# Multiscale modelling of Arabidopsis thaliana phyllotaxis perturbation

Yassin Refahi

Université Montpellier 2
Science et Technique du Languedoc

# Thèse

pour obtenir le grade de

Docteur de l'Unversité Montpellier 2

Discipline : Informatique
Ecole Doctorale : Informations, Structures, Systems (I2S)

Par

Yassin REFAHI

# Modélisation multi-échelle des perturbations de la phyllotaxie d'Arabidopsis thaliana

Sous la direction de:

Christophe Godin
Jan Traas
Etienne Farcot

Thèse soutenue le 15 novembre 2011

**Jury :**

| | | | | |
|---|---|---|---|---|
| M. | Eric RIVALS, | - | CNRS | , Président |
| M. | Jean-Luc Gouzé | - | INRIA | , Rapporteur |
| Mme. | Macha Nikolski | - | CNRS | , Rapporteur |
| M. | Yves Caraglio | - | CIRAD | , Examinateur invité |
| M. | Christophe Godin | - | INRIA | , Directeur |
| M. | Jan TRAAS | - | INRA | , Co-directeur |
| M. | Etienne Farcot | - | INRIA | , Encadrant |

Université Montpellier 2
Science et Technique du Languedoc

# Thesis

submitted for the degree of

Doctor of Philosophy

In

Computer Science

by

Yassin REFAHI

# Multiscale modelling of Arabidopsis thaliana phyllotaxis perturbations

advisors:

Christophe Godin
Jan Traas
Etienne Farcot

Thesis defended on November $15^{th}$ 2011 before the jury composed of:

**Jury :**

| | | | | |
|---|---|---|---|---|
| M. | Eric RIVALS, | - | CNRS | , Président |
| M. | Jean-Luc Gouzé | - | INRIA | , Rapporteur |
| Mme. | Macha Nikolski | - | CNRS | , Rapporteur |
| M. | Yves Caraglio | - | CIRAD | , Examinateur invité |
| M. | Christophe Godin | - | INRIA | , Directeur |
| M. | Jan TRAAS | - | INRA | , Co-directeur |
| M. | Etienne Farcot | - | INRIA | , Encadrant |

Je dédie cette thèse à mon père et à ma mère.

# Remerciements

Je souhaiterais en premier lieu exprimer mes plus profonds remerciements à mon directeur de thèse, Christophe Godin. Il a été pour moi un directeur de thèse exceptionnel. Tout en me laissant une grande liberté dans mes choix, ses conseils avisés et critiques de mon travail m'ont beaucoup aidé à orienter mon travail dans les bonnes directions.

Je remercie sincèrement mon co-directeur de thèse, Jan Traas. Je le remercie pour la confiance et la liberté qu'il m'a accordées.

J'adresse mes vifs remerciements à mon encadant de thèse, Etienne Farcot. J'ai eu la chance d'avoir un encadrant dont les qualités humaines et professionnelles sont exceptionnelles. Je suis vraiment fier d'être son étudiant. Etienne a toujours pris du temps pour m'expliquer et discuter. Merci pour sa grande disponibilité, sa gentillesse et ses encouragements permanents. J'exprime ma profonde reconnaissance pour tout le temps qu'il m'a consacré.

Mes remerciement vont également à Yann Guédon. Ses conseils experts, et ses remarques avisées ont largement contribué à améliorer ma thèse. Je ne peux pas oublier de saluer ses qualités humaines, sa gentillesse, et sa patience.

Je voudrais remercier Eric Rivals d'avoir accepté d'être le président du jury de la thèse.

Je remercie également Jean-Luc Gouzé et Macha Nikolski pour avoir accepté d'être rapporteurs de ma thèse et d'en faire un compte-rendu. La discussion de vive voix et les remarques constructives de Jean-Luc m'ont aidé à améliorer le manuscrit.

Mes remerciements s'adressent également à Yves Caraglio, pour avoir examiné la thèse et pour l'intérêt qu'il a porté à la thèse.

Cette thèse étant pluridiciplinaire à l'interface de la biologie, des mathématiques et de l'informatique, j'ai eu la chance de dialoguer avec les biologistes. Je voudrais tout particulièrement remercier Teva Vernoux et Fabrice Besnard qui m'ont fournies les données biologiques utilisées dans la thèse et qui se sont montrés disponibles pour toutes mes questions. Cela m'a été une collaboration très enrichissante et passionnante. J'adresse mes remerciement à Vincent Mirabet qui m'a fourni les séquences générées par son modèle de phyllotaxie. Je le remercie pour sa disponibilité. Ces remerciements sont à partager avec Pradeep Das pour m'avoir fourni des données biologiques utilisées dans mes travaux.

Je tiens à remercier François Parcy pour avoir bien voulu participé à mon comité de thèse, pour sa gentillesse, et ses conseils.

Pendant ma thèse, j'ai eu la chance d'enseigner à l'université de Montpellier 2. Je tiens à remercier Philippe Janssen, Anne-muriel Arigon, et Therese Libourel pour la confiance qu'ils m'ont accordée et leur disponibilité.

Je souhaiterais remercier les membres de l'équipe Virtual Plants: Christophe Pradal et Frederic Boudon (pour les nombreuses discussions et leur aide), Jérôme, Eric, Daniel, Thomas (pour leur aide et la gentillesse), et aussi tous les autres membre de l'équipe: Pascal, Chakkrit, David, Szymon, Michael, Mik, Romain, Damien, Jean, Léo, Pierre, Mathilde, Anne-Laure, ... .

Je voudrais remercier mes parents pour le soutien inconditionnel qu'il m'ont accordé ainsi que leurs encouragements permanents et leur amour. Je remercie également mes sœurs Elnaz et Sorour, mes beaux frères, Kambiz et Tooraj pour leur soutien sans oublier mon petit neveu Parham pour la joie qu'il me procure. Un grand MERCI à ma merveilleuse sœur Sorour qui m'a toujours soutenu dans les moments les plus difficiles.

# Contents

# Introduction

Plants play a major role in almost all aspects of human and animal life. They provide food for other species, material for construction, chemicals for pharmaceutical industry, oxygen for the atmosphere, not to mention the aesthetic and cultural uses of flowers. Because they are constantly growing throughout their life, their shape is of crucial importance for their description and can be extremely complex. Despite this complexity, their form and structure is highly regular, at the level of axis and organs. This regularity has attracted the attention of scientists for centuries and is referred to as phyllotaxis.

Phyllotaxis has been studied from a geometrical and descriptive point of view during the last centuries. Mechanistically, phyllotaxis is the result of dynamical processes which take place in a tissue composed of undifferentiated cells called meristem. The shoot apical meristem (SAM) located at the extremity of the stem generates all above-ground elements of plants such as leaves, and flowers. The SAM is divided into functional domains, the central zone and the peripheral zone, in particular. These domains are marked by the expression of specific genes. Alterations of these domains have a strong effect on phyllotaxis, which is thus regulated at the molecular level.

The above observation indicates that to achieve an in-depth understanding of phyllotaxis, we have to consider it at different scales. From the above description, three scales emerge. Each scale can be characterized by the smallest geometrical elements involved. We call *macroscopic* the scale that involves elements as small as organs. The mesoscopic or *meristematic* scale can be similarly defined in terms of newly emerging organs called primordia. Finally, the *cellular* scale considers individual cells and the processes involving genes and hormones.

Studying phyllotaxis at these three resolutions, leads to a complexity that exceeds that of classical reductionist approaches. This highlights the need for mathematical and computational approaches. From computer science point of view, these approaches present challenges, which will lead us to develop original models and algorithms in the course of this thesis.

Though phyllotaxis is a highly regular phenomenon, a closer look indicates that it can be subject to non negligible perturbations. During the long history of studying phyllotaxis, only few studies have treated this aspect.

In this dissertation we study phyllotaxis perturbation at different scales using computational modelling. To achieve this goal, we rely on biological data at different scales. The organization of the dissertation also corresponds to the above mentioned three scales. The first two chapters introduce the state of art of phyllotaxis and gene networks that will help understand the methods and results obtained in this thesis.

Since the advent of molecular biology, a classical way to decipher the role of genes in plant development is to consider the effect of mutations. Using this approach some biologist colleagues have been led to gather data on a mutant of the model plant *Arabidopsis thaliana*. In particular, they observed perturbations in the measured divergence angles (i.e. the angle between two successive organs) in both wild-type and the mutant. In Chapter 3, we study these perturbations at macroscopic scale, using data provided by these colleagues. These perturbations turn out to be highly structured and frequent. We propose that the perturbations can be explained by permutations of successive organs on the stem. To study this hypothesis we develop a combinatorial model and algorithms.

The relatively high frequency of observed perturbation patterns in Chapter 3, suggests that any model of phyllotaxis should be able to reproduce them. In Chapter 4, using the algorithms and tools proposed in Chapter 3, we assess the ability of a perturbation model to regenerate permutations of organs. We find that to study the simulation data, we need to augment our corpus of methods to account for perturbations that can not be explained by permutations. This analysis of the perturbation model highlights the crucial role of the central zone size in the regulation of phyllotaxis.

The main genes regulating the central zone of the meristem and their interactions are known, and different models have been proposed in the literature. In Chapter 5, we further elaborate on these models, using digitized meristems in order to take into account the spatial information in modelling gene networks. The first simulations on a three dimensional tissue discarding the cellular growth and division lead us to this hypothesis that the expression zone of a gene can be determined by the geometry of meristem. To further explore this hypothesis we conduct simulations on a temporal series of three dimensional digitized meristems.

To sum up, in this dissertation, we study perturbation of phyllotaxis at different scale. We show how the processes occurring at different scales are related to each other. By investigating the origins of phyllotaxis at lower scales, we show how interactions at a scale can result in complex patterns at a higher scale.

# Chapter 1

# Modelling phyllotaxis at different scales: a state of art

## Introduction

One of the most conspicuous patterns in plants is the organisation of leaves around the stem. The arrangement of leaves and by extension of other botanical organs such as flowers, is called phyllotaxis. The stunning regularity of phyllotactic patterns has drawn the attention of scientists since centuries [Adler *et al.*, 1997]. The highly regular patterns are often compared to crystals, and Bravais brothers referred to them as *living crystals* [Bravais *et al.*, 1837].

The organs of plants are all generated by a specific tissue, the *meristem*. Apical meristems, in particular, are small populations of undifferentiated cells, located at the extremities of axes. The cells in these tissues constantly grow and divide throughout the life of the plant, starting to differentiate as they are shifted away by this uninterrupted growth. They eventually become the plant organs, leading to the phyllotactic patterns we just mentioned.

Phyllotaxis is a dynamic process. To better understand its dynamic properties, the activity of meristems has to be followed in time, and in particular the activity of genes that are expressed and interact in meristematic tissues have to be followed in time. To shed light on genes regulating phyllotaxis, models of phyllotaxis should include descriptions of the dynamics of genetic processes in meristems.

In this chapter, we present some of the main models that have been proposed to describe and analyse phyllotaxis. This presentation will only be partial, because the literature on the subject is huge, with the earliest works going back to Ancient Period, and the very recent times having been a period of intense scientific activity on the subject [Adler *et al.*, 1997]. We thus

have selected the studies that seemed most relevant for the purpose of this thesis, which aims at improving our knowledge of plant morphogenesis at different scales, relying on the tools provided by computer science.

The structure of this chapter is based on the three scales that correspond to the three first paragraphs above. First, we consider phyllotaxis at the macroscopic scale. At this scale the literature mostly provides us with descriptive models, which categorise the geometry of organ positions on the stem. Then we describe the main classes of models that have been proposed to generate phyllotactic patterns, at the scale of the whole meristem, where the smallest geometric elements are newly emerging organs, or *primordia*. Finally, we discuss models that take into account the dynamics of molecular processes that control phyllotaxis.

## 1.1 Phyllotaxis at macroscopic scale

Phyllotaxis was described and studied from direct observation of botanical organ arrangements. The geometrical analysis of the observed phyllotactic patterns allowed to characterize phyllotactic modes. In the following, we discuss phyllotactic modes.

### 1.1.1 Fibonacci numbers and the golden ratio

The mathematician Leonardo Fibonacci (1175-1240) of Pisa considered the growth of an idealized population of rabbits. The problem was to count the number of pair of rabbits within a year given one pair of rabbits (male and female) by assuming that a born pair of rabbits produce another pair each month, and rabbits never die. To solve this problem he introduced the sequence

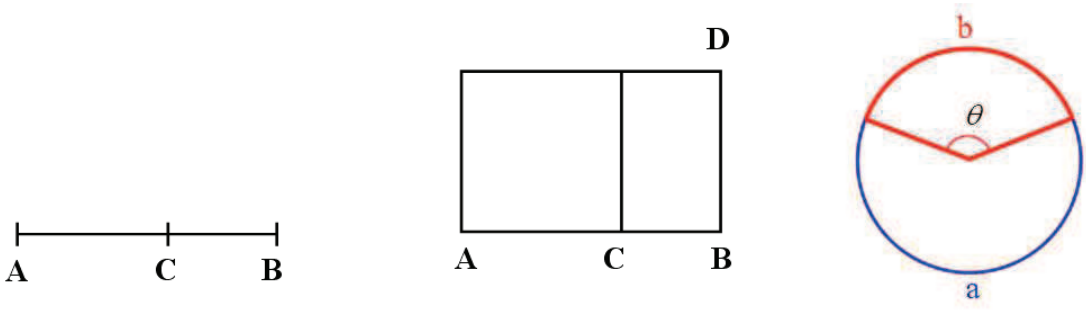
$$(1, 1, 2, 3, 5, ..., F_{k-1}, F_k, F_{k+1}, ...)$$

where $F_{k+1} = F_{k-1} + F_k$, $F_0 = 1$, and $F_1 = 1$ [Adler *et al.*, 1997].

Johannes Kepler (1571-1630) a German mathematician and astronomer was the first who observed the existence of Fibonacci numbers in plants. He was specially interested in number five [Adler *et al.*, 1997, Shipman *et al*, 2005]. He observed that the ratio of two consecutive elements of Fibonacci sequence converges and he concluded that:

$$\lim_{k\to\infty} \frac{F_{k+1}}{F_k} = \phi,$$

where $\phi$ turns out to be the golden ratio, $\phi = (\sqrt{5}+1)/2 \approx 1.61803399$. The golden ratio corresponds to a division of a line into two segments such that the ratio of the length of the line to the length of the bigger segment is equal to the ratio of the length of the bigger segment to the length of the smaller segment [Cooke, 2006, Livio, 2002, Mitchison *et al.*, 1977], see Figure 1.1.




(a) $\phi = \frac{AB}{AC} = \frac{AC}{CB}$ (b) $\phi = \frac{AB}{BD} = \frac{BD}{BC}$ (c) $\phi = \frac{a+b}{a} = \frac{a}{b}$

Figure 1.1: (a) Golden section is a line divided into two segments, the ratio of the length of the line to the length of the bigger segment is equal to the ratio of the length of the bigger segment to the length of the smaller segment, the ratio is called golden ratio. (b) Golden rectangle drawn from the golden section $ACB$, the ratio of the length of the longer side to the length of the shorter side is equal to golden ratio. If we draw the largest possible square within, the remaining rectangle is also a golden rectangle. (c) If we divide the circumference of a circle into two arcs, such that the ratio of the length of the largest arc to the length of the smaller arc is equal to the ratio of circumference to the length of the largest arc, the smaller of the two angles created is called golden angles, $\theta = 360° - \frac{360°}{\phi} \approx 137.5°$.

### 1.1.2 Phyllotactic patterns

In plants different types of phyllotactic patterns can be identified. One of the characteristics of the phyllotactic patterns is the number of organs initiating at the same node which is called *jugacy* or *whorl size* [Jean *et al.*, 1994, Smith *et al.*, 2006]. Jugacy is denoted by $j$ in the following. If only one organ appears at a time ($j = 1$), the pattern is called unijugate. Another quantitative characteristic of patterns, is the *divergence angle*. In case of unijugate phyllotaxis the divergence angle is the angle between two successive organs and is approximately constant. If the divergence angle is equal to 180° the pattern is called *distichous*, if the divergence angle is different from 180°, the pattern is called *spiral*. Spiral patterns are the most frequent patterns observed in plants [Jean *et al.*, 1994]. Sunflower, tobacco, and the model plant *Arabidopsis thaliana* have spiral phyllotaxis [Reinhardt *et al.*, 2001], see Figure 1.2. In spiral plants, linking organs in their order of initiation creates a curve that is called *generative spiral*. The time between initiation of two successive organs is called *plastochron*. Each organ has neighbouring organs called its *contacts*. Connecting the organs to the contacts creates secondary spirals (in clockwise and counterclockwise orientation) called *parastichies*. The number of parastichies are often consecutive elements of the Fibonacci sequence and such arrangements are called Fibonacci


Figure 1.2: Spiral phyllotaxis in *Arabidopsis thaliana*, the divergence angle is approximately 137.5°, the golden angle.

patterns, see Figure 1.3. The spiral phyllotaxis is characterized by the number of parastichies. In the study of phyllotactic patterns, usually two nearest contacts (leading to two parastichies numbers) are considered, since two parastichies are enough to characterize spiral patterns.

(a) Illustration of parastichies of Aonium.

(b) Illustration of parastichies of pine cone.

Figure 1.3: (a) Aonium plant. 2 spirals in counterclockwise orientation and 3 spirals in clockwise orientation. (b) Side view of pine cone. 8 counterclockwise spirals and 13 clockwise spirals. (2, 3), (8, 13) are consecutive elements of the Fibonacci sequence. Pictures from [www.math.smith.edu/phyllo].

Richards [Richards, 1951] studied the radial spacing of successive primordia along the generative spiral. He introduced plastochron ratio which is the ratio of the distance of one primordium from the center of meristem to that of the preceding primordium in the generative spiral, see Figure 1.4. Therefore, the plastochron ratio is a measure of radial growth of meristem per plastochron.

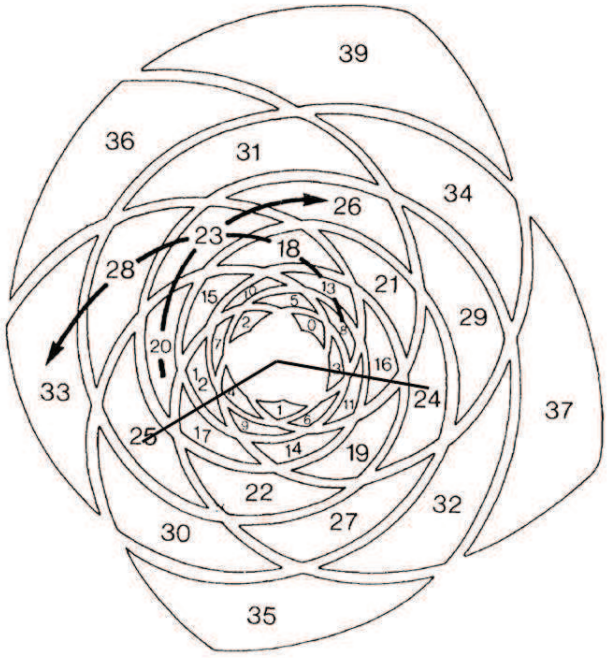



Figure 1.4: Representation of the spiral pattern and platochron ratio. The numbers correspond to order of organ initiation, 0 corresponds to the youngest organ. The arrows indicate clockwise and countercloskwise spirals. The plastochron ratio is the ratio of distances of two consecutive organs from the center, for example the distance of the organ number 25 from the center divided by the distance of the organ number 24 from the center. Picture from [Mitchison *et al.*, 1977].

Let $P_i$ and $P_{i+1}$ be two consecutive primordia in generative spiral and $d_i$ and $d_{i+1}$ be the distances of the primordia $P_i$ and $P_{i+1}$ from meristem center, the plastochron ratio ($PR$) is:

$$PR = \frac{d_{i+1}}{d_i}.$$

Phyllotactic patterns can be characterized by $PR$. When $PR$ is close to 1, the plant generates high parastichies numbers, for example sunflower can generate up to (55, 89) (clockwise, counterclockwise or vice versa) spirals [Jean *et al.*, 1994, Reinhardt *et al.*, 2001]. Whereas the potato has $PR = 1.3$ and generates (2, 3) (clockwise, counterclockwise or vice versa) spirals [Reinhardt *et al.*, 2001]. Patterns with high numbers of parastichies are called high order patterns.

There are sequences in which terms are the sum of two preceding ones as the Fibonacci sequences, but the first terms of the sequence are different from the first terms of the Fibonacci sequence. A sequence beginning by 1 and $p$ in which each term is the sum of two previous terms is called $(p-2)$th *accessory* sequence [Jean *et al.*, 1994, Smith *et al.*, 2006]:

$$(1, p, p+1, ..., a_{k-1}, a_k, a_{k+1}, ...), \quad a_{k+1} = a_{k-1} + a_k.$$

The first accessory sequence $(1, 3, 4, 7, 11, ...)$ is also called Lucas sequence.

A sequence constructed in the same way, beginning by 2 and $2q + 1$, $q > 1$, is called *anomalous* [Jean *et al.*, 1994, Smith *et al.*, 2006]:

$$(2, 2q + 1, 2q + 3, ..., a_{k-1}, a_k, a_{k+1}, ...), \quad a_{k+1} = a_{k-1} + a_k.$$

In addition to distichous and spiral phyllotactic patterns, there also exists whorled and multijugate patterns. In whorled and multijugate patterns several organs initiate at the same node. This means that the jugacy, $j$, (defined as the number of organs appearing at the same node) $j \geq 2$. In whorled phyllotactic arrangements $j$ organs appear at the same whorl at angles $360°/j$ from each other. The organs of the next whorl initiate in the spaces at the centre of the spaces between the organs of the previous whorl. The divergence angle between an organ of a whorl and the nearest organ of the next whorl is $\theta = 180°/j$. The divergence angle $\theta$ can be smaller than $180°/j$. In this case, a spiral arrangement can be observed and is called multijugate. Multijugate patterns are characterized by the number of parastichies which are multiples of the number of parastichies of unijugate patterns, see Figure 1.5. In multijugate patterns, the jugacy $(j)$ turns out to be equal to the greatest common divisor of the parastichy numbers and is in fact the number of generative spirals. See Figure 1.7 for schemas of the phyllotactic patterns.

Flowers have often whorled phyllotaxis, with sepals and petals arranged in whorls. Within the whorled patterns, the number of organs in each whorl further characterizes the patterns, as patterns with 2 or 3 organs per whorl are called decussate and tricussate respectively, see Figure 1.6.

To sum up, the phyllotactic patterns can be described by the divergence angle $(\theta)$ and the number of organs initiating simultaneously $(j)$ as follows:

$$j{=}1 : \begin{cases} \text{distichous, if } \theta = 180°, \\ \text{spiral, if } \theta \neq 180°, \end{cases}$$

$$j \geq 2 : \begin{cases} \text{whorled, if } \theta = 180°/j, \\ \text{multijugate, if } \theta \neq 180°/j. \end{cases}$$

The transition from decussate phyllotaxis to spiral one can be observed in many plants. For example, leaves of Arabidopsis at first are arranged with a divergence angle close to 90° generating a decussate pattern. Arabidopsis gradually generates spiral phyllotaxis as it develops from vegetative growth to flowering. The siliques (i.e. Arabidopsis fruits) form a spiral phyllotaxis with a divergence angle close to 137.5°.

