, A. Kanabe, and S. Kondo. Interactions between zebrafish pigment cells responsible for the generation of Turing patterns. *Proc Natl Acad Sci USA*, 106(21):8429–8434, 2009.

[83] F. Neumann and I. Wegener. Minimum Spanning Trees Made Easier Via Multi-Objective Optimization. *Natural Computing*, 5(3):305–319, 2006.

[84] M. Ogawa, H. Shinohara, Y. Sakagami, and Y. Matsubayashi. Arabidopsis CLV3 Peptide Directly Binds CLV1 Ectodomain. *Science*, 319(5861):294, 2008.

[85] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes*. Cambridge University Press, Cambridge, UK, 3rd edition, 2007.

[86] I. Prigogine and R. Lefever. Symmetry Breaking Instabilities in Dissipative Systems. *J Chem Phys*, 48(4):1695–1700, 1968.

[87] R. Quast, R. Baade, and D. Reimers. Evolution Strategies applied to the problem of line profile decomposition in QSO spectra. *Astron Astrophys*, 431(3):1167–1175, 2005.

[88] R. Raffard, K. Amonlirdviman, J. D. Axelrod, and C. J. Tomlin. Automatic parameter identification via the adjoint method, with application to understanding planar cell polarity. In *IEEE Conference on Decision and Control (CDC 2006)*, pages 13–18, Piscataway, NJ, USA, 2006. IEEE Press.

[89] G. V. Reddy and E. M. Meyerowitz. Stem-Cell Homeostasis and Growth Dynamics Can Be Uncoupled in the Arabidopsis Shoot Apex. *Science*, 310(5748):663–667, 2005.

[90] D. Reinhardt, M. Frenz, T. Mandel, and C. Kuhlemeier. Microsurgical and laser ablation analysis of interactions between the zones and layers of the tomato shoot apical meristem. *Development*, 130(17):4073–4083, 2003.

[91] S. J. Ruuth. Implicit-explicit methods for reaction-diffusion problems in pattern formation. *J Math Biol*, 34(2):148–176, 1995.

[92] U. Sauer. High-throughput phenomics: experimental methods for mapping fluxomes. *Curr Opin Biotechnol*, 15(1):58–63, 2004.

[93] M. A. Savageau. Allometric morphogenesis of complex systems: Derivation of the basic equations from first principles. *Proc Natl Acad Sci USA*, 76(12):6023–6025, 1979.

[94] H. Schoof, M. Lenhard, A. Haecker, K. F. X. Mayer, G. Jürgens, and T. Laux. The Stem Cell Population of Arabidopsis Shoot Meristems is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes. *Cell*, 100(6):635–644, 2000.

[95] L. A. Segel and J. L. Jackson. Dissipative structure: An explanation and an ecological example. *J Theor Biol*, 37(3):545–559, 1972.

[96] A. J. C. Sharkey, editor. *Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems*. Springer, London, UK, 1999.

[97] D. R. Smyth. Flower development. *Curr Biol*, 11(3):R82–R84, 2001.

[98] Y. Stahl and R. Simon. Plant stem cell niches. *Int J Dev Biol*, 49(5–6):479–489, 2005.

[99] D. S. Stoffer, D. E. Tyler, and A. J. McDougall. Spectral analysis for categorical time series: Scaling and the spectral envelope. *Biometrika*, 80(3):611–622, 1993.

[100] D. S. Stoffer, D. E. Tyler, and D. A. Wendt. The Spectral Envelope and Its Applications. *Stat Sci*, 15(3):224–253, 2000.

[101] S. H. Strogatz. *Nonlinear Dynamics and Chaos*. Studie in Nonlinearity. Westview, Cambridge, MA, USA, 2000.

[102] R. Thomas. Regulatory networks seen as asynchronous automata: A logical description. *J Theor Biol*, 153(1):1–23, 1991.

[103] C. J. Tomlin and D. Axelrod. Biology by numbers: mathematical modelling in developmental biology. *Nat Rev Genet*, 8(5):331–340, 2007.