Figure 1.5: Bijugate Fibonacci pattern ($j = 2$). This gymnocalycium has parastichy numbers $(10, 16) = 2(5, 8)$ which are two consecutive elements of $2 \times (1, 2, 3, 5, 8, 11, ...)$. Picture from [www.math.smith.edu/phyllo].

Figure 1.6: Whorled pattern (tricusatte). Picture from [www.math.smith.edu/phyllo].

## 1.2 Phyllotaxis at the meristem scale: dynamical models

The high regularity of phyllotactic patterns has drawn the attention of scientists from different domains. Biologists, physicists, mathematicians, as well as computer scientists have proposed models of this phenomenon. Different mechanisms have been suggested to explain how the organs are formed and how their positions are determined at the flank of the meristem. Despite the variety of mechanisms, they all involve the influence of pre-existing organs on the position of future organ initiations.

In 1868 Hofmeister [Hofmeister, 1868] proposed a rule describing the initiation of a new organ. He suggested that the new organ initiates periodically (i.e. the plastochron is regular) at the largest available space left by older organs. Snow and Snow [Snow *et al.*, 1931] further investigated this idea by surgically isolating the site of an incipient primordium. An incipient primordium (or more generally an incipient organ) is a primordium in its first stages of development). Though the next primordium was formed at the expected position with a divergence angle close to 137.5°, the second next primordium appeared closer to the isolated primordium than usual (with a divergence angle about 165°), see Figure 1.8. Snow and Snow used the displacement of the next second primordium as an evidence to propose a *first available* mechanism in which an organ appears when there is a minimum available space on the flank of meristem

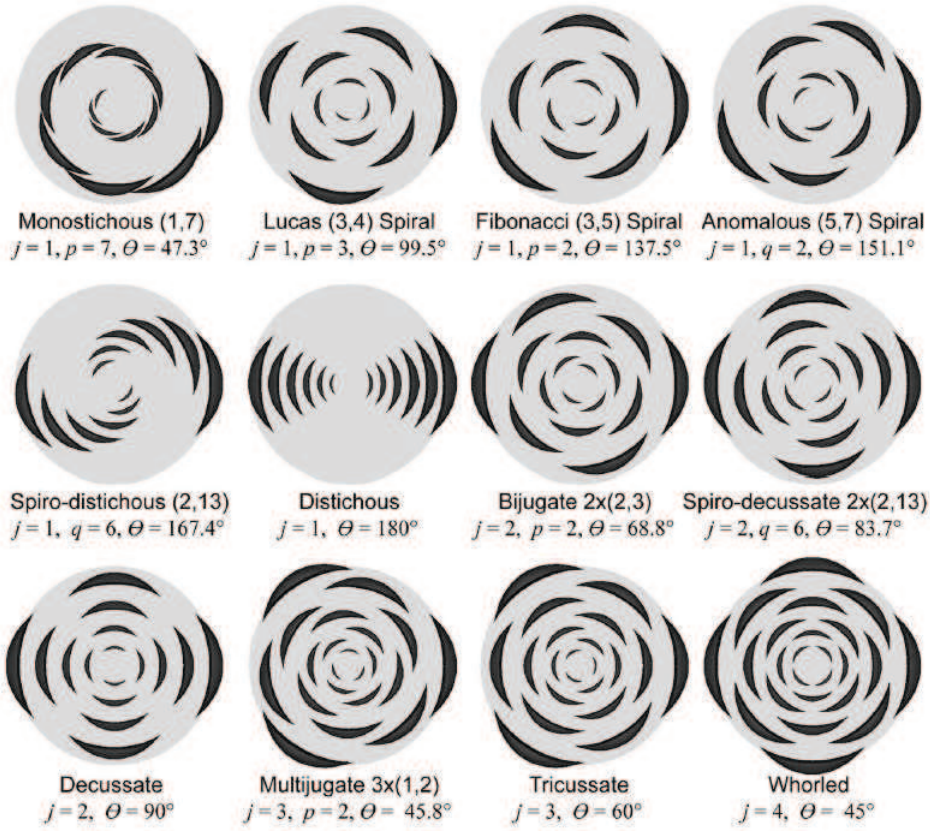



Figure 1.7: Schemas of phyllotactic patterns. For each pattern, jugacy $j$ and divergence angle $\theta$ are given. For description see the text. Picture from [Smith *et al.*, 2006].

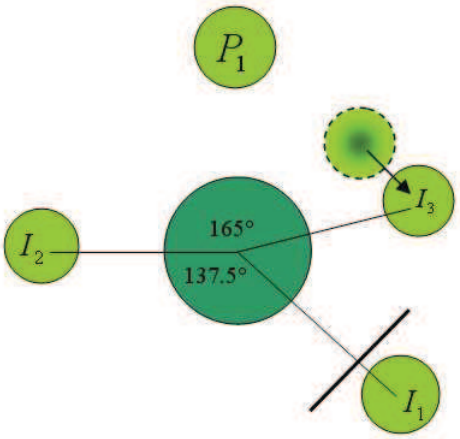



Figure 1.8: Schemas of the experiment of Snow and Snow. The incipient primordium $I_1$ is isolated by an incision. The next primordium $I_2$ initiates with a divergence angle close to 137.5°, the second next primordium $L_3$ initiates with a divergence angle of 165° closer than usual to $I_1$.

with a minimum distance from older organs. This mechanism is quite similar to that of Hofmeister except the periodic initiation imposed by the Hofmeister mechanism, which is absent in the proposition of Snow and Snow.

To explain the phyllotactic patterns, other mechanisms were also proposed. Another biophysical mechanism is based one the differential growth between the outer layers of the SAM and the internal layer of the SAM. In this mechanism the outer layers grow faster than the internal layer which leads buckling at the surface, resulting in organ formation [Green, 1992, Green, 1996].

In 1913, Schoute proposed a theory to determine the place of initiation of organs. According to his theory, a primordium produces a chemical inhibitor that prevents the formation of new organs nearby. This idea is known as "inhibitory field theory". According to this theory, primordia and the meristem summit are surrounded by inhibitory fields that can be gradients of nutrients or signalling molecules. The fields generated by older primordia prevents initiation of a new organ.

Douady and Couder [Douady *et al.*, 1996(a), Douady *et al.*, 1996(b), Douady *et al.*, 1996(c)] proposed two dynamical models with inhibitory fields around primordia. They used these models to provide an in-depth study of phyllotactic patterns. Notably, they were able to reproduce most of the known spiral and whorled patterns described in the previous section, showing that the hypothesis of inhibitory fields was a satisfactory explanation for the emergence of phyllotaxis from this lower scale dynamical system. In Chapter 4 we study a modified version of one of their models to investigate the role of noise in phyllotaxis. Therefore, in the following, we explain the models in detail.

### 1.2.1 Models proposed by Douady and Couder

In a series of three papers [Douady *et al.*, 1996(a), Douady *et al.*, 1996(b), Douady *et al.*, 1996(c)] Douady and Couder revisited the ideas of Hofmeister [Hofmeister, 1868] and Snow and Snow [Snow *et al.*, 1952] for the formation of a new organ. They proposed two dynamical models to reproduce phyllotactic patterns. The first one is based on the conditions of initiation of a new organ put forward by Hofmeister and the second one is based on the conditions proposed by Snow and Snow.

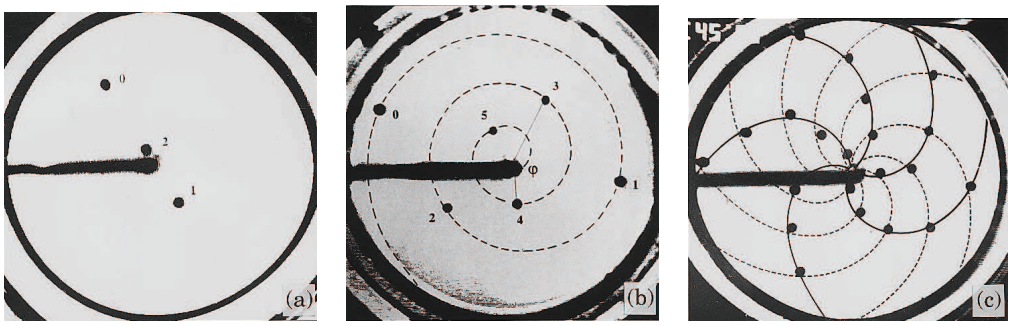
#### First model: revisiting the hypotheses proposed by Hofmeister

Douady and Couder re-examined the hypotheses of Hofmeister [Hofmeister, 1868] and reproduced spiral and distichous patterns in a physics experiment as well as in numerical simulations of a dynamical model.

According to the hypotheses proposed by Hofmeister, an organ appears periodically, with a fixed time period $T$ (i.e. the plastochrone is regular), at the peripheral zone of meristem in the largest space left by older ones. Because of the growth, the organs move away radially from the centre of the SAM with a velocity $V(r)$, $r$ being the distance of the primordium form the SAM center. Once an organ is formed, its angular position does not change over time outside of a region of radius $R_0$. The SAM is supposed to be axisymmetric.

#### The experiment

Douady and Couder performed a physics experiment in which the SAM was modelled as a dish filled with silicon oil placed on a vertical magnetic field. The initiation of the organs was modelled by droplets of ferromagnetic liquid falling at the center of the dish at regular time intervals $T$. The droplets were polarized in the presence of the magnetic field which had a radial gradient: minimum at the center and maximum at the periphery. The polarized droplets repelled one another. They were attracted to the periphery of the dish where the magnetic field was maximum, thus they moved away from the centre. By tuning $T$ the experiment created distichous phyllotactic mode (for large values of $T$) and spiral phyllotactic modes (for small values of $T$), see Figure 1.9. Douady and Couder report that droplets interacted with each other inside a region of radius of $R_c$. Outside this region the droplets were too far from each other to interact and they only drifted away radially from the center. The authors showed that the patterns produced were controlled by an adimensional parameter $G = V_0 T/R_c$ where $V_0$ was the advection velocity which is controlled by magnetic field. They also showed that $G = \ln(PR)$, where $PR$ is the plastochrone ratio.




(a) Distichous mode (b) Spiral mode (c) Spiral mode of higher order

Figure 1.9: Different phyllotactic patterns obtained by the experiment. (a) Distichous mode, divergence angle is 180° (b) Spiral mode with parastichy numbers (1, 2). The generative spiral is represented by the dashed line. (c) Spiral pattern of higher order with parastichy numbers (5, 8). Picture from [Douady *et al.*, 1996(a)].

**Model**

Based on the above mentioned hypotheses, Douady and Couder proposed a computational model. In this model, the system was further simplified by supposing that the SAM was planar and organs were circles of radius $d_0$. To make the organs appear in the largest space left by older ones, it is assumed that older organs exerted a repulsive force on incipient organs. This can be viewed as a simple way to model space occupation of organs.

More precisely, in this model organs appeared on a circle $C$ with radius $R_0$ with a time periodicity of $T$ on a planar meristem. They moved away from the center with velocity $V(r)$. Each organ created a repulsive energy, $E(d)$, where $d$ is the distance between the primordium and a sampling point $S$ on the circle $C$, see Figure 1.10. The repulsive energy represents the inhibitory fields previously mentioned. Let $\{P_1, ..., P_l\}$ be the set of existing primordia, the combined repulsive energy $h$ on the point $S$ is given by:

$$h(S) = \sum_{i=1}^{l} E(d(P_i, S)), \tag{1.1}$$

where $d(P_i, S)$ is the distance between the point $S$ and the primordia $P_i$, $1 \leq i \leq l$. The functions $E(d)$ that were used in the simulations were $1/d$, $1/d^3$ or $\exp(-d/l)$. Douady and Couder report that these functions generated qualitatively the same patterns.

To determine the location of a new primordium on circle $C$, the function $h$ is calculated for a discrete set of points on circle $C$. An organ appears at each period $T$ at the point for which the function $h$ is minimum. As mentioned before, the simulations of this model generated distichous and spiral phyllotaxis. Since only one organ appears after each time interval $T$, the model could not generate whorled phyllotactic patterns.

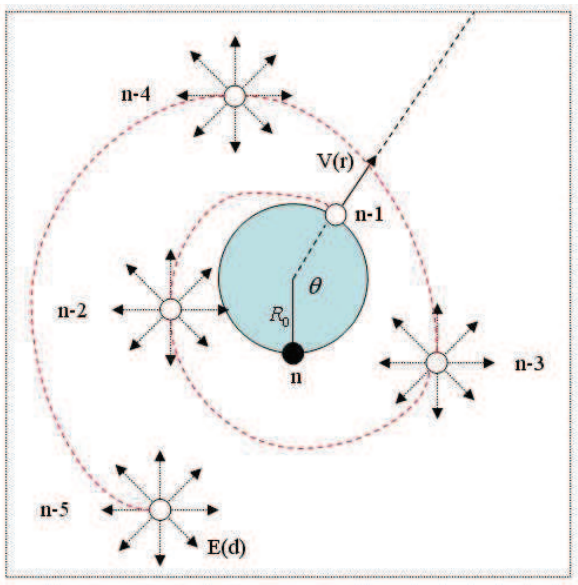



Figure 1.10: Schema of the planar model of the SAM. The incipient organ of number $n$ is represented in black. The position of the organ number $n$ is a function of positions of previous ones, $n-1$, $n-2$, ... . It is defined by the divergence angle $\theta$, and the ratio of their distances to the center, *plastochrone ratio*. The organs move radially from center with a velocity of $V(r)$, $r$ being the distance from the center. The inhibitory field lines of potential $E(d)$ are shown at each organ. The red dotted line represents the generative spiral.

**Second model: revisiting the hypotheses proposed by Snow and Snow**

Snow and Snow [Snow *et al.*, 1952] proposed a modification to the set of hypotheses put forward by Hofmeister. They simply removed the time periodicity of initiation of organs and suggested that the initiation of an organ takes place at the peripheral zone of the SAM where and when there is enough space for its formation. Similar to the hypotheses of Hofmeister, organs moved away with a radial velocity of $V(r)$. The angular positions of primordia did not change over time outside of a region of radius $R_0$.

Based on this new set of hypotheses Douady and Couder proposed a second model in which they assumed that the shape of the SAM is paraboloidal and primordia have azimuthally elongated shape. Douady and Couder took these assumptions into account by a introducing a supplementary parameter called the conicity. Similar to the first model, each primordium created a repulsive energy $E(d)$ and the combined repulsive energy $h$ on the point $S$ generated by the existing primordia $\{P_1, ..., P_l\}$ is given by the Equation 1.1.

To find where a new primordium appears, they calculated continuously the value of the function $h$ on the sampling points. A new primordium appears at the point for which the

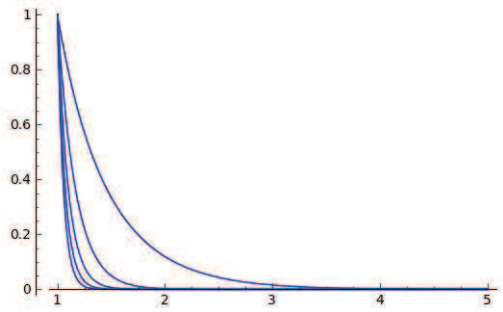



Figure 1.11: Profiles of the repulsive energy function $E(d)$, Equation (1.2), for $d_0 = 1$ and $\alpha \in \{1, 3, 5, 7, 9\}$. As $\alpha$ increases the curve becomes steeper ($\alpha = 1$ is the rightmost curve and $\alpha = 9$ the leftmost).

function $h$ is minimum whenever the value of function $h$ is below a certain threshold $E_S$.

The function $E(d)$ used in the second model is

$$E(d) = \frac{-1 + \left(\tanh\ \alpha \frac{d}{d_0}\right)^{-1}}{-1 + (\tanh\ \alpha)^{-1}}, \tag{1.2}$$

where tanh is the hyperbolic tangent and $\alpha$ represents the stiffness of the interaction, see Figure 1.11. The function is a decreasing function of distance and the higher is $\alpha$ the stiffer are interactions.

The authors showed that the state of the system can be described by the following parameter:

$$\Gamma = d_0/R_0. \tag{1.3}$$

If the conicity is taken into account

$$\Gamma = l_1/R_0,$$

where $l_1 = d_0/N^{1/2}$ is the orthoradial size of primordia, $1/N$ representing the conicity. See Figure 1.12 in which the bifurcation diagram of divergence angles as a function of $\Gamma$ is presented. Each point in the diagram corresponds to a steady value of divergence angles obtained in simulations. As one can observe in the figure, the model reproduces different steady phyllotactic patterns characterized by $\Gamma$.

Biological relevance of the inhibitory fields is not mentioned in the models proposed by Douady and Couder. They only considered an abstract form of such fields. Therefore, much research since then has considered alternative forms of inhibitory fields, as we summarize now, as well as more complex models including sub-cellular processes, as described in the next section.

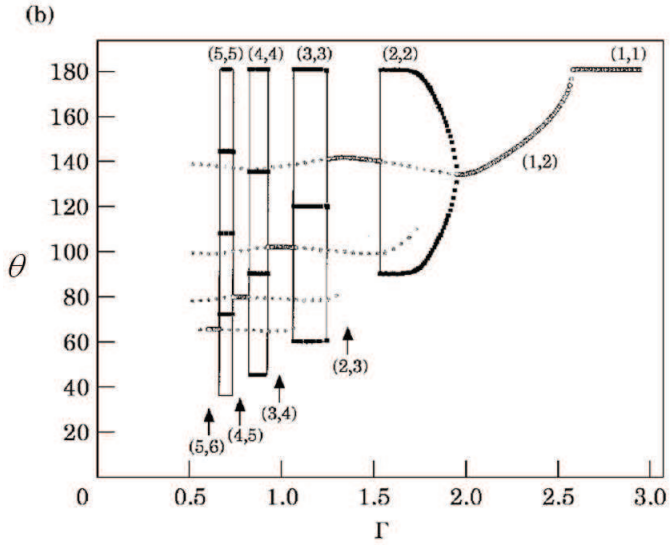



Figure 1.12: Diagram of divergence angle $\theta$ as a function of $\Gamma$. Spiral patterns are labelled by the number of clockwise and counterclockwise parastichies $(m, n)$. Whorled patterns are labelled by the number organs at each whorl $(j, j)$. $1/N = 3$, and $\alpha = 8$. Figure from [Douady *et al.*, 1996(b)].

Smith *et al.* [Smith *et al.*, 2006] proposed a model in which the spatial distribution and age of older organs determine the initiation place of a new organ. Similar to the second model proposed by Douady and Couder, the pre-existing organs created repulsive energy modelled by an inhibition function. An organ appeared where and when the sum of inhibitory influence of pre-existing organs was below a predefined threshold $E_s$. To reproduce whorled and multijugate patterns a simulation model with two inhibition functions was proposed. Both functions represented the inhibitory contribution of older primordia by taking into account their spatial distribution and their age. In addition, in one of the functions, the inhibitory effect of newly initiated organs increases gradually over time. This allowed to create phyllotactic patterns in which several organs initiate simultaneously in a robust manner.

More precisely, let $\{P_1, ..., P_l\}$ be the set of existing organs and $t_i$ be the age of organ (primordium) $P_i$, $1 \leq i \leq l$, two inhibition functions are defined as:

$$h_1(S) = (1.0 - e^{-b_d t_i}) \sum_{i=1}^{l} \frac{1}{d(P_i, S)} e^{-b_1 t_i}, \tag{1.4}$$

$$h_2(S) = \sum_{i=1}^{l} \frac{1}{d(P_i, S)} e^{-b_2 t_i}, \tag{1.5}$$

where $d(P_i, S)$ is the distance between organ $P_i$ and a sampling point $S$ on the apex. $b_1$ and $b_2$ control the exponential decrease of inhibition and $b_d$ controls the inhibitory influence of the

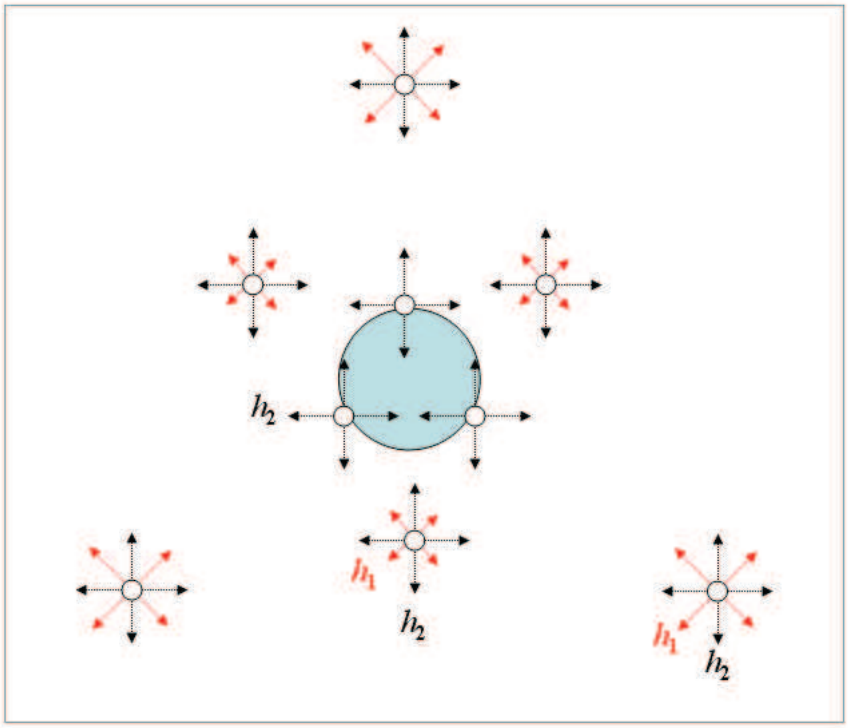



Figure 1.13: Schema of the two inhibition functions creating a tricussate pattern. Inhibiting influence of each organ is shown by field lines at each organ. Red lines correspond to $h_2$ function and black field lines correspond to $h_1$ function. Inhibition functions $h_1$ and $h_2$ are defined in Equations (1.4) and (1.5) respectively.

organ changing gradually over time.

The function $h_1$ represents a phased inhibiting affect of a newly formed organ and allows the simultaneous appearance of several organs as in whorled and multijugate phyllotactic patterns. The problem with this delay mechanism is that the threshold $E_s$ can be reached for neighbouring sampling points with short plastochron. In consequence, several organs can appear close to each other. To prevent this problem the inhibition function $h_2$ is added to the system. A primoridum appears when and where both functions $h_1$ and $h_2$ are both below the predefined threshold $E_s$, see Figure 1.13. This model is able to generate more robust whorled and multijugate patterns compared to the model proposed by Douady and Couder.

Chapman and Ferry [Chapman *et al.*, 1987] proposed a complementary mechanism of inhibitory fields. This mechanism included an organ initiating factor. The depletion of the factor by pre-existing organs prevented appearance of a new organ in the vicinity the organs. This mechanism was considered to be responsible for creating inhibitory fields. This idea in recently supported by biological facts. It is showed that the high concentration of auxin, a plant hormone,

is an essential factor for organ formation [Reinhardt *et al.*, 2000]. [Reinhardt *et al.*, 2003(a)] proposed that new organs deplete auxin, this generates a concentration gradient of auxin that is equivalent to inhibitory fields. The role of auxin in the regulation of phyllotaxis is discussed in the next section.

## 1.3 Molecular regulators of phyllotaxis

### 1.3.1 Regulation of the central zone of meristem

The aerial tissues of plants are generated from the shoot apical meristem (SAM). The SAM is divided into functional domains that are characterized by different patterns of gene expression. The central zone (CZ) situated at the tip of the meristem consisting of stem cells is the domain of expression of CLAVATA3 (CLV3) gene. Stem cells in central zone divide and the daughter cells enter the periphery of meristem, called peripheral zone (PZ), where leaf and flower primordia are formed. In a predictable zone just below the CZ, the organizing center (OC), the master regulator gene WUSCHEL (WUS) is expressed, see Figure 1.14. WUS upregulates the expression of CLV3 while in return in conjunction with the receptor kinase CLAVATA1 (CLV1) represses the expression of WUS. This negative feedback network is believed to regulate the maintenance of the central zone and its genetic identity. This controls the maintenance of the stem cell identity in meristems [Sablowski, 2007, Sharma *et al.*, 2003], see Figures 1.15 and 1.16. In Chapter 5, we study the role of negative feedback network between WUS, CLV1, and CLV3 in regulating the size of central zone. In particular, we study the zone of expression of WUS on a three-dimensional model of the SAM. Therefore, in the following we present a state of art about pattern formation involving WUS and models that have been proposed to describe these processes.

Primordia that arise on the flanks of the SAM can develop into flowers and in this case are called floral meristem. This process is controlled by genes [Lohmann *et al.*, 2002, Parcy *et al.*, 1998, Parcy *et al.*, 2005]. Different stages can be identified during development of a flower. Smyth *et al.* [Smyth *et al.*, 1990] divided the development process into 12 stages. Stages are defined in terms of observable morphological events in flower development. The first stage is the appearance of flower primordium as a buttress on the flank of the SAM. Then floral primordium grows out and becomes separate from the SAM (stage two). Stage three is marked by the appearance of sepals. In stage four sepals develop to overlie the floral meristem. The floral meristem develops until it becomes a completely developed flower. In this dissertation, we focus on the SAM, particularly in Chapter 5, we study the gene regulatory networks in the first 3 stages. Therefore, we do not describe stages five to twelve [Smyth *et al.*, 1990], see Figure 1.17.

In all dynamical models of phyllotaxis mentioned before, the central zone plays a crucial role. For instance, in the models of Douady and Couder, the key parameter $\Gamma$ which is the

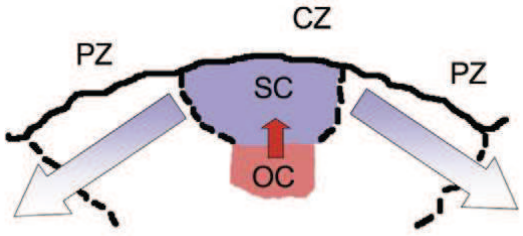



Figure 1.14: Functional domains of shoot apical meristem. Stem cells and organizing center are marked by blue and red colors respectively. CZ and PZ indicate the central zone and the peripheral zone. The cells in central zone divide and enter the peripheral zone where the differentiate and form primordia. The arrows represent the transition to differentiation. Picture from [Sablowski, 2007].

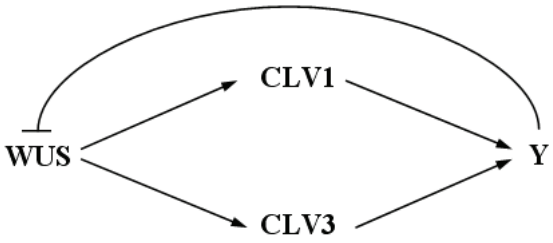



Figure 1.15: Negative feedback network between WUS, CLV1 and CLV3. WUS induces both CLV1 and CLV3. CLV1 in conjunction with CLV3 induces a signal indicated by Y that represses WUS.

ratio of the diameter of primordia to the central zone diameter, as we mentioned this parameter can explain all the phyllotactic regimes of the model as bifurcation events when it is varied, see Figure 1.12. Also, biological experiments have shown that modifications of this region, either of genetic origin or by direct alteration of the meristem, typically had a major impact on phyllotaxis. For these reasons, several models of the gene regulatory network that controls this region have been proposed in the recent years [Geier, *et al.*, 2008, Gordon *et al.*, 2009, Jönsson *et al.*, 2005, Hohm *et al.*, 2010]. In the following, we first present the biological facts related to the central zone and then we present these models.

### 1.3.2 Pattern formation in the SAM involving WUS

Recent studies on morphogenesis at the meristem scale rely on an integrated combination of experimental and theoretical approaches, a strategy sometimes used in the context of systems

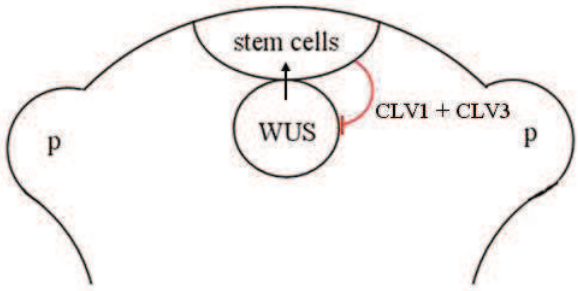



Figure 1.16: WUS expression domain in a cross section of the SAM. Above the WUS expression zone is the expression zone of CLV3 , labelled as stem cells. WUS up-regulates the expression of CLV3 which in return down-regulates WUS by binding to its receptor CLV1. P labels primordia such as sepal, leaf or flower primordia. Figure adapted from [Gordon *et al.*, 2009].

biology [Traas *et al.*, 2010]. In this section, we first present the main biological results that are known on the control of the central zone, and then the modelling studies available on this topic.

**Summary of biological facts**

A significant property of meristem is its ability in self-organization. Classical microsurgical experiments hightlighted the high plasticity of the SAM, for example needle pricking of the CZ does not stop the activity of meristem [Pilkington, 1929]. Reinhardt *et al.* [Reinhardt *et al.*, 2003(b)] conducted experiments in which parts of the tomato SAM were ablated using laser technology. After ablation of CZ and OC, the meristem reorganized itself; WUS was induced at low levels in a ring-shaped region around the lesion. At a later time after ablation, WUS expression domain became confined to one or two side(s). For each domain a new meristem center was initiated.

[Reinhardt *et al.*, 2003(b)] also used microsurgical techniques to reveal the role of L1 layer in the SAM development. In their experiments, after partial ablation of L1, the growth center of the meristem shifted to the intact side. This indicates the importance of L1 layer in specifying growth center of meristem.

Analysis of mutants gives indications to understand the function of genes. Loss-of-function CLV3 (or CLV1) and gain-of-function WUS mutants have a greatly enlarged SAM as well as an enormous CZ. Experiments performed by [Reddy *et al.*, 2005] showed that the expansion of CZ after reduction in CLV3 activity results from re-specification of PZ cells to CZ cells. This re-specification is followed by enlargement of the SAM either through increased cell division or reduced differentiation, or both. They suggest that CLV3 prevents differentiation of PZ cells into CZ cells leading to restriction of CLV3 domain. CLV3 prevents the enlargement of the SAM through inhibition of cell division rate.

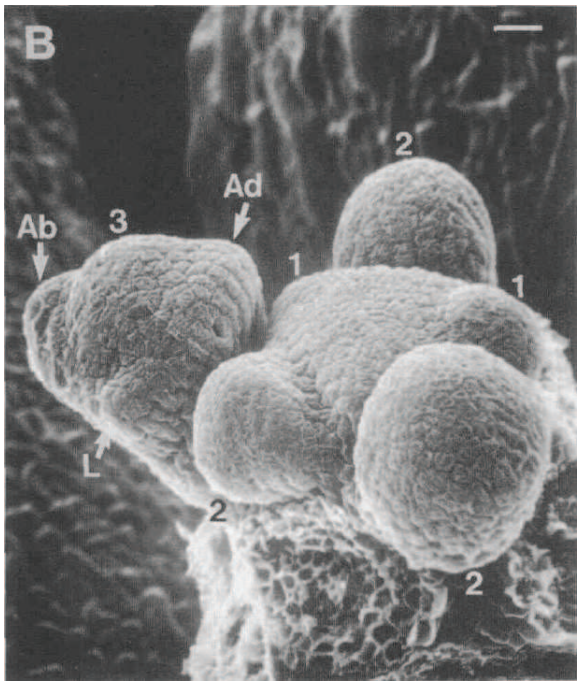



Figure 1.17: Stages of floral meristem. The stage of each bud is indicated by a number from 1 to 3. Ab, Ad, and L indicate abaxial, adaxial and left sepals. Picture reproduced from [Smyth *et al.*, 1990].

It appears that all the studies mentioned above show that once formed, the CZ is very robustly maintained. The creation of this zone, when a new primordium is being formed, on the other hand, has not yet received any exploration. This will be discussed from a modelling point of view in Chapter 5.

### 1.3.3 Overview of modelling approaches

To understand more precisely how this feedback loop controls the spatial extent of the CZ, mathematical models have been developed.

Jönsson *et al.* [Jönsson *et al.*, 2003] propose a model described by a set of ordinary differential equations (ODE) in which the WUS signal is combined with a signal originating from the L1 layer to induce CLV3 expression. Given the expression domain of WUS the model organizes the CLV3 expression domain. In a more elaborate study [Jönsson *et al.*, 2005] they propose two models for the organization of the WUS expression domain which are capable of recreating WUS expression domain in a two dimensional cell-based meristem tissue. The dynamics of the models is described by ordinary differential equations. One of the models, based on the Brusselator reaction diffusion system (see Chapter 2) is capable of reorganizing the WUS expression domain after an ablation of CZ cells in a lattice platform [Reinhardt *et al.*, 2003(b)]. In Chapter 5, these

two models are presented in detail and we use a modified version of one of them to study the role of geometry on the WUS expression zone.

[Hohm *et al.*, 2010] propose a dynamic model using partial differential equations (PDE) for pattern formation in the SAM. By restricting the stem cell identity to outer layers and including a spatial parameter which makes cells that reside closer to the meristem tip more competent to activate WUS expression, the model mimics the WUS and CLV3 expression domains. It is also capable of reorganizing the WUS expression domain after CZ ablation [Reinhardt *et al.*, 2003(b)]. Whereas [Jönsson *et al.*, 2005] considered an abstract repressive signal emanating from the whole L1 region, this model considers explicitly the CLV3 signal. They also address the question of WUS creation, although performing their simulations on a fixed, fully formed, digital meristem.

Recently hormonal feedback on gene regulatory networks were studied through a systems biology approach. Gordon *et al.* [Gordon *et al.*, 2009] conducted experiments and developed a mathematical model providing a new insight of hormonal feedback on WUS expression. In their model, they study the cytokinin perception network and they implicate WUS expression. Their model predicts that cytokinin induces WUS through both clv-dependent and clv-independent pathways. This prediction is further confirmed by biological experiments. To account for the spatial patterning of WUS, the authors propose that the cytokinin response distribution acts as a spatial reference to inform cells of their position in space.

A different kind of model, where variables are no longer concentrations but the size of cell populations within the SAM, has been proposed by Geier *et al.* [Geier, *et al.*, 2008]. They measured the functional domains size and mitotic indexes in different conditions. Mitotic index measures the proliferation of a population of cells. It is the ratio of the number of cells in mitosis to the total number of cells. Via computational modelling they addressed the question of how stem cell domain is regulated by changes in differentiation rate or proliferation rate. They propose three models described by ordinary differential equations in which they treat implicitly spatial structure of the SAM by using cell pools that are connected via differentiation rate. They test the models to predict CLV3 and WUS over-expression phenotypes and CZ homeostasis under changing growth conditions, but none of them is capable of correctly predicting both of them at the same time.

The presented computational models include some spatial information to re-generate WUS expression zone. Imposed implicit spatial pre-patterning are often behind the models to reproduce the zone of expression of genes. In the case of [Gordon *et al.*, 2009] this takes the form of a pre-defined Gaussian shaped cytokinin distribution. In both [Hohm *et al.*, 2010] and [Jönsson *et al.*, 2005], indicator functions are used to represent a specific behaviour (production rate in both cases) of the external layers of the meristem. The latter choice is motivated by the biological fact that the L1 layer of the meristem has a specific genetic identity, see for

example [Lu *et al.*, 1996].

Besides the CLV and WUS genes, many other molecular actors play a role in the CZ specification and maintenance: known genes, like TFL1, STM, SCR, ZWILLE to name a few [Sablowski, 2007, Weigel *et al.*, 2002], and probably others yet to be discovered. However, in most cases the precise molecular interactions in which these genes are involved is not known, and neither is their precise expression domain in time and space. This indicates why any model proposed at the time being has to consider only a restricted set of molecular signals. These signals are either known from biological experiments, or proposed as hypotheses to be tested.

All existing models are simulated on two dimensional static tissues. Both the geometry of the meristem, which is three dimensional, and the change it undergoes during development, are thus ignored in these models. Contrary to the previous point, recent data and computational methods provide us with precise information on the meristem geometry in 3D, at different time points, and remain to be used for modelling purposes [Fernandez *et al.*, 2010].

To take advantage of these new tools, it seems appropriate to rely on a model where geometry plays an important role, even if the molecular details are reduced to a simplified network involving only few variables, and limited regulatory interactions. Among the models described above, those proposed in [Jönsson *et al.*, 2005] match quite well with this criterion, and we thus focus on them in Chapter 5.

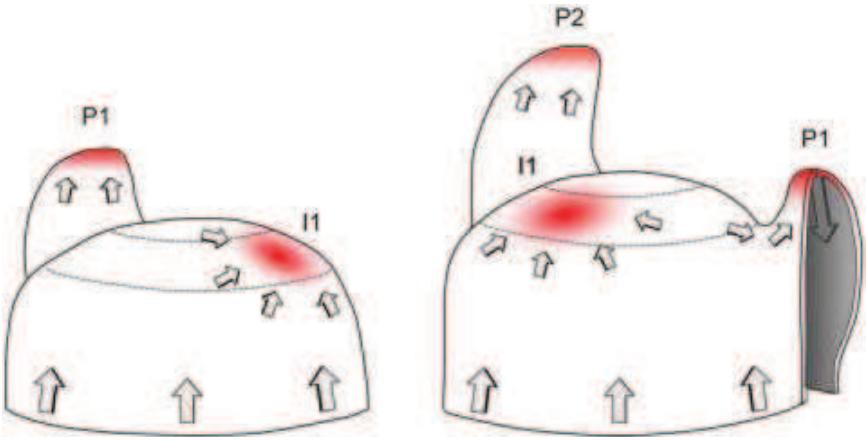
### 1.3.4 Regulation of phyllotaxis by hormones

Plant hormone auxin plays an important role in development and organ initiation. For example, auxin is implicated in embryogenesis, cellular differentiation and induces organ formation at the SAM [Reinhardt *et al.*, 2000]. Since auxin contributes to new organ formation, it can be involved in phyllotaxis by determining the position of organ initiation.

Auxin is actively transported throughout the plants by cellular import and export proteins. The two main proteins in auxin transport are the influx carrier AUXIN RESISTANT1 (AUX1) and the efflux carrier PIN1, located at the plasma membrane [Reinhardt, 2005]. These import and export proteins create fluxes of auxin that leads to the formation of local auxin maxima. Local auxin maxima can represent potential sites for new organ initiations.

Reinhardt et al. [Reinhardt *et al.*, 2003(b)] conducted experiments and proposed a model for phyllotaxis regulation. Auxin is transported upwards into the SAM. Pre-existing primordia absorbed auxin, acting like sinks. In consequence, auxin is depleted from nearby area around primordia. This leads to heterogeneous auxin distribution in the SAM. Auxin accumulation cannot take place in the vicinity of pre-existing primordia but at a minimal distance from them. This determines the position of new organ in the SAM, see Figure 1.18.

A series of phyllotaxis models inspired by the above conceptual model has been pro-




(a) (b)

Figure 1.18: Model of phyllotaxis regulation by auxin transport. (a) PIN1 proteins direct auxin fluxes in L1 layer. This creates auxin accumulation marked by red color at the future organ initiation position at the PZ. (b) Auxin is diverted by P1 primordium downwards leading to auxin depletion from nearby. This leads to another auxin accumulation at the PZ, I1. Picture from [Smith *et al.*, 2006].

posed [Barbier, 2005, Heisler *et al.*, 2006, Smith *et al.*, 2005]. These models include active and passive auxin transport and the SAM growth. They involve a mechanism to allocate PIN1 proteins to plasma membrane. Barbier [Barbier, 2005] modelled the auxin active transport by supposing that the primordia and the SAM summit attract auxin fluxes, the model is based on the work of [Barbier *et al.*, 2006]. Jönsson *et al.* [Heisler *et al.*, 2006] and Smith *et al.* [Smith *et al.*, 2005] proposed that the amount of PIN protein allocated to a plasma membrane of a cell is determined by concentration of auxin in neighbouring cells such that auxin is transported from low concentration to high concentration against its gradient. These models are able to generate different types phyllotactic patterns.

Stoma *et al.* [Stoma *et al.*, 2008] proposed a different mechanism to active auxin transport called *flux-based*. According to this mechanism the amount of PIN protein allocated to a membrane is determined by net auxin flux between neighbouring cells. This leads the PIN proteins to amplify and stabilise small existing fluxes, which eventually leads to the formation of local maxima. By computational modelling, Stoma *et al.* show that the above mentioned flux-based polarization hypothesis is able to explain auxin transport at the shoot meristem.

Another important family of plant hormones that is involved in different plant development mechanisms is cytokinins. Cytokinins are in particular involved in regulating phyllotaxis [Vernoux *et al.*, 2010]. In the cytokinin signalling pathway, a gene called AHP6 plays a repressive role [Jennifer *et al.*, 2007], and the expression pattern of the associated protein turns

out to be surrounding the newest primordia, forming a sort of annular shape of a few cells around them (data not shown). It was thus tempting to look for a relation between the gene AHP6 and the molecular origins of inhibitory fields of phyllotaxis models. Some biologist colleagues (T. Vernoux and F. Besnard, from ENS Lyon) have gathered a number of experimental data on the loss of function *ahp6* mutant of *Arabidopsis thaliana*, and the activity of the AHP6 gene in the shoot apical meristem. The phyllotaxis of the *ahp6* mutant is particularly perturbed, a fact which led us to study these perturbations in Chapter 3 for both *ahp6* and wild-type data.

## 1.4 Conclusion

Phyllotaxis can be considered as one of the oldest branches of biomathematics [Adler *et al.*, 1997]. Since centuries, scientists from different domains have contributed to shed light on this beautiful natural phenomenon. In this chapter, we first introduced different phyllotactic modes that one can observe in nature. Divergence angle, the number of organs initiating simultaneously as well as the number of parastichies have been defined. Different dynamical models have been proposed in the literature, to reproduce phyllotactic patterns at meristem scale. We have described some of these models that are important for this dissertation. Finally, we have described the main molecular processes that are known to regulate phyllotaxis in the meristem: the active transport of auxin, and the regulation of the central zone. In both cases, we have summarized the main modelling approaches that were used to describe these processes.

Even though the three scales that we have presented are different scales of a unique phenomenon, the models that have been introduced to understand them are usually not explicitly connected. For instance, none of the publications cited in this chapter presents models at more than one of the three scales. The mathematical and computational techniques which are used at these three scales are also different. One of the main goals of this thesis is to improve our understanding of the connections that exist between these different scales, by the use of computational and mathematical techniques. The contents of this chapter provided a summary of the main biological concepts and modelling approaches related to phyllotaxis, at these three scales.

A multiscale study of phyllotaxis is of course too broad to be handled in a sensible manner within the scope of this work, and we will focus on more restricted aspects of phyllotaxis in the following. However, except for Chapter 2 which presents the main formalisms used for gene network modelling, Chapters 3, 4, and 5 concern the three scales we just mentioned, in the same order.

# Chapter 2

# Computational modelling of genetic and biochemical networks

## Introduction

In Chapter 1, we have seen that plant morphogenesis, and phyllotaxis in particular, can be analysed at different scales, from cellular processes to the organization of the stem. If phyllotaxis has been studied for centuries, mostly at macroscopic scales, it is only in the 1960's that the principle of gene regulation has been first understood [Jacob *et al.*, 1961]. Several classes of models of gene regulatory networks have been proposed since this discovery, relying on different formalisms and assumptions. This diversity can be explained by the relative youth of the subject, although it is also undoubtedly related to the inherent complexity of these systems. Indeed, depending on the specific process under consideration, or on the type of available experimental data, different types of models will have to be considered.

In this chapter, we present an overview of the main classes of gene regulatory networks. The importance of genetic and biochemical networks in understanding biological phenomena together with the considerable number publications on this subject indicate the need for such a review in this dissertation.

In view of the diversity of the models of gene regulatory processes, in the current chapter we chose to introduce the models that were extensively used. Though only one formalism is mainly used in this dissertation, an introduction to other formalisms can help better understand the general context of the models. This can help understand why and where a formalism can be chosen.

First, we discuss the most important biological aspects of gene regulation. Then, we present

different mathematical formalisms, first at the single cell level, and second with spatial extension. Finally, we briefly discuss the implications of these different formalisms in the context of this thesis.

## 2.1 Basic biology of gene regulation

Genes are heredity units of living organisms. They are segments of DNA (Deoxyribonucleic acid) which is the molecular storehouse of heredity information. DNA is a double-stranded, helical molecule and is composed of four nucleotides: adenine (A), thymine (T), guanine (G), and cytosine (C). DNA is replicated at cell division so that daughter cells have the same genetic information as the mother cell. In eukaryotes (organisms containing nucleus in their cells) DNA is located in nucleus. Prokaryotes do not have a nucleus and DNA is not in a disclosed area within the cell.

Genomes (i.e. the entirety of DNA code) have been recently sequenced for some organisms. Sequencing is the process of determining the nucleotide order of DNA. The genome of *Arabidopsis thaliana*, the model plant in plant biology and genetics, was the first sequenced plant genome in the year 2000. Later in a subsection we discuss briefly the advantages that this model plant offers for research in genetics.

The information in DNA is copied by an enzyme, called RNA polymerase to an intermediate product called messenger RNA (mRNA). This process is called *transcription*. In eukaryotes, mRNA moves outside the nucleus and attaches to molecular complexes called ribosoms. Before attaching to a ribosom, mRNA can be modified, for example by having pieces cut out. Each ribosom reads the information in mRNA and creates proteins, a process called *translation*. We will call *gene expression* the whole process of gene transcription to mRNA, followed by translation of mRNA intro proteins.

Proteins have several tasks, for example, they can catalyse chemical reactions in cells (enzymes), or they might be involved in cell signalling and signal transduction (e.g. insulin, or auxin), or they may have structural functions. *Transcription factors* (TF) are a particular class of proteins, which regulate the expression of genes.

To understand this, one must enter more into details, and mention the fact that genes are usually composed of two regions: a coding region and a regulatory region. The coding region is the part that will be transcribed to mRNA. In other words, the coding region bears the information that will allow the production of proteins. The regulatory region, on the other hand can bind with some proteins, the TFs we just mentioned. These binding events can occur either directly with a single TF, or with complexes formed of several TFs. As their name suggest, the function of the regulatory region is to regulate the expression of the gene. When they are bound

to a regulatory region, TFs can activate (up-regulate) or repress (down-regulate) the expression of the gene, by facilitating or blocking the binding of RNA polymerase to the coding region, and thus transcription, see Figure 2.1. This means that an increase in the concentration of TFs can lead to more or less related gene expression, respectively.