[104] M. R. Tucker, A. Hinze, E. J. Tucker, S. Takada, G. Jürgens, and T. Laux. Vascular signalling mediated by ZWILLE potentiates WUSCHEL function during shoot meristem stem cell development in the Arabidopsis embryo. *Development*, 135(17):2839–2843, 2008.

[105] A. Turing. The chemical basis for morphogenesis. *Philos Trans R Soc Lond, B*, 237(641): 37–72, 1952.

[106] E. O. Voit. *Computational Analysis of Biochemical Systems.* Cambridge University Press, Cambridge, UK, 2000.

[107] G. von Dassow, E. Meir, E. M. Munro, and G. M. Odell. The segment polarity network is a robust developmental module. *Nature*, 406(6792):188–192, 2000.

[108] M. Yamaguchi, E. Yoshimoto, and S. Kondo. Pattern regulation in the stripe of zebrafish suggests an underlying dynamic and autonomous mechanism. *Proc Natl Acad Sci USA*, 104(12):4790–4793, 2007.

[109] Y. Zhao, L. Medrano, K. Ohashi, J. C. Fletcher, H. Yu, H. Sakai, and E. M. Meyerowitz. HANABA TARANU Is a GATA Transcription Factor That Regulates Shoot Apical Meristem and Flower Development in Arabidopsis. *Plant Cell*, 16(10):2586–2600, 2004.

[110] E. Zitzler and L. Thiele. Multiobjective Optimization Using Evolutionary Algorithms - A Comparative Case Study. In *Conference on Parallel Problem Solving from Nature (PPSN V)*, pages 292–301, Amsterdam, 1998.

[111] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca. Performance Assessment of Multiobjective Optimizers: An Analysis and Review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.

[112] E. Zitzler, L. Thiele, and J. Bader. SPAM: Set Preference Algorithm for Multiobjective Optimization. In G. Rudolph et al., editors, *Conference on Parallel Problem Solving From Nature (PPSN X)*, volume 5199 of *LNCS*, pages 847–858. Springer, 2008.

Acronyms

SAM	shoot apical meristem
A. thaliana	*Arabidopsis thaliana*
GRN	gene regulative network
CMA-ES	Covariance Matrix Adaption Evolution Strategy
MO-CMA-ES	Multiobjective Covariance Matrix Adaption Evolution Strategy
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
mRNA	messenger RNA
DE	differential equation
ODE	ordinary differential equation
PDE	partial differential equation
RD	reaction-diffusion
IMEX	implicit-explicit
SCD	stem cell domain
OC	organizing center
WUS	*WUSCHEL*
WUS	WUSCHEL
CLV	*CLAVATA*
CLV	CLAVATA
CLV1	*CLAVATA1*
clv1	*clavata1*
CLV2	*CLAVATA2*

CLV3	*CLAVATA3*
clv3	*clavata3*
CLV3	CLAVATA3
FacX	*factor X*
FacX	factor X
LRR	leucine rich repeat
CRN	*CORYNE*
ARR	*ARABIDOPSIS RESPONSE REGULATOR*
GFP	green fluorescent protein
SYD	*SPLAYED*
SNF2	sucrose non-fermenting 2
ATPase	adenosine triphosphatase
BARD1	BRCA1 associated RING domain 1
HAN	*HANABA TARANU*
WG	*wingless*
HH	*hedgehog*
EN	*engrailed*
CC	connected component
ANN	artificial neural network

Curriculum Vitae

Personal Information

Tim Hohm
Born April 8, 1979 in Hannover, Germany
Citizen of Germany

Education

2006–2009	doctoral student at Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland
2004–2006	research assistant at center of advanced european studies and research (caesar), Bonn, Germany
1999–2003	diploma studies in computer science (Dipl.-Inform.) at the University of Dortmund, Germany
1998–1999	alternative civilian service (Zivildienst), Arbeiter Samariter Bund, Essen, Germany
1998	Abitur at Helmholtzschule, Essen, Germany