Living organisms are an assembly of different cells having different characteristics, for example plant leaf cells have a different function from meristematic cells. However nearly all these different cells have the same DNA content. Now the question is: what makes a leaf cell different from a meristematic cell? In other words, we need to know how cell identity is determined. One of the major differences between different kinds of cells is the genes being expressed. To study cell identity in different tissues, it is thus important to understand gene regulation [Gibson *et al.*, 2001, Ptashne *et al.*, 2002].

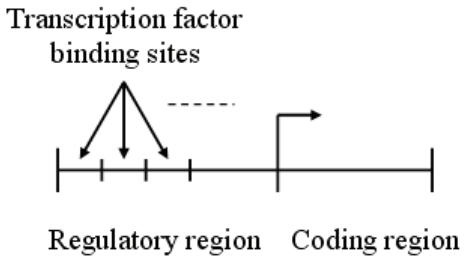
This understanding requires a dynamical point of view on these systems since, for example, mRNAs and proteins are constantly degraded. Ribo-nuclease (RNase) competes with ribosom to bind to mRNA. If it wins the mRNA will be degraded, if mRNA binds to ribosome it will be translated. Proteins are also degraded by proteasomes which are large protein complexes that degrade and recycle proteins. The process is mediated by ubiquitin which binds to the protein and directs it to the proteasomes.

In summary, gene expression is under the influence of various processes of protein degradation and production, the latter being controlled by transcription factors, which are themselves proteins. These interactions between proteins have led to the term *gene regulatory networks* to describe these systems. Their complexity has led to various mathematical representations, that will be presented later in this chapter.

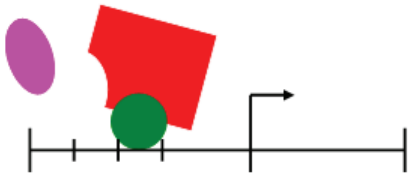
**Arabidopsis thaliana as a model plant**

The complexity and large DNA of cultivated plants make their genetic analysis difficult. However, there are similarities between the DNA of all plants. Biologist communities have relied on these similarities, and chosen to focus on plants with reasonably small genomes. The results of such study can hopefully be used to face the complexity challenge of more complicated genomes.

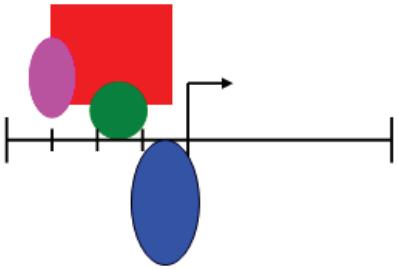
In genetics and developmental biology, *Arabidopsis thaliana* (see Figure 2.2) plays the role that mice and fruit flies (Drosophila) play in animal biology. Though *Arabidopsis* does not have major direct significance for agriculture (in fact it is considered a weed), it has many advantages as a model plant for rapid genetic and molecular analysis. It has one of the smallest higher plant genomes. Rice and maize have genomes four times and twenty times respectively bigger than the genome of *Arabidopsis thaliana*. The genome consists of 5 chromosomes and about 25500 genes [INRA webpage]. The small size of *Arabidopsis* (mature plant is about 30 cm - 40 cm tall), rapid life cycle (about two months) as well as prolific seed production (about 10000 per plant)



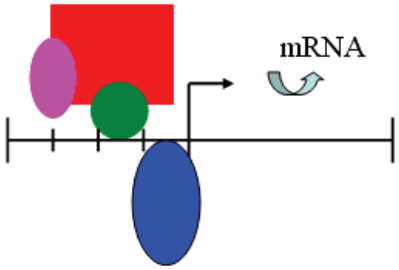

(a) Several transciption factors can regulate a single gene.



(b) Transcription factors form complexes and bind to regulatory region.



(c) RNA polymerase binds to the coding region.




(d) RNA polymerase transcribes the DNA to mRNA.

Figure 2.1: Gene regulation steps. Picture adapted from [Bower *et al.*, 2001].

makes it an ideal model organism for studying plant development. It can be easily cultured in small spaces (nearly 1000 individual per square metre). It is self-pollinated, which allows to produce strains with multiple mutations [INRA webpage], through cross-proliferation.

For all these reasons, this plant has become the most widely used species for plant biologists, and all the data that will be used in this thesis comes from this plant.

## 2.2 Spatio-temporal modelling and simulation

Describing and studying the dynamics of systems in both space and time is called *spatio-temporal modelling*. This leads to a complex modelling framework having both *spatial* and *temporal* properties when compared to describing the dynamics as a function of time only such as chemical kinetics.

Resolving a dynamic process in space highly increases the number of variables. Consider the simple reaction below:

$$A + B \longrightarrow C.$$

This reaction can be modelled using three variables: the concentrations of $A$, $B$ and $C$. Now we imagine that we model it in space by assuming that $A$ and $B$ are produced at certain locations in space and diffuse freely. In consequence, their concentration changes in space. This highly

Figure 2.2: Model plant *Arabidopsis thaliana*. Image from INRA web page [INRA webpage].

increases the number of variables in simulations. Let $N$ be the number of points for which we need the local concentrations, the number of variables rises to $3N$. In fact there could be no limit to the number of points in space and one might be interested in spatial patterns in concentrations fields. Taking infinitely many points is in fact considering the system as a *continuum*. Furthermore, the reactions happening at different points influence each other through diffusion of $A$, $B$ and $C$, or other forms of transport (e.g. auxin transport [Krupinski *et al.*, 2010]) or signalling (e.g. juxtacrine signalling [Owen *et al.*, 1999]).

In spatio-temporal modelling we have often four dimensions, time and three spatial dimensions. Sometimes the space can be one or two dimensional. While time is crucial for the existence of dynamics, spatial aspects are highly important to the study of patterning processes as well as many other processes. For example, one can think of predators and preys, the expression zone of genes, distribution of plant hormones or the spread of a disease because of an epidemic. When we model such systems we should take into account the spatial properties of systems.

Computer simulations allow us to conduct *virtual* experiments studying systems with spatio-temporal properties. The importance of *in silico* experiments is underlined when the *in vivo* experiments are impossible because of the complexity of the system or even because of ethical considerations.

Spatio-temporal modelling has many aspects. Different approaches are used in order to deal with the challenges of such kind of modelling. In this chapter we focus on spatio-temporal modelling gene regulatory networks. Particularly we are interested in modelling gene regulatory networks on three dimensional digitized meristem evolving with time.

**Modelling Categories**

Systems with spatial and temporal properties can be categorized in several ways according to their characteristics. Sbalzarini [Sbalzarini, 2009] proposes three categories: phenomenological v.s. physical, discrete v.s. continuous, and deterministic v.s. stochastic. Each category has its own challenges and different techniques are proposed to deal with them.

A phenomenological description of a system models its overall behaviour, without modelling the underlying mechanisms. They can be used to study the reaction of a model to a perturbation. In physical models on the other hand, a bottom-up approach is used to model the mechanistic behaviour of the system. These models use physical laws to make predictions about the system.

The spatial resolution of the system can be discrete or continuous in space or time. Partial Differential Equations (PDE) are used to formulate continuous models. In the following, we will also introduce cellular automata as simulation tools for a discrete version of PDE models.

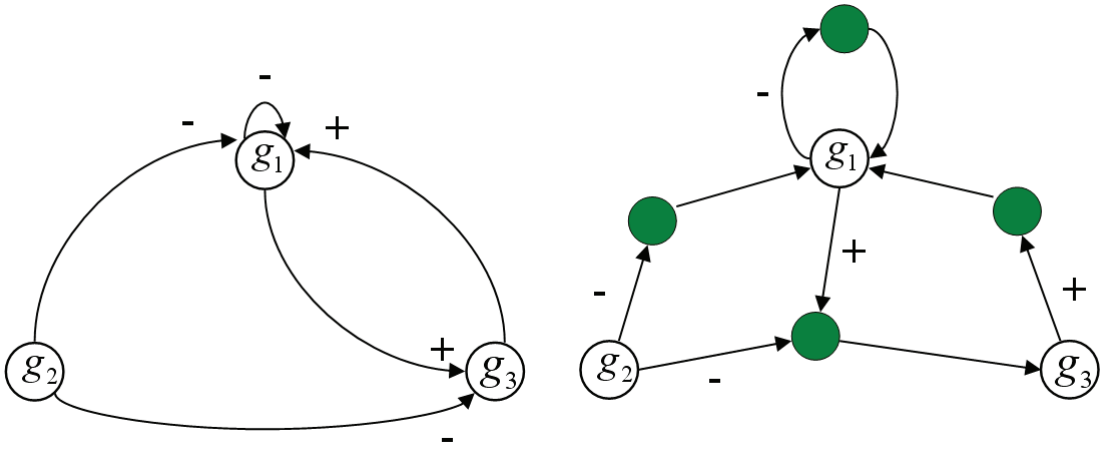
In stochastic models the behaviour of the system is determined not only by the present state and its inputs but also by random fluctuations.

Each category has its own complexities and these complexities create a real challenge. Often computer simulations are the only way to deal with the challenges, as will appear in the rest of this thesis, where computational developments have proved to be a decisive tool in our investigations, using models at various scales, and thus belonging to several of the categories described above.

## 2.3 Computational modelling of gene regulatory networks: formalism review

> "Biologist can be divided into two classes: experimentalists who observe things that cannot be explained, and theoreticians who explain things that cannot be observed"
>
> - George Oster, joking, at commemoration of the 10th anniversary of the death of Aharon Katzir-Katchalsky (biologist and modeller).

The above quotation indicates that theoreticians and experimentalists may not always be in harmony with each other. We will use experimental data in the following chapters. The following section is however exclusively theoretical, and will present the main formalisms that are used to describe gene regulatory networks. There are several reviews discussing computational modelling of gene regulatory networks and their applications in biology [Alvarez-Buylla *et al.*, 2007, Bower *et al.*, 2001, Endy, 2001, de Jong, 2002, Smolen *et al.*, 1999]. Here we present a general overview of the existing approaches. All formalisms are not included in this overview, we rather focus on the formalisms that seem to be more important and more useful for our work. The different classes of models presented hereafter are of course not independent, and we will try to underline the connections that exist between them along the way.




(a) Directed graph. (b) Directed hypergraph.

Figure 2.3: (a) Directed graph representing a gene regulatory network, $V = \{g_1, g_2, g_3\}$, $E = \{(g_1, g_1, -), (g_1, g_2, -), (g_1, g_3, +), (g_3, g_1, +), (g_3, g_2, -)\}$ . (b) Directed hypergraph illustrating a gene regulatory network with cooperative interactions. The edges are represented by green nodes. $V = \{g_1, g_2, g_3\}$, $E = \{(g_1, [g_1], [-]), (g_1, [g_2], [-]), (g_1, [g_3], [+]), (g_3, [g_1, g_2], [+, -])\}$. The plus, $+$, and minus, $-$, symbols sometimes are omitted and are replaced by $\rightarrow$ and $\dashv$ edges, respectively.

### 2.3.1 Interaction graph

The simplest way to represent gene interactions might be a graph where vertices correspond to the genes and edges represent interactions between two or more genes. Consider a directed graph $G = (V, E)$, where $V = \{g_1, ..., g_n\}$ is the set of vertices and $E = \{e_1, ..., e_m\}$, is the set of edges where $e_k \in E$, $e_k = (g_i, g_j, s_k)$, $1 \leq k \leq m$, $1 \leq i, j \leq n$, $s_k \in \{+, -\}$. $V$ corresponds to the set of $n$ genes, and $E$ corresponds to the set of interactions $e_k = (g_i, g_j, s_k)$ such that $e_k = (g_i, g_j, +)$ ($e_k = (g_i, g_j, -)$ respectively) indicates that the gene $g_i$ is activated (inhibited respectively) by gene $g_j$.

Hyper graphs can be used to present cooperativity of transcription factors to regulate gene expression. In this case, the edges can be defined by $e_k = (g_i, \mathcal{G}_j, S_k)$ where $\mathcal{G}$ represents a list of genes that regulate $g_i$ and $S_k$ is a list of signs indicating the regulatory influences of genes in $\mathcal{G}_j$, see Figure 2.3.

Once interaction graph of a gene regulatory network is set up, it can be used to make predictions about the regulatory network. Missing regulatory interactions or redundancy can be detected by searching paths between genes. Feedback networks can be detected by identifying cycles and the structure of the graph can give information about of genes. Clustering algorithms can be used to identify co-regulated genes as well as identifying genes manifesting the same

temporal pattern [Ben-Dor *et al.*, 1999].

As a final remark for this section, interaction graphs do not provide any information regarding the dynamics of gene regulation. In the following section we present the simplest models that take the dynamics into account.

### 2.3.2 Boolean model

A coarse representation of state of a gene can be considering a gene either expressed or not expressed. This can be described by a Boolean variable with value 1 representing active gene and value 0 representing inactive gene. Thus the state of the entire system is a list of zeros and ones corresponding to not expressed or expressed genes. The interactions are represented by boolean functions that return the next state of a gene, thus the next state of system, from the current gene states [Kauffman, 1969, Kauffman, 1993].

Let $X = (x_1, ..., x_n) \in \{0, 1\}^n$ represent the state of a gene regulatory system comprising $n$ elements where $x_i \in \{0, 1\}$, $1 \leq i \leq n$, is the state of element $i$. Given the state of the system at time $t$, $X(t)$, the state of the system at time $t + 1$, $X(t + 1)$ is:

$$X(t + 1) = F(X(t)), \tag{2.1}$$

where $F(X(t)) = (f_1(X(t)), ..., f_n(X(t)))$, $f_i$, $1 \leq i \leq n$ is a Boolean function. The state space of a system of $n$ variables has $2^n$ states, therefore there are $2^{2^n}$ possible logical functions.

The state of the system can be expressed by a *truth table* which specifies the next state for each current state (all together $2^n$ states for a system of $n$ elements). Truth tables and Boolean functions are equivalent, that is from Boolean functions we can set up the truth table, and vice versa, for an example see Table 2.1. A Boolean network can be also represented by logical gates diagrams which are suitable for calculating the transitions between states, see Figure 2.4. The boolean model formalism yields a unique output for an input, thus it is *deterministic*. To calculate the state of the system at time $t+1$, the Boolean functions are evaluated *simultaneously*; such systems are called *synchronous*.

A Boolean function can have different formal expressions. We can simplify the Boolean expressions using Karnaugh maps [Thomas *et al.*, 1990]. Karnaugh maps are practical to simplify expressions with not many variables. In fact a Karnaugh map is another presentation of truth table of a Boolean function see Table 2.2. For each combination of variable values, which is in fact a state, the corresponding value of the function is written in the table. The "1"s in the Karnaugh map are grouped in largest possible rectangular groups such that the number of "1" in the group must be a power of 2. This creates groups of "1"s, then the groups are converted into boolean expressions in two steps:

| $X(t)$ | $X(t+1)$ |
|---|---|
| (0, 0, 0) | (0, 1, 1) |
| (0, 0, 1) | (1, 0, 1) |
| (0, 1, 0) | (1, 1, 1) |
| (0, 1, 1) | (1, 0, 1) |
| (1, 0, 0) | (0, 0, 1) |
| (1, 0, 1) | (1, 0, 0) |
| (1, 1, 0) | (1, 0, 1) |
| (1, 1, 1) | (1, 0, 0) |

$$
\begin{aligned}
x_1(t+1) &= x_2(t) \text{ or } x_3(t),\\
x_2(t+1) &= x_1(t) \text{ nor } x_3(t),\\
x_3(t+1) &= x_1(t) \text{ nand } x_3(t).
\end{aligned}
$$

Table 2.1: Boolean network example; truth table and corresponding Boolean function. $X(t) = (x_1(t), x_2(t), x_3(t))$ and $X(t+1) = (x_1(t+1), x_2(t+1), x_3(t+1))$. See Figure 2.4 to see the corresponding wiring diagram.

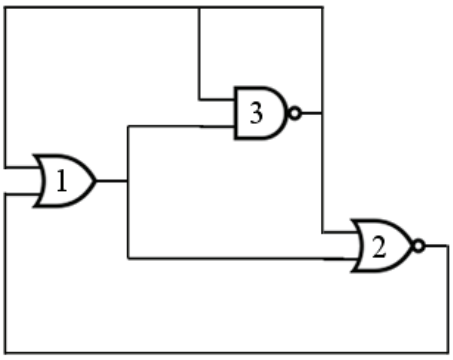



Figure 2.4: Logical gates diagram corresponding to the Boolean network in Table 2.1. The gates 1,2, and 3 are defines the $x_1$, $x_2$, and $x_3$ in the example.

| $f$ | 00 | 01 | 11 | 10 | $x_1, x_2$ |
|---|---|---|---|---|---|
| 0 | • | • | • | • | |
| 1 | • | • | • | • | |
| $x_3$ | | | | | |

| $f$ | 00 | 01 | 11 | 10 | $x_1, x_2$ |
|---|---|---|---|---|---|
| 00 | • | • | • | • | |
| 01 | • | • | • | • | |
| 11 | • | • | • | • | |
| 10 | • | • | • | • | |
| $x_3, x_4$ | | | | | |

Table 2.2: For three and four variable, the input of Karnaugh maps are illustrated. We use one function $f$ per map. The bullets points can be either "0" or "1".

| f | 00 | 01 | 11 | 10 | $x_1, x_2$ |
|---|---|---|---|---|---|
| 00 | 0 | 0 | 1 | 1 | |
| 01 | 0 | 0 | 1 | 1 | |
| 11 | 0 | 0 | 0 | 1 | |
| 10 | 0 | 1 | 1 | 1 | |
| $x_3, x_4$ | | | | | |

Figure 2.5: A Karnaugh map example. Three groups of "1"s are marked. The groups are converted into boolean expressions in 2 step. First, if a variable and its inverse are in the same group, they are removed. This leads to three groups of variables $\{x_1, \bar{x_3}\}$, $\{x_1, \bar{x_2}\}$, and $\{x_2, x_3, \bar{x_4}\}$. The most compact expression for function $f$ is the sum of the products of remaining variables in each group, that is $f(x_1, x_2, x_3, x_4) = x_1.\bar{x_3} + x_1.\bar{x_2} + x_2.x_3.\bar{x_4}$ where ".", "+", and " $\bar{}$ " represent the and, or and not Boolean operators.

- First, if a variable and its inverse are contained in the same group, they are removed.
- Second, each group is converted into a product of remaining variable, the boolean expression is the sum of the products.

For an example see Figure 2.5.

Once a boolean model has been specified, for instance using a truth a table or a Karnaugh map as above, one is usually interested in the dynamics it describes. The first important notion is that of a *trajectory*: given an initial state vector $X^0$, the trajectory of $X^0$ is the series $(X^N)_{N\in\mathbb{N}}$ of iterates under iterations of the function $F$, Equation (2.1):

$$X^1 = F(X^0), \quad \ldots \quad X^N = F(X^{N-1}) = F^N(X^0), \qquad N \in \mathbb{N}$$

Because the state space $\{0,1\}^N$ is finite, and the dynamics is deterministic, each trajectory is in fact finite, *i.e.* it must end up with a periodic sequence. This can be seen for instance by considering the $(2^n+1)$th iterate, for an arbitrary initial condition $X^0$: because each point has a unique successor (its image by $F$), $X^{2^n+1}$ must be equal to one of the $X^j$, for $j < 2^n + 1$, and thus the trajectory contains a cycle.

One particular case of biological interest is that of cycles of length 1, more commonly called *fixed points*, or *equilibrium state*. Indeed, these points correspond to steady states of the network, which are not modified by the dynamics. In general a boolean network may have several of these fixed points, and it has been suggested that they could be interpreted as different cell types in an organism [Kauffman, 1993]. With the notion of truth table above, equilibrium states can be conveniently identified as those lines in the table where $X(t) = X(t+1)$.

The brief presentation we have just made concerns the class of synchronous boolean networks, because all variables are updated simultaneously at each time step, and variables are boolean. Although powerful, and leading to biological interpretations as we just exemplified, these models rely on very strong simplifying assumptions. In some cases however some extensions have been proposed to relax these constraints.

A generalized logical model has been proposed [Thomas *et al.*, 1973, Thomas *et al.*, 1990] where variables may have multiple discrete values (called logical variables) and transitions between states can be asynchronous. In these models, variables take a finite number of values, often interpreted as discretized values $\hat{x}_i$ of a concentration variable, $x_i$. If element $i$ of the gene regulatory networks regulates $p_i$ other variables, it might have at most $p_i$ thresholds:

$$\theta_i^{(1)} < ... < \theta_i^{(p_i)}.$$

Then $\hat{x}_i$ is defined as follows:

$$\hat{x}_i = k - 1, \text{ if } x_i < \theta_i^{(k)}.$$

Therefore $\hat{x}_i \in \{0, ..., p_i - 1\}$. It follows that the state space is no longer boolean, but the product of finite sets $\prod_i \{0, ..., p_i - 1\}$. The function $F$ from Equation (2.1) is thus defined on this new sate space.

The asynchronous character of these models takes the following form: at each time step only one of the variables is updated, say variable $\hat{x}_i$, using the coordinate function $f_i$. Different methods have been proposed to select which variable should be updated at each time step. The most common are either to fix an order on the variables, and updating them successively in that order, or to endow every variable with a delay, or an internal clock (*i.e.* a scalar function of time), and updating the variables whose clock has the lowest value at a given time step [Thomas *et al.*, 1973].

### 2.3.3 Differential equations

Ordinary differential equations are one of the most widespread formalisms to describe the behaviour of a dynamic system in science. They are also used to describe the time evolution of elements in gene regulatory systems. In this case, the state of the system is a list of concentrations of the elements, and the equations are built based on the theory of biochemical

kinetics. Here we present a general form for this type of equations, based on reviews found *e.g.* in [Alon, 2007, Bower *et al.*, 2001, Gibson *et al.*, 2001, de Jong, 2002]

Let us consider a regulatory network including $n$ elements and $X = (x_1, ..., x_n)$ be the state of the system i.e. a vector of concentrations of the elements. A formulation for the time evolution of the system, which generalizes most existing models, is as follows:

$$\frac{dX}{dt} = G(X) - \Gamma X, \tag{2.2}$$

where $G(X) = (g_1(X), ..., g_n(X))$ is a production rate, and $\Gamma$ is diagonal matrix with $\Gamma_{i,i} > 0$, representing the degradation rate of $x_i$ [de Jong, 2002].

The function $g_i(X) : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$ is usually defined as:

$$g_i(X) = \sum_{j \in J} \kappa_{ij}\, b_{ij}(X), \tag{2.3}$$

where $\kappa_{ij} > 0$ is a rate parameter, $J$ is a set of regulating terms, and the function $b_{ij}(x)$ are usually products of terms of the form $h^{\pm}(x_l; \theta_{li}, k_{li})$, with $h^{\pm}$ being very often the *Hill function*, $h^+$ for activation and $h^-$ for inhibition:

$$h^+(x_l; \theta, k) = \frac{x_l^k}{x_l^k + \theta^k}, \qquad h^-(x_l; \theta, k) = \frac{\theta^k}{x_l^k + \theta^k} = 1 - h^+(x_l; \theta, k).$$

The positive parameters $\theta$ and $k$ represent respectively a threshold of the concentration of element $l$ for the regulation of element $i$, and a steepness parameter, see Figure 2.6. When $k = 1$, we have Michaelis-Menten equation which is the rate of the following reaction schema [Zeron, 2007]:

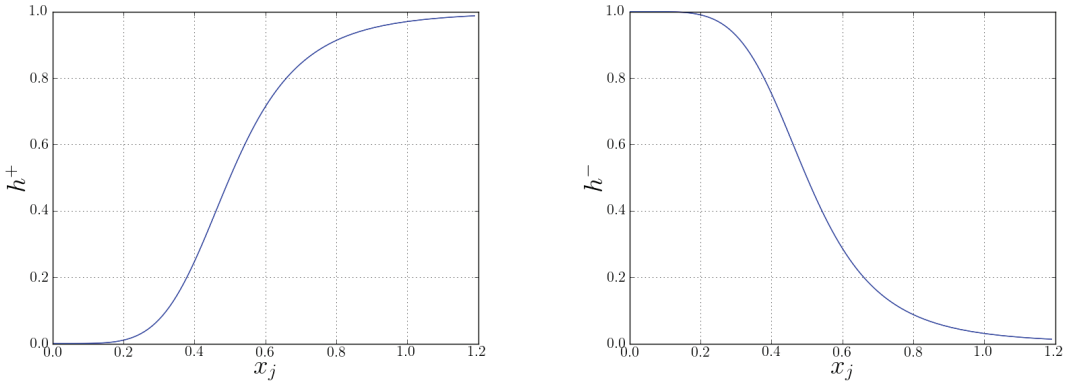
$$F + D \rightleftharpoons F : D \to F + R,$$

where $F$ is transcription factor, $D$ is regulatory region of gene, $F : D$ is the complex of $F$ and $D$, and $R$ is the mRNA. The Hill function generalizes the Michaelis-Menten equation to the case of multiple binding sites between a transcription factor and the regulatory region.

One implicit assumption for equations of the form (2.2) is that the production rate of a gene is directly dependent on the concentration of its regulators. In reality however, gene regulation involves a number of steps, as we have seen in the preceding section: a protein is produced from a segment of mRNA, which is itself produced from DNA through the activity of the regulating proteins. To take these intermediate steps into account, delays can be introduced in the equations [de Jong, 2002], leading to the abstract form:

$$\frac{dx_i}{dt} = g_i(x_1(t - \tau_{i1}), ..., x_n(t - \tau_{in})) - \Gamma_{ii}\, x_i,$$

where $1 \leq i \leq n$, and $\tau_{i1}, ..., \tau_{in} > 0$ represent the discrete time delays.




(a) $h^+(x_j, \theta_j, k)$ (b) $h^-(x_j, \theta_j, k)$

Figure 2.6: regulation functions: (a) Hill curve function $h^+$ to express activation. (b) $h^- = 1 - h^+$ to express inhibition. $\theta = 0.5$, $k = 5$.

Because the production rate function $G$ is non-linear, an analytic solution of the Equation (2.2) is usually not possible. The introduction of delays as above makes an analytic solution even more difficult. Therefore, computer simulations are an essential tool to track the evolution of state of the system over time. As an illustration of this phenomenon, our developments and conclusions in Chapter 5 will heavily rely on numerical techniques.

A particularly relevant problem is the adjustment of parameter values in ODE models. Except some well studied systems such as $\lambda$ phage switch, we do not have quantitative information about kinetic parameters of the equations. There are several ways to estimate parameters of the model [Hohm *et al.*, 2009]:
(i) Adjusting the parameter values by hand such that the model reproduce the qualitative behaviour [Jönsson *et al.*, 2005].
(ii) Studying the system analytically and trying to find indications to choose parameters giving the desired behaviour [Koch *et al.*, 1994],
(iii) Optimizing parameters to minimize the difference between the model simulation output and the output obtained from biological experiments.

In Chapter 5, we will have to use all the three approaches above:
(i) In Section 5.5.1, we adjust the parameter values of a model by hand, to reproduce qualitatively a gene expression pattern on a meristem.
(ii) In Section 5.5.2, we consider a simple model of gene regulation, on an abstract tissue made on a row of cells. This simplification allows us to derive some analytical results on the systems behaviour.
(iii) In Section 5.5.3, we use gene expression patterns that have been designed by an expert

biologist, in order to fit parameters that reproduce these patterns with the best accuracy, using standard optimization techniques.

Besides intensive use of numerical techniques, alternative approaches have been proposed, where the Hill functions appearing in the expression of $G$ are replaced by step functions, as we briefly describe in the next section.

### 2.3.4 Piecewise-linear differential equations

In piecewise linear differential equation models, the regulation of elements by sigmoidal functions is approximated by step functions. The differential equation resulting from this simplification are piecewise linear. Different aspects of such models are studied [de Jong *et al.*, 2004, Edwards, 2000, Farcot, 2006, Glass *et al.*, 1973, Glass, 1975, Gouzé *et al.*, 2003]. Here we present the general form.

The general equations of the model is identical to (2.2), with the Hill functions replaced by Heaviside functions $s^+$ and $s^-$. The functions $s^+$ and $s^-$, can be obtained by taking the limit of the exponent parameter $k$ in Hill functions, and are defined by:

$$s^+(x_j,\theta_j)=\begin{cases}1, \text{ if } x_j > \theta_j\\ 0, \text{ if } x_j < \theta_j\end{cases}$$

and $s^-(x_j,\theta_j) = 1 - s^+(x_j,\theta_j)$.

We will not use these models in the following, but mention them here for sake of completeness, and invite the interested reader to consult the literature mentioned above. The two main remarks that should be made here are, first, that not only these models can be derived from smooth o.d.e.'s as explained above, but they can also be obtained from discrete models by adding a continuous notion of time, and the discretization of concentration values described at the end of section 2.3.2. In other words, they form a natural intermediate class between the type of formalisms presented in the two preceding sections. As a second remark, the qualitative nature of these equations (which stems from the Heaviside terms appearing in the production terms) allow for specific simulation algorithms, which help in particular to decipher some properties that are partially parameter independent, hence reducing the computational cost of the typical procedures described at the end of the previous section [de Jong *et al.*, 2004].

All the models presented in this section are based on description of the dynamics of gene regulation in a single cell. However, cells in a living organism exchange information by various means, a fact that is particularly important in this thesis. As mentioned in general terms in Section 2.2, there exists different frameworks for modelling spatial interactions between dynamical elements. In the context of developmental biology and the mathematical models of gene

induced patterning of organisms, reaction-diffusion models constitute the most studied class. We use variants of these models in Chapter 5, hence we present some basic facts on them in the following section.

## 2.4 Spatial pattern formation with reaction diffusion systems

### 2.4.1 Classical models of reaction diffusion

British ingenious mathematician and computer scientist Alan Turing was the first who showed that the interaction of two diffusive substances with different diffusion rates can create spatial patterns. He proposed a mathematical system in which the pattern arises due to a competition between activation by a slowly diffusing substance, *activator*, and inhibition by a rapidly diffusing substance, *inhibitor*. Turing in his important paper, *The chemical basis of morphogenesis* [Turing, 1952], called activator and inhibitor *morphogens* and he said that his model was very theoretical. However, now the capability of Turing models to generate biological patterns has been used in numerous studies [Koch *et al.*, 1994, Murray, 1989, Murray, 2003].

Turing model in its most generic form can be written as:

$$\begin{cases} \dfrac{\partial A}{\partial t} = D_A \nabla^2 A + f(A,B) & (2.4) \\ \dfrac{\partial B}{\partial t} = D_B \nabla^2 B + g(A,B), & (2.5) \end{cases}$$

where $A$ and $B$ are the concentrations[1] of morphogens, $D_A$ and $D_B$ are the diffusion coefficients of $A$ and $B$ respectively, and $\nabla^2$ is Laplace operator. The above equations are in fact two diffusion equations coupled via two non-linear reaction terms $f$ and $g$. The functions $f$ and $g$ can have different forms for example the Gray-Scott model [Gray *et al.*, 1983, Gray *et al.*, 1984, Gray *et al.*, 1985], the Gierer-Meinhardt model [Gierer *et al.*, 1972, Koch *et al.*, 1994] and the Brusselator model [Prigogine *et al.*, 1968].

In the following, we present two examples of reaction-diffusion systems, the Gierer-Meinhardt model [Gierer *et al.*, 1972] and Brusselator model [Prigogine *et al.*, 1968]. In chapter 5, we present a gene pattern model that uses Brusselator model.

The Gierer-Meinhardt model is an example of activator-inhibitor model described by the following equations:

$$\begin{cases} \dfrac{\partial A}{\partial t} = D_A \nabla^2 A + \rho_A \dfrac{A^2}{(1+\kappa_A A^2)B} - \delta_A A + \sigma_A, & (2.6) \\ \dfrac{\partial B}{\partial t} = D_B \nabla^2 B + \rho_B A^2 - \delta_B B + \sigma_B, & (2.7) \end{cases}$$

[1] To simplify the notations, in this dissertation we use the same symbol for a chemical substance and its concentration.

where $A$ is the activator and $B$ is the inhibitor[2], $D_A$ and $D_B$ are diffusion coefficients, $\delta_A$ and $\delta_B$ are degradation rates, $\rho_A$ and $\rho_B$ the cross-reaction coefficients, $\sigma_A$ and $\sigma_B$ are basic production terms, and $\kappa_A$ is a saturation constant. The term $\frac{A^2}{(1+\kappa_A A^2)}$ introduces an explicit limitation of the activator $A$ by supposing that the production of activator might saturate ($A$ saturates if $A^2 \approx \kappa_A$). By not considering the saturation, the term $\frac{A^2}{(1+\kappa_A A^2)B}$ is replaced by $\frac{A^2}{B}$, and in a more general from by $\frac{A^r}{B^s}$, and similarly the term $\rho_B A^2$ can be replaced by $\rho_B \frac{A^t}{B^u}$. The authors showed that the following condition must be respected to generate a non-homogeneous solution, i.e. a pattern:

$$\frac{st}{u+1} > r - 1.$$

In the Equations (2.6) and (2.7) $r = t = 2$ , $u = 0$, $s = 1$.

The inhibitor should diffuse faster than the activator, therefore $D_B >> D_A$. For $\kappa_A = 0$ the system creates irregularly arranged peaks and for $\kappa_A > 0$ the system can reproduce stripe like patterns.

The Brusselator model that we present now below was developed by Nobel prize winner Ilya Prigogine in Brussels. This model corresponds to chemical reactions as follows:

$$\begin{aligned} U &\to A \\ V + A &\to B + X \\ 2A + B &\to 3A \\ A &\to Y, \end{aligned}$$

where $A$ and $B$ are concentrations of chemical substances that vary spatially and temporally. The concentrations of $U$, $V$, $X$, and $Y$ are kept constant. We can derive mathematical equations corresponding to the above reactions using mass action law which by including the diffusion lead to the Brusselator model:

$$\frac{\partial A}{\partial t} = D_A \nabla^2 A + U - (V+1)A + A^2 B \tag{2.8}$$

$$\frac{\partial B}{\partial t} = D_B \nabla^2 B + VA - A^2 B, \tag{2.9}$$

In all forms of reaction-diffusion models, the basic principle that leads to spatial patterning abilities can be informally explained in a few words, and was at the heart of Turing's original proposal [Turing, 1952]. Assume that a given reaction diffusion system has a homogeneous equilibrium solution, *i.e.* a steady state that is identical at all points in space, and thus with no pattern. This implies in particular that the Laplacian terms in equations (2.4) and (2.5) are zero, and thus that this equilibrium is the solution of the underlying reaction system, with no diffusion.

[2]The activating (resp. inhibiting) substance is recognizable from the fact that the terms it is involved in are increasing (resp. decreasing) functions, except for diffusion or degradation terms.

Then, it is possible to perform a linear stability analysis of this equilibrium, *i.e.* to study its first order response to small perturbations. As usual [Hirsch *et al.*, 2004, Turing, 1952], this is done by linearising the reaction terms in the equations, and studying the real parts of the associated eigenvalues. When one of the latter is positive, the equilibrium is unstable. It turns out that the diffusion terms can lead to a situation where the homogeneous equilibrium, though stable in the pure reaction system, becomes unstable. This means that small perturbations of an homogeneous distribution of substance will lead to an irreversible change of the concentrations, that depart from this uniform situation.
Because the linear analysis is only valid locally, the specific form of the non-linear terms will then control towards what type of other equilibrium the system evolves. At this point non-linear analysis, and most often numerical simulations, are once again of the highest importance. In brief, the most common patterns that are observed are spots, stripes, and spirals (or labyrinthine combinations of these) [Murray, 1989, Murray, 2003].

As a remark that will be useful for Chapter 5, let us underline the fact that the majority of reaction-diffusion models found in the literature are studied on a 2D domain. In particular, the types of patterns we just mentioned do not generalize straightforwardly in 3D. For instance, the spot like patterns are at the basis of a model that uses Brusselator equations to generate a pattern mimicking the central zone of a shoot apical meristem [Jönsson *et al.*, 2005]. This choice makes sense on a planar domain, where a single spot may indeed be reminiscent of the spot like shape of the WUS domain. However, in 3D, numerical simulations show patterns of the Brusselator which are different from the gene expression patterns in the central zone[3].

Remark also that we have presented PDE models in this section, a fact that is materialized by the $\nabla^2$ terms in the equations. When dealing with tissues at a cellular resolution, the coordinates in physical space are no longer continuous, but discrete indices for cells. In this case, $\nabla^2$ is naturally considered as the usual discrete version of the Laplacian, and the PDE become ordinary differential equations with as many variables as there are cells in the tissue for each biochemical substance. This will be made explicit in Chapter 5. The level of generality of the discussion above is such that all mentioned facts remain true in this context, which is in fact also considered in Turing's paper [Turing, 1952].

Finally, because the same mathematical principle underlies all reaction-diffusion based patterning phenomena, and (numerical) experimental evidence indicate that similar patterns are produced by many models (with entirely different mechanistic interpretation), we will mostly consider variants of these models for their qualitative patterning properties, in situations where precise knowledge of the mechanistic underlying processes is missing or incomplete.

[3]See e.g. the intermingled tubes appearing in the simulation: `http://www.youtube.com/watch?v=y3fSpQUc2TQ`

It is also worth noticing here that we will consider in detail, in Chapter 5, a model that strictly speaking could be termed reaction-diffusion, although it does not exactly correspond to the models discussed here. The main difference will be that reaction terms will be reduced to their simplest expression, and diffusion will thus play a prominent role.

### 2.4.2 Using cellular automata for reaction-diffusion systems

Cellular automata (CA) were originally introduced by von Neumann as a formal framework for the study of complicated natural phenomena [Sarkar, 2000]. A cellular automaton consists of a regular array of cells. Each cell is in a finite number of states, such as “0” or “1”. For each a number of finite cells is defined as neighbouring cells. At discrete time steps, each cell updates its state according to its state at previous time step and the state of its neighbours.

The simplest description of a CA is a one-dimensional array, with two states and two neighbours. As we mentioned the state of a cell and states of its neighbours that can be eight different states determines the state of the cell (two different states) at next time step. This leads to $2^8 = 256$ different CA. Such automata are called "elementary cellular automata" by S. Wolfram [Wolfram, 2002]. Each of 256 CA is indexed by a binary number. The decimal representation is called the "rule" for the automaton.

In this chapter, we do not treat the computational aspects of CA as computation universality, but rather we present how cellular automata can be used in modelling natural phenomena. CA are systems with (discrete) space dimensions and (discrete) time evolution, they are used to model spatio-temporal phenomena. There are many applications of CA in modelling biological systems [Alber *et al.*, 2002, Balzter *et al.*, 1998, Ermentrout, 1993, Kier *et al.*, 1996, Luthi *et al.*, 1998, Wurthner *et al.*, 1999]. Here we present a method of simulating reaction-diffusion systems using CA.

The Equations (2.4) and (2.5) defined the Turing reaction-diffusion model which had two morphogens. Reaction-diffusion systems can include any number of chemical substances and in their general form can be written as:

$$\dot{x_i} = D_i \nabla^2 x_i + f_i(x_1, ..., x_n), \tag{2.10}$$

where $x_i$ is concentration, $1 \leq i \leq n$, and $n$ is the number of chemical substances of the system. $D_i$ is the diffusion coefficient of $x_i$. The functions $f_i$ are non-linear. The above equation is the general form of the Turing model as presented in Equations (2.4) and (2.5).

Using cellular automata to simulate reaction-diffusion systems consists in discretizing the space in regular cells and discretizing the time in regular times steps. In CA the state of each cell is also discretized. The state of the system at time $t + \Delta t$ is calculated based on the state of the cell and states of its neighbours at time $t$. This consists in three

steps [Weimar, 1996, Weimar, 2002(a), Weimar, 2002(b)]: first to approximate diffusion, we calculate a sum of concentrations in neighbours. Then, we add the reaction terms, and finally we round it. These steps are formulated by the following equations. Each equation corresponds to one step:

$$\bar{x}(t,\mathbf{r}) = \sum_{\mathbf{r}^* \in \mathcal{N}} a(\mathbf{r}^*)\, x(t, \mathbf{r}+\mathbf{r}^*) \tag{2.11}$$

$$\hat{x}(t,\mathbf{r}) = \frac{1}{c}(\bar{x}(t,\mathbf{r})) + \Delta t\, f(\bar{x}(t,\mathbf{r})) \tag{2.12}$$

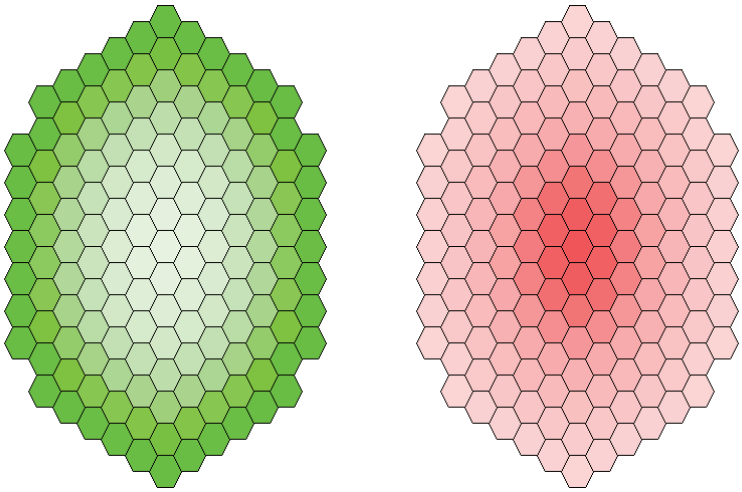
$$x(t+\Delta t, r) = R[\hat{x}(t,\mathbf{r})], \tag{2.13}$$

where $\mathbf{r}$ gives the position in space, $\mathcal{N}$ defines the neighbourhood, for example for seven point Laplacian $\mathcal{N} = (0,0,0),(\pm 1,0,0),(0,\pm 1,0),(0,0,\pm 1)$, $a(\mathbf{r}^*)$ are Laplacian coefficients. $c$ is used to make the coefficients take integer values. The Equation (2.12) adds $\Delta t\, f(\bar{x})$ to the approximated diffusion term. When $\Delta t$ is too big, $\Delta t\, f(\bar{x})$ can be replaced by a term obtained by using higher order numerical integration [Weimar, 2002(a)]. $R$ is a rounding operator and is used to convert $\hat{x}(t,r)$ to a discrete value. There are different type of roundings, e.g. rounding deterministically to the next discrete value [Weimar, 1998].

In cellular automata the steps described by Equations (2.12) and (2.13) are implemented as a table lookup. $\bar{x}$ in Equation (2.11) takes values in a finite set, since $x$ can be in a finite number of states. We calculate the result of Equations (2.11) and (2.12) for each possible value of $\bar{x}$ and we store it in a lookup table. At runtime we only need to calculate the sum in Equation (2.11) and then we retrieve $\hat{x}$ and $x$ from lookup tables. The savings in simulation time can be significant since the functions $f_i$ are only evaluated when we construct the lookup tables. Therefore, the simulation time becomes independent of the non-linear $f_i$ functions. We used cellular automata to simulate a reaction-diffusion model. We simulated one of the models proposed by Jönsson *et al.* [Jönsson *et al.*, 2005], see subsection 1.3.3 in Chapter 1. As we mentioned in Chapter 1, the model is proposed for the WUSCHEL expression domain. In this model a signal originates from the cells in the outermost layer of meristem and represses WUSCHEL gene. The meristem is supposed to be two-dimensional and is discretized by hexagons, see Figure 2.7. See the Appendix B for a simulator of gene regulatory networks using cellular automata. This simulator uses the above mentioned procedure where the space representing a tissue is discretized by hexagonal cells.

## 2.5 Conclusion

In this chapter, we have given a brief overview of gene regulation mechanisms for the non-biologist. Then, we have recalled some basic facts about the main classes of models that are



(a) Repressive signal (b) WUSCHEL

Figure 2.7: Using cellular automata as a simulation tool for modelling reaction diffusion systems. (a) A signal originates form outermost layer of meristem. Its concentration is represented by green colour. This signal represses the expression of WUSCHEL. (b) The concentration of WUSCHEL is represented by red color.

used nowadays to describe and simulate these systems, with a specific focus on spatial patterning and reaction diffusion systems.

It is clear that these different formalisms are related to one another. We have tried to indicate in particular that discrete models can be seen as a convenient approximation of continuous, ordinary or partial differential equations. Notably, we have shown in the last section that discrete models like cellular automata, when considered with care, can provide efficient tools to simulate complex spatio-temporal dynamics.

We also have mentioned the connections that exist between boolean, continuous and piecewise-linear models at the single cell level. Despite these links, it is a striking fact that no full consensus exists on what models should be used to describe gene network dynamics. The main reason is that, contrary to physics for instance, biology is usually difficult to describe by reductionist approaches, as phenomena at very distinct scales in space and time play a major role, and interact with each other. As a consequence, we believe that it is inevitable to make use of a variety of mathematical and computational tools, depending on the specific questions that are addressed, or the experimental data that is used.

We try to illustrate these two facts in the rest of this thesis. First, the multi-scale aspect of vegetal morphogenesis will appear as the three following chapters correspond to three resolutions in space: the macroscopic scale (where atomic elements will be organs), the scale of the meristem (where atomic elements will be primordia), and finally the scale of the underlying biochemical

processes (where the atomic spatial elements will be the cells). Secondly, it will also appear that computational and mathematical tools of a quite different nature will be used in these different chapters, favouring their pertinence over the sake of formal uniformity in the whole thesis.

# Chapter 3

# Modelling and Analysis of Perturbation Patterns in Spiral Phyllotaxis: Application to *ahp6* mutant in *Arabidopsis thaliana*

## Introduction

In Chapter 1 we have presented phyllotaxis as a highly organized arrangement of botanical elements. The geometrical regularity of this phenomenon has drawn attention of scientists for centuries. However, phyllotaxis can be subject to perturbations. Contrasting with the considerable research dealing with the crystal like structure of patterns, few studies have considered the perturbations. For instance, statistical tests have been proposed to distinguish between random and regular phyllotactic patterns, or combinations thereof [Jeune *et al.*, 2004, Jeune *et al.*, 2006]. The statistical methods are applied to the marginal distribution of angles without taking into account the succession of angles and their local dependencies. Perturbations in the phyllotactic patterns have also been observed in a study about transitions between different phyllotactic modes in real plants [Couder, 1998].

In this chapter we study the perturbations of in phyllotaxis considering both wild-type and a phyllotactic mutant. Phyllotactic mutants offer interesting perspectives, notably because they indicate specific genes as potential regulators of phyllotaxis. Further investigations of these genes function will then help us understand the biological mechanisms that lie at the basis of phyllotaxis. In other words, studying perturbations and abnormal events is required to understand

regularity of phyllotaxis.

In the case of this study, we have observed that the perturbations in the mutant sequences are very subtle, and cannot be easily described at the individual plant level. This has led our biologist colleagues to measure sequences of divergence angles for both wild-type and the mutant. It turned out that both the mutant and the wild-type displayed perturbations.

In the following, we describe the experimental measurement of angles. We then discuss how an exploratory analysis of the data has led us to hypothesize that the perturbations can result form permutations of successive organs along the stem. This hypothesis is further validated by a statistical approach, which is based on hidden Markov chains. Then we develop a combinatorial model in which we suppose explicitly that perturbations are due to the permutations. Algorithms are proposed to detect permutations in sequences of angles, and generate all candidate sequences from noisy data. In these algorithms, uncertainty is taken into account by assuming that each measured angle can correspond to several theoretical angles among those predicted by the model; for a given theoretical angle, the corresponding observed angles are modelled by a Gaussian-like distribution for circular data.

## 3.1 Measurement protocol and data set

Fabrice Besnard and Teva Vernoux from ENS Lyon, designed a measurement protocol inspired from previous studies [Peaucelle *et al.*, 2007, Ragni *et al.*, 2008] and built a dedicated device, consisting of a mobile protractor on a vertical structure see Figure 3.1. The main inflorescence of fully grown *Arabidopsis* plants were placed in the center of the protractor and the radial position of each siliques (i.e. *Arabidopsis* fruit) was noted, from bottom to top. They considered the insertion of siliques on the stem, so that possible twisting of pedicels during flower development does not bias the angle value. Sequences always started at the first silique just above the rosette of leaves. They assumed that the transient phase corresponding to the transition from decussate phyllotaxis to spiral one does not overlap the segment of stem measured. Last siliques were discarded when inter-node were not sufficiently elongated or when the height of the stem exceeds the size of the device. Consequently perturbation patterns can be truncated at the beginning or at the end of sequence. They then deduced by first-order differencing the successive divergence angles between pair of consecutive siliques according to two possible orientations (clock-wise or counterclock-wise) of the generative spiral. The orientation of the spiral was determined in an initial exploratory analysis mainly by identifying segments of divergence angles close to canonical Fibonacci divergence angle of 137.5°. In the case of highly perturbed mutant individuals, we used a posteriori our model to select the most likely spiral orientation. We studied wild-type and *ahp6* mutant that belong to the same Columbia ecotype, and thus share the same genetic

background. Our data set consists of successive divergence angles of 82 wild-type plants and 89 *ahp6* mutants. Altogether 5220 angles were measured (2405 angles for wild-type and 2815 for *ahp6* mutant).

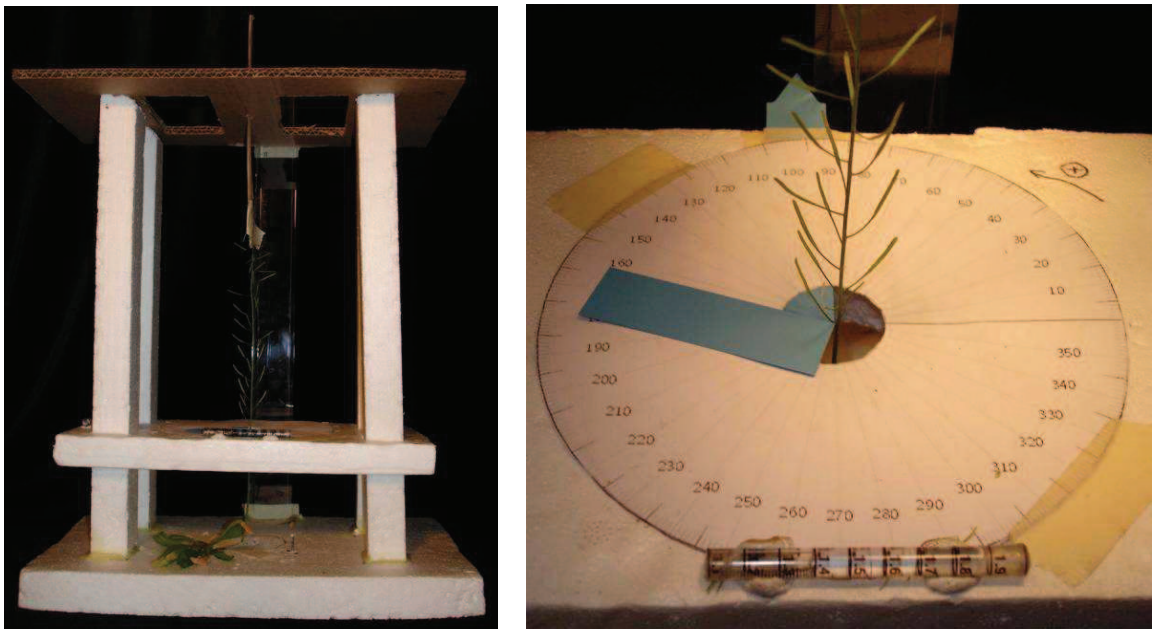


Figure 3.1: Dedicated device to measure divergence angles. It is inspired from previous studies [Peaucelle *et al.*, 2007, Ragni *et al.*, 2008] and is built by Fabrice Besnard and Teva Vernoux from ENS Lyon.

## 3.2 Exploratory analysis

The measured divergence angle sequences took most of the time the form of a baseline where successive measured angles fluctuates around canonical Fibonacci angle 137.5° interspersed by subsequences of non-canonical angles (i.e. different from 137.5°), see Figure 3.3 . The non-canonical angles indicate perturbations of the phyllotaxis (see the histogram of measured angles in Figure 3.2). Surprisingly, phyllotaxis of wild-type plants were frequently highly perturbed. Previous analysis also reported the occurrence of non-canonical angles in wild-type *Arabidopsis* plants [Peaucelle *et al.*, 2007, Ragni *et al.*, 2008]. However, these studies interpreted it as reversal in the orientation of the generative spiral, which would produce segments of $360° - 137.5° = 222.5°$. This interpretation did not seem to fit our data, as *ahp6* plants had much more defects in phyllotaxis, with higher frequencies of non-canonical angles different from 222.5° see Figure 3.2.

For a given plant, let $\alpha$ denote the canonical divergence angle, in case of Fibonacci phyllotaxis $\alpha = 137.5°$, whereas $\alpha = 99.5°$ in case of Lucas phyllotaxis. An ideal sequence without any

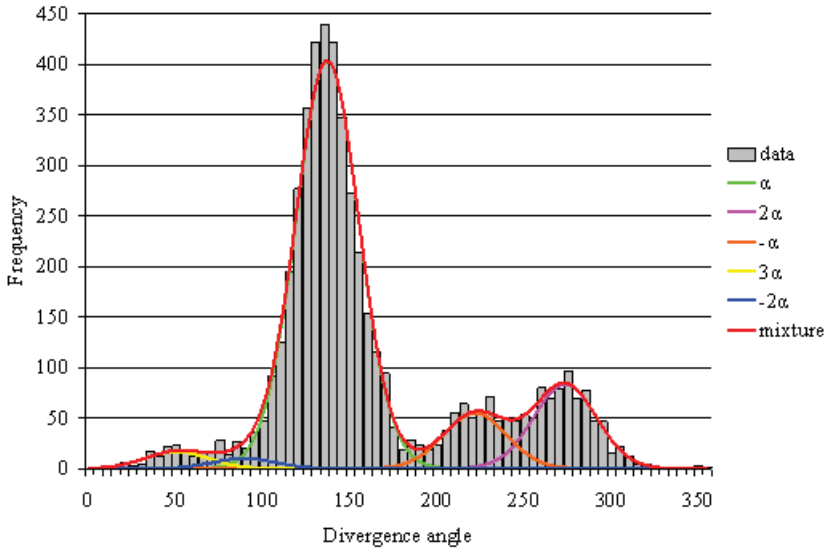



Figure 3.2: Distribution of measured divergence angles. At least four classes are apparent. $\alpha = 137.5°$

perturbation would simply be a repetition of the canonical angle, of the form $(\alpha, \alpha, ..., \alpha)$.

The exploratory analysis of our measured angles highlighted two characteristics of the measured divergence angle sequences:

- The measured divergence angles covered almost all the possible values (between 0 and 360°) with highest frequencies around the canonical Fibonacci angle, i.e. 137.5°. At least four classes of divergence angles were apparent but they were not well separated see Figure 3.2.

- Sub-sequences of non-canonical divergence angles were identified along measured sequences and were more frequent in the mutant. They seemed to be highly structured (i.e. non random repeated divergence angles).

In particular, a motif corresponding approximately to $(2\alpha, -\alpha, 2\alpha)$ was frequently observed in wild-type and even more often in mutants see Figure 3.3. This motif, which was already observed in [Couder, 1998], can be simply explained as the result of a permutation of two consecutive organs on the stem, without changing their angular positions see Figure 3.2. This led us to hypothesize that the segments of non-canonical angles could be explained by permutations involving 2 or more consecutive organs.

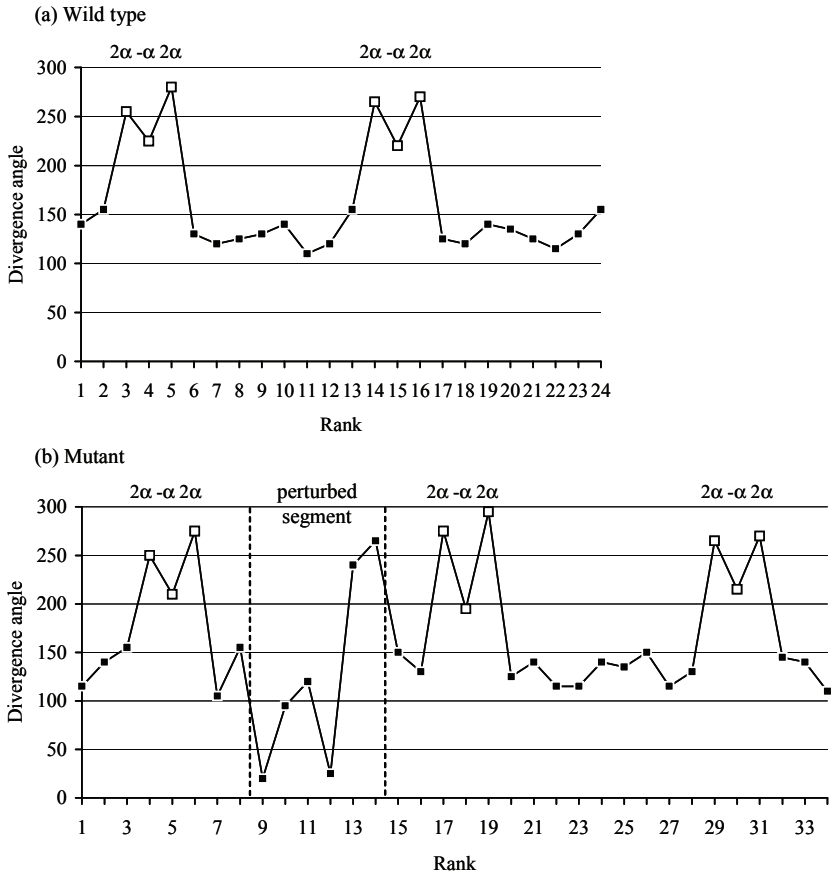
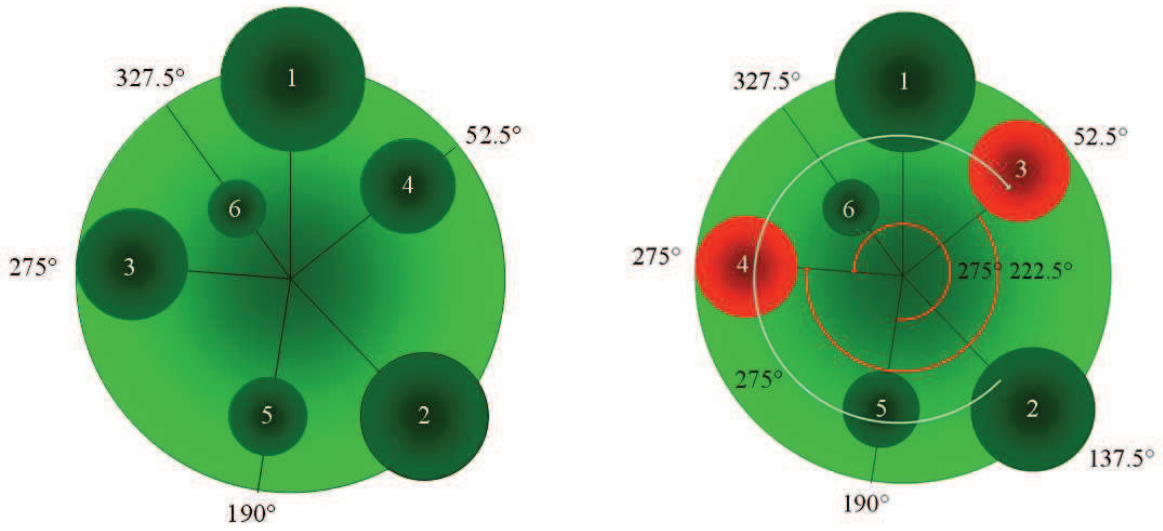



Figure 3.3: Identification of M-shaped motifs corresponding to isolated 2-permutations. The perturbed segments cannot be easily explained on the mutant individual. $\alpha = 137.5°$.

## 3.3 Statistical modelling

A statistical analysis, based on hidden Markov chain models, was performed by Yann Guédon (CIRAD/INRIA Virtual Plants team, Montpellier). Later we will use the results of this approach to parametrise our algorithms, therefore we summarize here the approach.

In a first step, a hidden first-order Markov chain was estimated on the basis of the pooled wild-type + mutant measured divergence angle sequences. In this hidden Markov chain, the states of the non-observable Markov chain represents "theoretical" divergence angles while the von Mises observation distributions attached to each state of the non-observable Markov chain represents measurement uncertainty. The von Mises distribution [Mardia *et al.*, 2000], also known as the circular Gaussian distribution, is a univariate Gaussian-like periodic distribution for a variable $x \in [0, 360°)$. We used the estimated hidden first-order Markov chain to optimally label the measured sequences.




(a) Sequence of divergence angles:$(\alpha, \alpha, \alpha, \alpha, \alpha)$ (b) Sequence of divergence angles: $(\alpha, 2\alpha, -\alpha, 2\alpha, \alpha)$

Figure 3.4: Schema of co-initiation of two organs leading to a 2-permutation on meristem. In (a) the organs are initiated at constant Fibonacci divergence angle ($\alpha = 137.5°$). In (b) the organs 3 and 4 are initiated simultaneously. The co-initiation can lead to a permutation of organs 3 and 4 on the stem. The permutation leads to the $(2\alpha, -\alpha, 2\alpha)$ motif in the sequence of divergence angles. Note that $\alpha$, $2\alpha$ and $-\alpha$ corresponds to 137.5°, 275°, and $360° - 137.5° = 222.5°$ respectively.

In a second step, the memories of a variable-order Markov chain were optimally selected [Csiszár *et al.*, 2006]. See Appendix C for formal definition of variable-order Markov chains and hidden variable-order Markov chain and Appendix D for the principle of the selection of memories. This can be interpreted as a way to identify local dependencies between successive angles. The second step led us to identify frequent patterns corresponding to permutations assumption: e.g. $(2\alpha, -\alpha, 2\alpha)$ corresponds to the organ order 1 3 2 4, and $(2\alpha, -\alpha, 3\alpha, -\alpha, 2\alpha)$ corresponds to the organ order 1 3 2 5 4 6. The latter pattern corresponds to two successive permutations.

In a third step, a hidden variable-order Markov chain was estimated where the underlying variable-order Markov chain has the memories previously selected. The optimally labelled divergence angle sequence was then computed for each observed sequence using the estimated hidden variable-order Markov chain see. For more details about these steps see [Refahi *et al.*, 2010].
The third step allowed us to improve the quality of labelling. One shortcoming of this approach is that some multiples of the canonical divergence angle which occur rarely (*e.g.* $4\alpha, 5\alpha$) as well as alternative phyllotaxis (e.g. Lucas with a canonical divergence angle of 99.5°) cannot be modelled. Yet another significant outcome of the third step is that discarding labels from the transition graph of the estimated hidden Markov chain and restricting it to the modelling of 2-

permutations yields to a finite state automaton that corresponds to the language over theoretical angles induced by 2-permutations of organs see Figure 3.5.

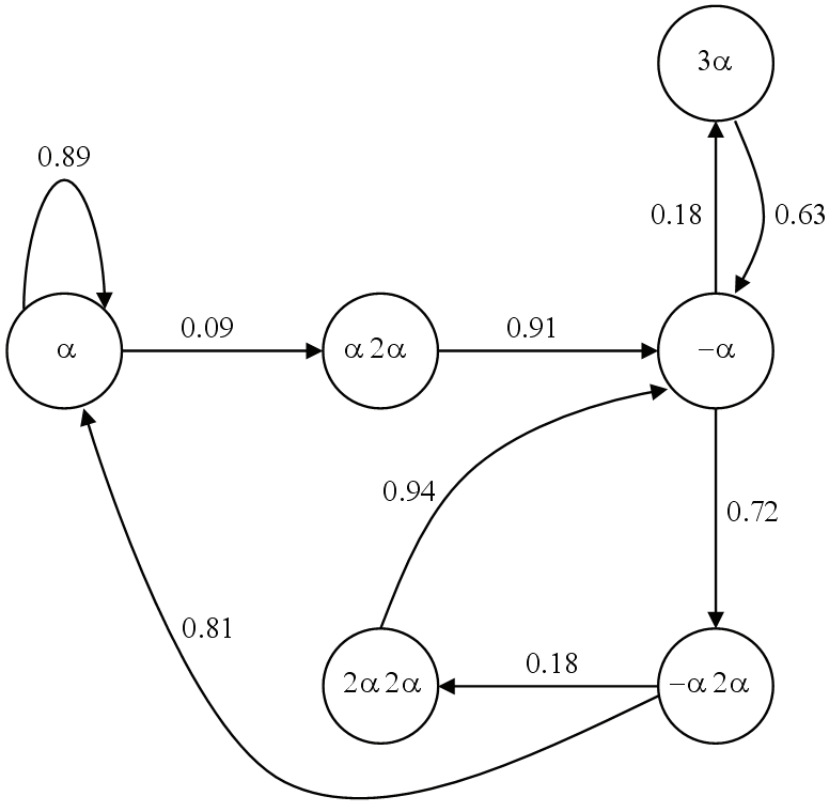



Figure 3.5: Transition graph of hidden variable-order Markov chains restricted to the modelling of 2-permutations. Each vertex represents a possible memory. The labels on edges correspond to the transition probabilities between states. Discarding these labels from this graph leads to a finite state automaton that corresponds to the language over theoretical angles induced by 2-permutations of organs.

This approach showed that without making *a priori* the assumption of permutations, the presence of permutations of two or three organs in our measured sequences is strongly supported by a statistical analysis. Indeed, the finite automaton obtained from the transition graph indicates that the observed perturbations are not likely to be caused by noise and are highly structured. We thus design a combinatorial model in order to label all angles in sequences and test systematically if the observed perturbations could result from permutations of organs.
Note that both the hidden Markov models and the combinatorial models we will present in the following can be understood as latent structure models, where in the combinatorial case the underlying Markov chain is replaced by the combinatorial model defined in the next section. We

have chosen in this chapter to put the accent on the combinatorial model that we have developed.

## 3.4 Combinatorial model formulation

The exploratory analysis put forward the hypothesis that the perturbations could be explained by permutations. This hypothesis is strongly supported by the statistical model as explained in section 3.3. In this section we develop a combinatorial model to explore this hypothesis.

### 3.4.1 Definitions

First we define formally a permutation.

**Definition 3.1.** *Given the set $T_k = \{0, ..., k\}$, a permutation $\sigma$ of $T_k$ is a bijective function $\sigma : T_k \to T_k$. One might see a permutation $\sigma$ as an ordered rearrangement of the elements in $T_k$ where $\sigma(i)$ is the element in position $i$. Hereafter we will denote $\sigma(i)$ by $\sigma_i$.*

Now by an example we describe the intuition that guided us through the formalisation of the model.

**Example 3.1.** *Suppose that the sequence $(0, 1, 2, 4, 3, 5, 6)$ is a sequence of organs order in which the organs number 3 and number 4 are permuted. This sequence of organs order corresponds to the sequence of divergence angles $S = (\alpha, \alpha, 2\alpha, -\alpha, 2\alpha, \alpha)$. $\alpha$ is the canonical divergence angle as stated before. As we can see, the divergence angles are multiples of $\alpha$. If we calculate the sequence of absolute angles and we add 0 to the beginning of the sequence, we obtain the sequence $(0, \alpha, 2\alpha, 4\alpha, 3\alpha, 5\alpha, 6\alpha)$. By dividing the elements of the latter sequence by $\alpha$, we obtain $(0, 1, 2, 4, 3, 5, 6)$ which turns out to be the initial sequence of organs order.*

Now in the following, we formalise the intuition of above example. Let $S = (\mu_1, ..., \mu_\ell)$ be a sequence of divergence angles. If permutations occur, all terms in the sequence $S$ are of the form

$$\mu_j \in \alpha\mathbb{Z}^* = \{i\alpha \,|\, i \in \mathbb{Z},\, i \neq 0\}.$$

We define the corresponding absolute angles as follows:

$$v_0 = 0, \quad v_i = \sum_{j=1}^{i} \mu_j, \quad V(S) = (v_0, ..., v_\ell), \quad i \in \{1, ..., \ell\}. \tag{3.1}$$

From $V$ we define a sequence containing the order of appearance of organs the first being 0. We name it *order index series* of $S$, denoted by $U(S)$, or simply $U = (u_0, u_1, ..., u_\ell)$ when $S$ is clear from the context:

$$u_i = \frac{1}{\alpha}(v_i - v_J), \quad 0 \leq i \leq \ell \qquad \text{where} \qquad J = \arg\min_{j \in \{0 \cdots \ell\}} v_j. \tag{3.2}$$

$U$ is in a bijective relationship with the sequences of divergence angles.

From the definition it is clear that if $J > 0$ then $v_J < 0$ since $v_0 = 0$. This may occur when the sequence starts with permuted organs, a fact related to the left truncation of an observed sequence with respect to complete sequence, see Section 3.1. If the sequence $S$ follows a spiral phyllotaxis, i.e. $S = (\alpha, ..., \alpha)$, we have

$$u_i = i, \quad \forall i \in \{0, ..., \ell\}.$$

In a permuted sequence we are interested in special blocks, called shuffled blocks.

**Definition 3.2.** *Given a permutation $\sigma$ of $\{0, ..., \ell\}$, $(\sigma_j, ..., \sigma_{j+k})$ is called a shuffled block of size $(k+1)$ if:*

$$\{\sigma_j, ..., \sigma_{j+k}\} = \{j, ..., j+k\}.$$

*In other words, $(\sigma_j, ..., \sigma_{j+k})$ is a permutation of $\{j, ..., j+k\}$.*

**Definition 3.3.** *A permutation is indecomposable, if and only if none of its prefixes is a shuffled block.*

**Example 3.2.** *Given $\sigma = (2, 1, 0, 5, 3, 4)$, the three shuffled blocks are: $(2, 1, 0)$, $(5, 3, 4)$ and $(2, 1, 0, 5, 3, 4)$. Amongst them $(2, 1, 0)$ and $(5, 3, 4)$ are indecomposable.*

Indecomposable permutations were introduced by Comtet [Comtet, 1972, Comtet, 1974]. Let $I_i$ denotes the set of indecomposable permutations of length $i$. Comtet proved that $p_i = |I_i|$ (i.e. the cardinal number of $I_n$) has the following generating function[1]:

$$f(t) = 1 - \frac{1}{\sum_{n-1}^{\infty} i! t^i},$$

and

$$\lim_{i \to \infty} p_i / i! = 1. \tag{3.3}$$

Furthermore, King [King, 2006] proved that $\forall i \geq 1$, $p_i/i! \geq 1/2$.

In [Comtet, 1974] Comtet derives the asymptotic series for these numbers

$$p_i = i!(1 - 2i^{-1} - (i(i-1))^{-1} - 4(i(i-1)(i-2))^{-1} + O(i^{-4}))$$

where $i$ is the length of indecomposable sequence. The first ten elements of the sequence are

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $p_i$ | 1 | 1 | 3 | 13 | 71 | 461 | 3447 | 29093 | 273343 | 2829325 |

[1] A generating function is formal power series whose coefficients give information about a sequence of numbers. A formal power series can be considered as a polynomial with infinitely many terms.

This sequence is registered in The On-Line Encyclopedia of Integer Sequences (OEIS)[2] and is referenced by A003319, where various properties and related sequences are given. Later we use $p_i$ to enumerate specific sequences that we call $n$-admissible sequences. We call a sequence $S$ of divergence angles $n$-admissible if a finite number of permutations, applied to disjoint blocks of at most $n$ successive organs, results in the order index series $(0, ..., \ell)$.

**Definition 3.4.** *A sequence $S = (\mu_1, ..., \mu_\ell) \in (\alpha\mathbb{Z}^*)^\ell$ is $n$-admissible, if and only if the maximum size of indecomposable shuffled blocks of its associated order index series is $n$, $n \in \{1, ..., \ell + 1\}$. More precisely $S$ is $n$-admissible for some $n \in \{1, ..., \ell + 1\}$, if and only if its associated order index series $U$ satisfies:*
$\forall i \in \{0, 1, ..., \ell\},\ u_i \neq i \Rightarrow \exists j, k \in \{0, 1, ..., \ell\},\ j \leq i \leq k,\ k - j + 1 \leq n,\ (u_j, ..., u_k)$ *is a shuffled block.*

Note that $n$ can be at most $\ell + 1$, since $n$ is the number of organs involved in permutations. If a sequence is $n$-admissible it is $m$-admissible for $m > n$.

**Property 3.1.** *If $S$ is $n$-admissible, then $U$ is a permutation of $(0, ..., \ell)$. The converse holds only for a certain $n \in \{1, ..., \ell\}$.*
*$U$ is a permutation of $(0, ..., \ell) \iff \exists n \in \{1, ..., \ell + 1\}$ s.t. $S$ is $n$-admissible.*

**Example 3.3.** *The sequence $S = (-\alpha, 2\alpha, 3\alpha, -\alpha, -\alpha, 3\alpha)$ is 3-admissible, but not 2-admissible. Indeed, its absolute angles are $V = (0, -\alpha, \alpha, 4\alpha, 3\alpha,\ 2\alpha, 5\alpha)$. Hence $v_J = v_1 = -\alpha$, and $U = (1, 0, 2, 5, 4, 3, 6)$. Then, $(u_0, u_1) = (1, 0)$ and $(u_3, u_4, u_5) = (5, 4, 3)$ are shuffled blocks of length at most 3, and suffice to reconstruct the canonical sequence $(0, 1, 2, 3, 4, 5, 6)$.*

Divergence angles of $n$-admissible sequences only belong to a finite subset of $\alpha\mathbb{Z}^*$:

**Property 3.2.** *The divergence angles of an $n$-admissible sequence take values in:*

$$D_n = \{i\alpha \,|\, (1 - n) \leq i \leq (2n - 1),\ i \neq 0\}\,.$$

*Proof.* Up to the multiplicative constant $\alpha$, $S$ is the first-order differenced sequence of $U$. Let $\mu_i$ be a divergence angle in an $n$-admissible sequence, we know from Equation (3.1)-(3.2) that $\mu_i = (u_i - u_{i-1})\alpha$. There are four possible cases for $u_{i-1}$ and $u_i$:

1. Neither $u_{i-1}$ nor $u_i$ are in any shuffled block, so $u_{i-1} = (i - 1)$, $u_i = i$ and $\mu_i = \alpha$.

2. $u_{i-1}$ is in a shuffled block but $u_i$ is not in a shuffled block, so $u_i = i$, and $u_{i-1} \in \{(i - n), ..., (i - 2)\}$ then $\mu_i \in \{2\alpha, ..., n\alpha\}$.

[2] http://oeis.org/

3. $u_{i-1}$ is not in a shuffled block but $u_i$ is in a shuffled block, so $u_{i-1} = (i-1)$ and $u_i \in \{i+1, ..., (i+n-1)\}$ then $\mu_i \in \{2\alpha, ..., n\alpha\}$.

4. Both $u_{i-1}$ and $u_i$ are in a shuffled block.

   - $u_{i-1}$ and $u_i$ are in the same shuffled block so $u_i, u_{i-1} \in \{j, ..., (n+j-1)\}$, for some $j < i-1$. Hence $\mu_i \in \{(1-n)\alpha, ..., (n-1)\alpha\} \backslash \{0\}$.
   - $u_{i-1}$ and $u_i$ are in two different but chained shuffled blocks so $u_{i-1} \in \{(i-n), ..., (i-2)\}$, $u_i \in \{i+1, ..., (i+n-1)\}$ then $\mu_i \in \{3\alpha, ..., (2n-1)\alpha\}$.

□

In the following, we refer to $D_n$ as the set of theoretical angles.

As we mentioned in the exploratory analysis, the apparent classes of measured divergence angles are not unambiguously separated, see Figure 3.2. These classes represent the theoretical angles and the overlap between adjacent classes leads to several candidate theoretical angles for a measured angle. Consequently several sequences of divergence angles might be attributed to a measured sequence. Later they will be stored in a tree.

**Definition 3.5.** *A labelled tree $T = (V, E, L)$, where $L : V \to D_n$, $D_n$ is the set of theoretical angles, is called an n-admissible tree if all leaves have a common depth $\ell \in \mathbb{N}$, and every path from the root to a leaf is labelled by an n-admissible sequence.*

### 3.4.2 Properties of n-admissible sequences

In this section we describe how $n$-admissible sequences are affected by operations such as order inversions or concatenation. They will be used afterwards to design algorithms.

It will be useful in the last section to scan sequences backward. One shall then rely on reversibility of $n$-admissible sequences:

**Proposition 3.1.** *Let $S = (\mu_1, ..., \mu_\ell)$ be a sequence of divergence angles. $S$ is n-admissible if and only if the reverse sequence $S' = (\mu_\ell, ..., \mu_1)$ is n-admissible.*

*Proof.* Let $U(S) = (u_0, u_1, ..., u_\ell)$ be the order index series of $S$, we know that $S = ((u_1 - u_0)\alpha, ..., (u_\ell - u_{\ell-1})\alpha)$ and $S' = ((u_\ell - u_{\ell-1})\alpha, ..., (u_1 - u_0)\alpha)$. According to Property 3.1 we have $\max(u_i) = \ell$. Moreover we know that

$$V(S')/\alpha = (0, (u_\ell - u_{\ell-1}), (u_\ell - u_{\ell-2}), ..., (u_\ell - u_0)) = u_\ell - (u_\ell, ..., u_0)$$

and the minimal element of this sequence is:

$$v'_J/\alpha = \min_{i\in\{0,\cdots,\ell\}}(u_\ell - u_i) = u_\ell - \max_{i\in\{0,\cdots,\ell\}}(u_i) = u_\ell - \ell,\ i \in \{0,1,...,\ell\}.$$

Then

$$U(S') = u_\ell - (u_\ell,...,u_0) - (u_\ell - \ell) = \ell - (U(S))'$$

where $(U(S))' = (u_\ell,...,u_0)$ is the reversed order index series of $S$. It is clear that the latter can be decomposed into shuffled blocks of length $\leq n$ iff $U(S)$ itself can. Since $U(S')$ is seen above to be a translation of this reversed sequence it also shares this property.

□

$n$-admissible sequences are the sequences of divergence angles between consecutive organs from bottom to top. The reverse of sequence of divergence angles can be interpreted as divergence angles between consecutive organs from top to down.

Now let us investigate the concatenation of $n$-admissible sequences. Let $S$ and $P$ be two sequences of divergence angles. In addition let $S$ be $n$-admissible. After two preliminary propositions we show that the concatenation of $S$ and $P$ can be $n$-admissible under some conditions.

**Proposition 3.2.** *Let $S = (\mu_1,...,\mu_\ell)$ be $n$-admissible and $V(S)$ and $U(S)$ be the sequence of absolute angles and the order index series respectively. Then $J < n$ and $0 \leq -v_J < n\alpha$, where $J$ is defined as in Equation (3.2).*

*Proof.* First we prove that $J < n$. By construction of order index series we know that $u_J = (v_J - v_J)/\alpha = 0$.
If $J = 0$ then $J < n$. We suppose that $J \neq 0$, therefore $u_J \neq J$ and by Definition 3.4, $\exists j,k \in \{0,1,...,\ell\},\ j \leq J \leq k,\ k-j+1 \leq n, u_J = 0 \in \{u_j...u_k\} = \{j...k\}$. It follows that $j = 0$, and $k - j + 1 = k + 1 \leq n$. Hence $J \leq k < n$.
Now we prove $0 \leq -v_J < n\alpha$. Since $u_0 = -v_J/\alpha \geq 0$ by definition, this amounts to $0 \leq u_0 < n$, in which only the second part remains to be proved. It holds obviously for $u_0 = 0$. Otherwise, $u_0$ is part of a shuffled block $\{u_j...u_k\} = \{j...k\}$, where $j = 0$, and $k < n$, whence $u_0 < n$.

□

The order index series of two concatenated sequences does not always begin with the order index series of the first sequence.

**Proposition 3.3.** *Let $S = (\mu_1,...,\mu_i)$, $P = (\mu_{i+1},...,\mu_{i+k})$ be sequences of divergence angles in $D_n$ and $U(S) = (u_0,...,u_i)$ and $U(P) = (u'_0,u'_1,...,u'_k)$ be the order index series of $S$ and $P$ respectively. Suppose that $S$ is $n$-admissible and $u'_0 = 0$. If $i \geq n$, then $U(S)$ is prefix of $U(S \cdot P)$ where $S \cdot P = (\mu_1,...,\mu_{i+k})$ denotes the concatenation of $S$ and $P$.*

*Proof.* Let $V(S) = (v_0, ..., v_i)$, $V(P) = (v'_0, ..., v'_k)$. From the Equation (3.1) we know that $V(S \cdot P) = (v_0, ..., v_i, v'_0 + v_i, ..., v'_k + v_i)$ and thus $V(S)$ is prefix of $V(S \cdot P)$. Let $J$, $J'$ and $J''$ denote the indexes of minimum elements of $V(S)$, $V(P)$ and $V(S \cdot P)$ respectively as defined in Equation (3.2). $U(S)$ is a subsequence of $U(S \cdot P)$ if and only if $v_J = v_{J''}$. $u'_0 = v'_0 - v_{J'} = 0 - v_{J'}$, $u'_0 = 0$ hence $v_{J'} = 0$. From Equation (3.1) we know that $v_{J''} = \min(v_J, (v_{J'} + v_i)) = \min(v_J, v_i) = v_J$.

□

**Proposition 3.4.** *Let $S = (\mu_1, ..., \mu_i)$, and $P = (\mu_{i+1}, ..., \mu_{i+k})$ be sequences of divergence angles and $U(S) = (u_0, ..., u_i)$ and $U(P) = (u'_0, u'_1, ..., u'_k)$ be the order index series of $S$ and $P$ respectively. Suppose that $S$ is $n$-admissible, $i \geq n$ and $u'_0 = 0$.*
*Then, the concatenated sequence $S \cdot P = (\mu_1, ..., \mu_{i+k})$ is $n$-admissible iff*

$$P|_{u_i} \doteq \Big(\mu_{i+1} + (u_i - i)\alpha,\ \mu_{i+2}, ..., \mu_{i+k}\Big) \quad \text{is } n\text{-admissible.} \tag{3.4}$$

*Proof.* We use again the identity $\mu_i = \alpha(u_i - u_{i-1})$.
Let $U = U(S \cdot P) = (u_0, ..., u_{i+k})$ denote the order index series $S \cdot P$. Since from Proposition 3.3 we know that $U(S)$ is a subsequence of $U(S \cdot P)$, we can easily show that $U(S \cdot P) = U(S) \cdot (u'_1 + i, ..., u'_k + i)$. From Property 3.1 we know that $\{u_0, ...u_i\} = \{0, ..., i\}$. Since $S$ is $n$-admissible and $u'_0 = 0$ then it is clear that $S \cdot P$ is $n$-admissible if and only if $(u_{i+1}, ..., u_{i+k})$ is a permutation of $(i+1, ..., i+k)$ that can be decomposed into disjoint shuffled blocks of length $\leq n$, or equivalently for $(u_{i+1} - i, ..., u_{i+k} - i)$ and $(1, ..., k)$. In other words, $S \cdot P$ is $n$-admissible iff $(u_{i+1} - i, ..., u_{i+k} - i)$ is the order index series of an $n$-admissible sequence. From the initial remark, the divergence angle sequence leading to this order index series can be written as

$$\begin{aligned}
&\alpha\Big(u_{i+1} - i, (u_{i+2} - i) - (u_{i+1} - i), ..., (u_{i+k} - i) - (u_{i+k-1} - i)\Big) \\
&= \alpha\Big(u_{i+1} - i, u_{i+2} - u_{i+1}, ..., u_{i+k} - u_{i+k-1}\Big),
\end{aligned}$$

where the multiplication by $\alpha$ is applied to each component. Then, the same remark again shows that this sequence is exactly $P|_{u_i}$.

□

**Example 3.4.** *$S = (\alpha, \alpha, 2\alpha, -\alpha)$ and $P = (2\alpha, \alpha, \alpha)$ are two sequences of divergence angles. Let $U(S) = (u_0, u_1, u_2, u_3, u_4) = (0, 1, 2, 4, 3)$ be the order index series of $S$. $S$ is 2-admissible. Denoting $P = (\mu'_1, \mu'_2, \mu'_3)$ we define $P|_{u_4} = ((\mu'_1 + (u_4 - 4)\alpha), \mu'_2, \mu'_3) = (\alpha, \alpha, \alpha)$. $P|_{u_4}$ is 2-admissible. $S \cdot P = (\alpha, \alpha, 2\alpha, -\alpha, 2\alpha, \alpha, \alpha)$ is 2-admissible. ($U(P|_{u_4}) = (0, 1, 2, 3)$, $U(S \cdot P) = (0, 1, 2, 4, 3, 5, 6, 7)$).*

### 3.4.3 Geometrical property of $n$-admissible sequences

Now we introduce a proposition for order index series of $n$-admissible sequences. Later it can be used to prune the set of theoretical angles corresponding to a measured angle.

**Proposition 3.5.** *Let $S = (\mu_1, ..., \mu_\ell)$ be an $n$-admissible sequence and $U = (u_0, ..., u_\ell)$ be the order index series of $S$. Then $|u_i - i| < n$, $0 \leq i \leq \ell$.*

*Proof.* $S$ can be decomposed into shuffled blocks. This proposition holds for element of any shuffled block. Since from Definition 3.4 a shuffled block $(u_j, ..., u_k)$ is a permutation of $\{j, ..., k\}$ such that $k - j + 1 \leq n$.

□

Given $S$ a sequence of divergence angles, one can represent its order index series $U(S) = (u_0, ..., u_\ell)$ by a set of points with coordinates $(i, u_i)$ in Cartesian plane. According to the latter proposition the points associated with an $n$-admissible sequence of length $\ell$ must be within the area defined by:

$$\{(x, y) | x, y \in \{0, ..., \ell\}, x < y + n, \, x > y - n\}.$$

See Figure 3.6.

**Remark 3.1.** *The condition stated in the Proposition 3.5 is a necessary condition for $n$-admissibility. The converse of the Proposition 3.5 is not true in general. As a counter example, let $U = (0, 1, 3, 5, 2, 4)$ be the order index series of some $n-$admissible sequence. $\forall i \in \{0, 1, ..., 5\}$, $|u_i - i| < 3$ but the sequence is not 3-admissible.*

The converse of the Proposition 3.5 is true when $n = 2$. For $n > 2$ the condition presented in the proposition, is far from being sufficient.

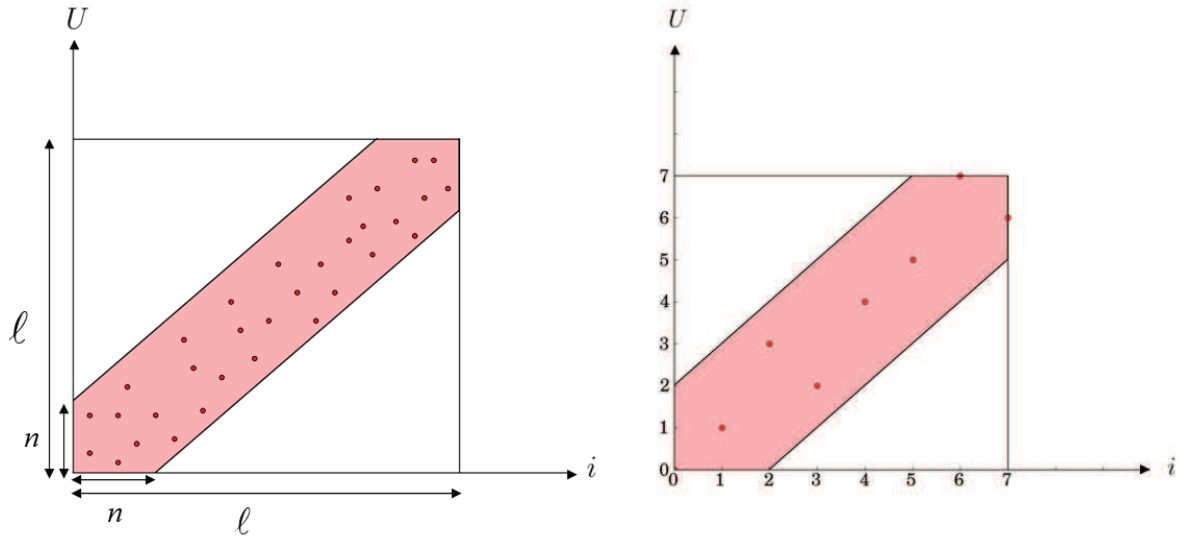
### 3.4.4 Enumeration of n-admissible sequences

In the preceding subsections, we introduced $n$-admissible sequences and we discussed properties such as reversibility and concatenation. In this section we study the probability of a random sequence of divergence angles to be $n$-admissible. To achieve this aim we first enumerate $n$-admissible sequences.

**Proposition 3.6.** *Let $N_n^\ell = \{S = (\mu_1, ..., \mu_\ell) \,|\, S$ is $n$-admissible$\}$ and $L_n^{\ell+1} = \{U(S) | S \in N_n^\ell\}$. Then $N_n^\ell \subseteq N_i^\ell$, $\forall i \geq n$ and $|N_n^\ell| = |L_n^{\ell+1}|$.*

*Proof.* An $n$-admissible sequence is $i$-admissible, $\forall i \geq n$ hence $N_i^\ell \subseteq N_n^\ell$. $U$ is bijective therefore $|N_n^\ell| = |L_n^{\ell+1}|$.

□




(a) Graphical representation of order index series in $xy$-plane.

(b) An example of representation of order index series in $xy$-plane.

Figure 3.6: Graphical representation of order index series in $xy$-plane. (a) $U(S)$ can be represented by a set of points with coordinates $(i, u_i)$ in $xy$ plane. The points must be within the area marked by red color. (b) Example of a graphical representation of order index series $U = (0, 1, 3, 2, 4, 5, 7, 6)$ corresponds to a 2-admissible sequence and it can be represented in $xy$-plane. The red points are located in the area marked by red color.

Note that order index series of an $n$-admissible sequence of length $\ell$, has $\ell + 1$ elements, that is why in above proposition we have $L_n^{\ell+1}$.

**Proposition 3.7.** *Let $t_n^{\ell+1} = |L_n^{\ell+1}|$. The following recursive relationship holds:*

$$t_n^{\ell+1} = \sum_{i=1}^{n} (p_i t_n^{\ell+1-i})$$

*where $p_i$ is the number of indecomposable permutations of length $i$.*

*Proof.* By definition order index series of an $n$-admissible sequence begins by an indecomposable permutation of size $i$, $1 \leq i \leq n$. The recursive formula is in fact the sum of number of order index series beginning with an indecomposable permutation of all possible length. □

From Proposition 3.7 it is clear that $|N_n^{\ell}| = t_n^{\ell+1}$.

As mentioned in Property 3.1 the elements of $L_n^{\ell+1}$ are permutations of $(0, ..., \ell)$. To see how large is the set of $n$-admissible sequences of length $\ell$ compared to the set of permutations of $(0, ..., \ell)$, we calculated $t_n^{\ell+1}/(\ell+1)!$, see Figures 3.7 and 3.8. Below we give the values of $t_n^{\ell+1}/(\ell+1)!$ for $n \in \{1, 2, ..., 10\}$ in two tables for $\ell = 9$ and $\ell = 34$).

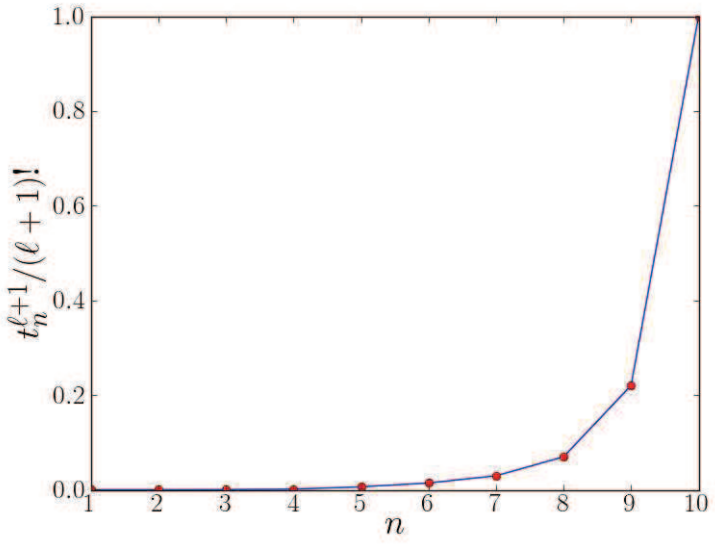



Figure 3.7: $t_n^{\ell+1}/(\ell+1)!$ for $\ell = 9$. The ratio is 1 for $n = 10$. Note that $n$ is the maximum number of organs involved in a permutation and $\ell$ is the number of angles of $n$-admissible sequence.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $t_n^{10}/10!$ | 2.7e-07 | 2.4e-05 | 2.9e-04 | 1.5e-03 | 6.2e-03 | 0.01 | 0.03 | 0.07 | 0.2 | 1.0 |

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $t_n^{35}/35!$ | 9.6e-41 | 1.4e-33 | 1.6e-29 | 1.1e-26 | 1.8e-24 | 1.2e-22 | 4.1e-21 | 9.3e-20 | 1.8e-18 | 1.9e-17 |

As we can see in the figures 3.7 and 3.8 and above tables the ratio $t_n^{\ell+1}/(\ell+1)!$ becomes very small when $n \ll \ell$. Hence the set of order index series of $n$-admissible sequences is a very small subset of the set of permutations of $(0, 1, ..., \ell)$ when $n \ll \ell$. There is a bijection between the set of order index series and the set of $n$-admissible sequences. This suggests that if a sequence of divergence angles is chosen randomly it has is a very small probability to be $n$-admissible. To further develop this idea we introduce a theorem.

**Lemma 3.1.** $p_{\ell+1} = (\ell+1)! - t_\ell^{\ell+1}$.

*Proof.* Note that $L_{\ell+1}^{\ell+1}$ is the set of all permutations of $\{0, ..., \ell\}$ see Property 3.1. One might notice that $I_{\ell+1} = L_{\ell+1}^{\ell+1} \backslash L_\ell^{\ell+1}$. Hence $|I_{\ell+1}| = |L_{\ell+1}^{\ell+1}| - |L_\ell^{\ell+1}|$. Therefore

$$p_{\ell+1} = (\ell+1)! - t_\ell^{\ell+1}.$$

□

**Theorem 3.1.** $\lim_{\ell\to\infty} t_n^{\ell+1}/(\ell+1)! \to 0$, $\forall n \leq \ell$.

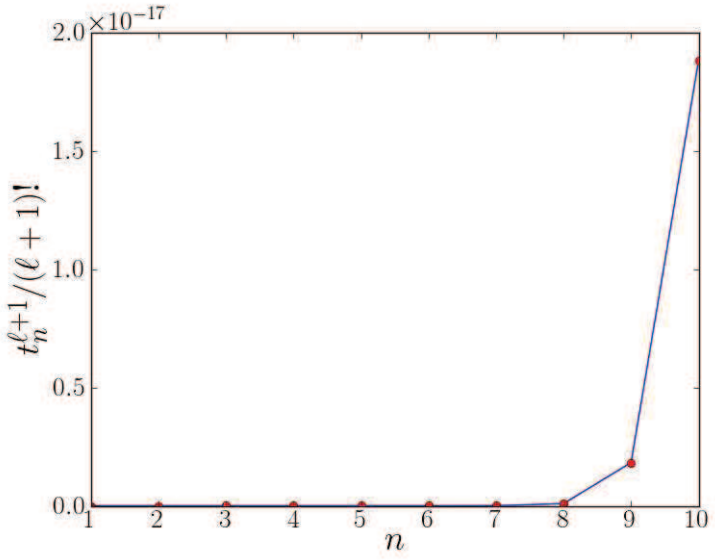



Figure 3.8: $t_n^{\ell+1}/(\ell+1)!$ for $\ell = 34$, which is the average length of measured sequences of divergence angles of our data set. The ratio becomes very small when $n \ll \ell$.

*Proof.* From Lemma 3.1 we deduce $\lim_{\ell\to\infty} \frac{p_{\ell+1}}{(\ell+1)!} = 1 - \lim_{\ell\to\infty} \frac{t_\ell^{\ell+1}}{(\ell+1)!}$. From Equation 3.3 we thus have $\lim_{\ell\to\infty} \frac{t_\ell^{(\ell+1)!}}{(\ell+1)!}$. Further more $t_n^{\ell+1} \leq t_\ell^{\ell+1}$, $1 \leq n \leq \ell$ since $L_n^{\ell+1} \subseteq L_\ell^{\ell+1}$. Therefore $\lim_{\ell\to\infty} t_n^{\ell+1}/(\ell+1)! \to 0$, $\forall n \leq \ell$.

□

## 3.5 Detecting n-admissibility in noisy sequences

In previous sections, $n$-admissibility is introduced and its different aspects are treated. In this section we develop an algorithm to recognise $n$-admissible sequences. One step further, we propose an algorithm to detect $n$-admissibility in measured sequences of divergence angles. These issues are first presented as problems then the algorithms are proposed.

### 3.5.1 Problems

**Problem 3.1.** *Given a sequence $S$ of theoretical divergence angles, the task is to determine whether $S$ is $n$-admissible.*

The following is a straightforward observation which is a special case of lemma 10.3 in [Parida, 2008]. It will be used to recognise $n$-admissible sequences.

**Lemma 3.2.** *Let $U$ be a permutation of $\{0, ..., \ell\}$. Then for all $0 \leq i < j \leq \ell$, $\{u_j, ..., u_k\} = \{j, ..., k\}$ iff $\min\{u_j, ..., u_k\} = j$ and $\max\{u_j, ..., u_k\} = k$.*

Note that if $U$ is order index series corresponding to an $n$-admissible sequence, the block $\{u_j, ..., u_k\} = \{j, ..., k\}$ is in fact a shuffled block of length $(k - j + 1)$.

**Example 3.5.** *Let $U = (0, 1, 3, 2, 4, 6, 7, 5, 8)$, according to Lemma 3.2:*

- $\{u_2, u_3\} = \{2, 3\} \iff (\min\{u_2, u_3\} = 2, \max\{u_2, u_3\} = 3)$
- $\{u_5, u_6, u_7\} = \{5, 6, 7\} \iff (\min\{u_5, u_6, u_7\} = 5, \max\{u_5, u_6, u_7\} = 7)$

Problem 3.1 can be seen as a special case of intervals problem [Parida, 2008]:

**Intervals problem.** *Given a sequence $U$ on integers $0, 1, 2, ..., \ell$, $I = [i, j]$ is an interval if for some $l$*

$$\{U_i, ..., U_j\} = \{l, l + 1, l + 2, ..., l + (j - i)\}, i < j.$$

*The task is to find all intervals $I$ on $U$.*

**Example 3.6.** *Given $U = (1, 0, 4, 3, 2)$, the four intervals in $U$ are:*

- $I_0 = [0, 1]$ *which can be underlined on $U$ as* $\underline{1, 0}, 4, 3, 2$.
- $I_1 = [2, 3]$ *which can be underlined on $U$ as* $1, 0, \underline{4, 3}, 2$.
- $I_2 = [2, 4]$ *which can be underlined on $U$ as* $1, 0, \underline{4, 3, 2}$.
- $I_3 = [0, 4]$ *which can be underlined on $U$ as* $\underline{1, 0, 4, 3, 2}$.

There are interval extraction algorithms [Parida, 2008] that can extract intervals for a given sequence $U$. However, the Problem 3.1 is much more specific than intervals problem. The existing interval extraction algorithms return intervals that are not necessarily indecomposable shuffled blocks. Therefore, existing interval extraction algorithms would return invalid sub-sequences for our problem (for example $I_1$ and $I_3$), whence the need for a more specific algorithm.

The Problem 3.1 deals with the sequences of divergence angles whose angles are in $D_n$ (see Property 3.2). A step forward would be recognising $n$-admissibility in measured sequences of divergence angles whose angles are not perfect multiples of $\alpha$ (more precisely do not belong to $D_n$), but are affected with noise. Let $x_i$ be a variable representing the $i$th angle in a sequence $(x_1, ..., x_\ell)$ of measured angles, and $\Gamma$ be a mapping that for each measured angle proposes a set of candidate theoretical angles among those in $D_n$:

$$\begin{array}{rccc} \Gamma: & [0, 360^\circ) & \longrightarrow & 2^{D_n} \\ & x_i & \longmapsto & C_i \subset D_n \end{array} \tag{3.5}$$

Note that several theoretical angle can be assigned to a measured angle (i.e. $|C_i| \geq 1$). In consequence, given the $\Gamma$ function, a set of measured angles will generate a possibly high number of candidate sequences.

**Problem 3.2.** *Let* $x = (x_1, ..., x_\ell) \in [0, 360°)^\ell$ *be a sequence of measured angles, and* $C = \prod_{i=1}^{\ell} C_i \subset D_n^\ell$ *where* $C_i = \Gamma(x_i)$ *and* $\prod_{i=1}^{\ell} C_i$ *is the Cartesian product of* $C_i$ $1 \leq i \leq \ell$. *The task is to find all n-admissible* $S = (\mu_1, \mu_2, ..., \mu_\ell)$ *in* $C$.

In the following, the we present how the function $\Gamma$ is defined for our data set of measured angles.

### 3.5.2 Assignment of theoretical angles to measured angles

In the previous subsection we introduced a general form of $\Gamma$ function as a mapping that for each measured angle proposes candidate theoretical angles. A simple way to assign candidate theoretical angles to a measured angle, is to define intervals around theoretical angles. Then, a measured angle can correspond to any theoretical angle whose interval contains it. In practice, we have chosen a common width for all these intervals, defined in terms of the von Mises distributions estimated as hidden distribution of angles in the statistical model, see Section 3.3. This common width was computed to ensure that the cumulative probability of the distribution, when restricted to the interval, had a fixed value $\rho$ (chosen close to 1 in practice).

Given a probability $\rho$, mathematical properties of the von Mises distribution allow us to compute the adequate interval width as a multiple of the common von Mises standard deviation $\nu$. Formally, the mapping $\Gamma$ is defined as follows,

$$\Gamma(x_i) = \{\mu \in D_n \mid \mu - k\nu \leq x_i \leq \mu + k\nu\}, \tag{3.6}$$

where $k$ is computed so that the cumulative probability $\rho$ has the desired value [Abramowitz *et al.*, 1970], see Figure 3.10.

In practice the von Mises distribution is defined in terms of a concentration parameter $\kappa$, from which the circular standard deviation $\nu$ can be computed [Mardia *et al.*, 2000]:

$$\nu = \frac{180\sqrt{-2\log\left(I_1(\kappa)/I_0(\kappa)\right)}}{\pi},$$

where $I_0(\kappa)$ and $I_1(\kappa)$ are the modified zero order Bessel function of the first kind and the modified first order Bessel function of the first kind respectively [Abramowitz *et al.*, 1970]. The common concentration parameter $\kappa$ estimated by the statistical model corresponds to a circular standard deviation $\nu = 18.5°$.

Using this statistical construction for the mapping $\Gamma$, we can also calculate a confidence level for each assignment of a theoretical angle to a measured angle. Let us denote $\omega(x_i, \mu)$ this confidence level. We define it as the probability of the corresponding assignment in the mixture of

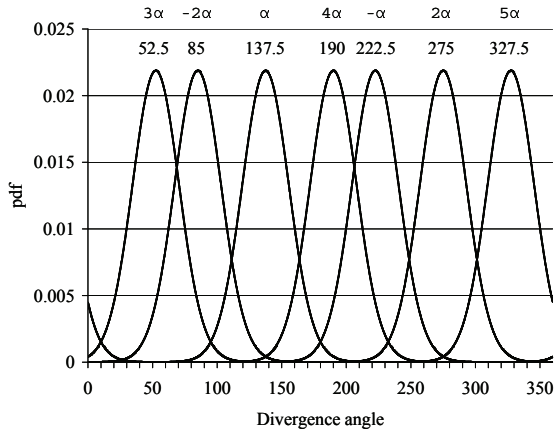



Figure 3.9: The estimated von Mises distributions used for the $\Gamma$ mapping.

von Mises distributions, based on the hypothesis that the theoretical angles are *a priori* equally probable (uniform distribution assumption), i.e. they are all given an identical weight. Let

$$g\left(x_i\,;\mu,\kappa\right) = \frac{1}{360 I_0(\kappa)} \exp\left(\kappa \cos(x_i - \mu)\frac{\pi}{180}\right)$$

denote the probability density function of the von Mises distribution with parameters $\mu$ (mean direction) and $\kappa$ (concentration parameter) see Figure 3.9. Then, in general, the probability density function of a mixture of von Mises distributions (in degrees) with common concentration parameter $\kappa$ is given by:

$$f(x_i) = \sum_{\mu \in D_n} \eta_\mu \, g(x_i; \mu, \kappa),$$

where the weights $\eta_\mu = P(x_i = \mu)$ are such that $\sum_{\mu \in D_n} \eta_\mu = 1$. In our case, the assumption of uniform weights leads to $\eta_\mu = 1/|D_n| = 1/(3n-2)$ for all $\mu \in D_n$, $D_n$ being defined in Property 3.2

This uniform distribution assumption is the most neutral assumption. In fact, the frequencies of occurrence of the possible angles are very different in the our combinatorial model and this implicitly weights the angles in the overall combinatorial model. Hence, it would be irrelevant to include explicit weights $\eta_\mu$ with contrasted values reflecting the frequencies.

In summary, for each possible theoretical $\mu \in D_n$, the posterior probability of assigning $\mu$ to $x_i$ is:

$$\omega(x_i, \mu) = \frac{g\left(x_i; \mu, \kappa\right)}{\sum_{\gamma \in D_n} g\left(x_i; \gamma, \kappa\right)}. \tag{3.7}$$

$\omega(x_i, \mu)$ can be calculated and compared with a predefined threshold (typical value 0.05) to decide whether $\mu$ should be kept or rejected as a candidate angle.

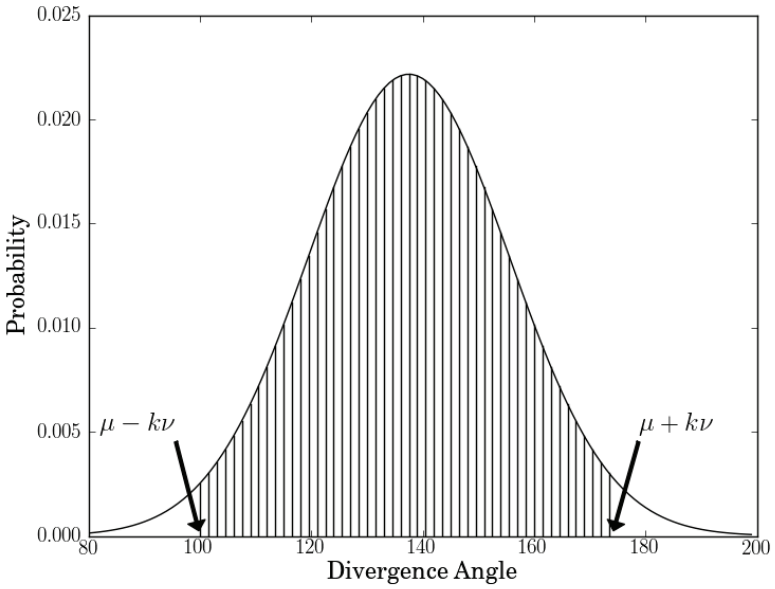



Figure 3.10: Interval around theoretical angles, $\mu = 137.5°$. The interval is used to define $\Gamma$ function.

### 3.5.3 Algorithms

First we propose an algorithm to solve Problem 3.1.

**`n-admissible algorithm`: detecting $n$-admissible sequences of theoretical divergence angles**

The algorithm that we propose uses Parikh mapping [Parida, 2008] to test $\{u_0, ..., u_\ell\} = \{0, 1, ..., \ell\}$. Parikh mapping maps each word to a vector of number of occurrences of the letters in the word. Let $\Sigma$ be an alphabet of size $m \in \mathbb{N}$,

$$\Sigma = \{\sigma_1, \sigma_2, \cdots, \sigma_m\}.$$

We suppose moreover that the symbols are ordered: $\sigma_1 < \sigma_2 < \cdots < \sigma_m$. For each word $w \in \Sigma^*$ let $\psi_i$ be the number of occurrences of $\sigma_i$ in $w$. Parikh mapping is a function

$$\Psi : \Sigma^* \to \mathbb{N}^m$$

$$\Psi(w) = (\psi_1, \psi_2, ..., \psi_m).$$

**Example 3.7.** *Let $\Sigma = \{a, b, c, d\}$ with the alphabetical order, and $w = abad$. Then $\Psi(w) = (2, 1, 0, 1)$.*

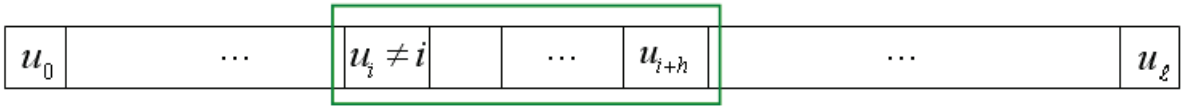



Figure 3.11: Graphical representation of the `n-admissible` algorithm. The sequence $U$ is scanned form left to right. When $u_i \neq i$, the conditions of Lemma 3.2, $\min\{u_i, ..., u_{i+h}\} = i$, $\max\{u_i, ..., u_{i+h}\} = i + h$, $1 \leq h \leq n$, is checked.

The algorithm uses an array $\Psi = (\psi_0, ..., \psi_\ell)$ where $\psi_i$, $0 \leq i \leq \ell$, is the number of occurrence of $i$ in $U = \{u_0, ..., u_\ell\}$. $U = \{0, 1, ..., \ell\}$ iff $\Psi = (\psi_0, ..., \psi_\ell) = (1, 1, ..., 1)$. A single scan of $U$ from left to right allows us to compute $\Psi$ in linear time.

Once using Parikh mapping $\{u_0, ..., u_\ell\} = \{0, 1, ..., \ell\}$ is verified, then the sequence $U$ is scanned from left to right. Wherever $u_i \neq i$, the set of at most $n$ succeeding elements including $u_i$, i.e. $\{u_i, ..., u_{i+h}\}$, $1 \leq h \leq n$, is checked according to Lemma 3.2 (i.e. $\min\{u_i, ..., u_{i+h}\} = i$, $\max\{u_i, ..., u_{i+h}\} = i + h$). If the conditions are not satisfied, the sequence is not $n$-admissible. This procedure scans the sequence from left to right once, therefore the time complexity of the algorithm is $O(\ell)$ where $\ell$ is length of the input sequence.

---

```
n-admissible algorithm:
input:  n, S # S: a sequence of length ℓ a priori composed of theoretical divergence angles
output:  Boolean (true or false)
Begin
if S_0 not in D_n:
   return False # since (μ_{i+1}+(u_i − i)α) in P|_{u_i} in Proposition 3.4 could be not in D_n
Construct the order index series U(S)
if {u_0, ..., u_ℓ} ≠ {0, 1, ..., ℓ} :    return False % using Parikh mapping
i:=0
while i ≤ ℓ :
   if u_i ≠ i :
      lo:=u_i
      up:=u_i
      j:=i+1
      while true:
          if j > ℓ :
               return false
```

```
          if j - i > n - 1:
               return false
          lo:=min(u_j, lo)
          up:=max(u_j, up)
          if (lo = i) & (up = j):
               i:=j+1
               break % (u_i, ..., u_j) shuffled block
          j:=j+1
   else:  i:=i+1
      i:=i+1
return true # the order index series U(S) can also be returned if needed
End
```

---

Now we propose an algorithm to solve Problem 3.2.

**`n-admissible tree algorithm`: detecting $n$-admissibility in measured sequences**

Now we can deal with Problem 3.2. A naive algorithm would construct all candidate sequences $S \in \prod_{i=1}^{\ell} C_i$, and then apply the `n-admissible algorithm`. The number of candidate sequences is $\prod_{i=1}^{\ell} |C_i|$, where the $|C_i|$ is the cardinality of $C_i$. Since $C_i$s are typically not singletons, $|C|$ increases exponentially with $\ell$. We know that $n$-admissible sequences can be decomposed into shuffled blocks. We take advantage of this property and we propose a branch and bound algorithm to solve Problem 3.2.

Let $x = (x_1, ..., x_\ell)$ be the input sequence of measured angles. The algorithm that we propose maintains a rooted tree $T$ initialised by *root* of depth 0. For each *leaf* of depth $d$ in $T$, if $d < \ell$, a window consisting of at most ($n$-1) angles, $(x_{d+1}, ..., x_{d+h})$, $1 \leq h < n$ is selected from the input sequence of measured angles. The window could potentially correspond to a shuffled block of theoretical angles. The `n-admissible algorithm` is applied to $P|_{u(leaf)}$ where $P \in \Gamma(x_{d+1}) \times ... \times \Gamma(x_{d+h})$. $P|_{u(leaf)}$ is computed according to Proposition 3.4. If $P|_{u(leaf)}$ turns out to be $n$-admissible then the algorithm adds $P$ to *leaf*. If $\nexists P \in \Gamma(x_{d+1}) \times ... \times \Gamma(x_{d+h})$, $1 \leq h < n$, such that $P|_{u(leaf)}$ is $n$-admissible, the algorithm returns ($T$, `False`). This means that no theoretical angle can be affected to $x_{d+1}$ and the algorithm fails at the angle $x_{d+1}$ . We call the angle $x_{d+1}$ a *non-explained* angle. The algorithm returns the rooted tree ($T$, `True`) when the depth of all leaves of $T$ is equal to the length of input sequence $\ell$. See Figure 3.12 for a graphical representation of the algorithm.

When the algorithm fails at angle $x_{d+1}$, we can use the `n-admissible tree algorithm` to analyse subsequence the $(x_{d+2}, ..., x_\ell)$.

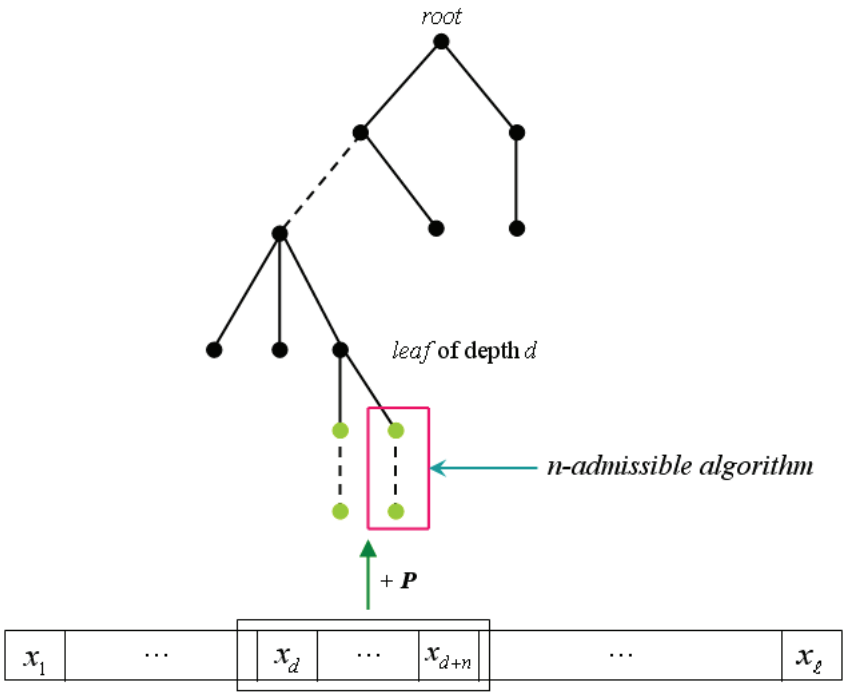



Figure 3.12: Graphical representation of `n-admissible tree algorithm`. For each *leaf* of depth $d$ in tree, a window consisting of $(x_{d+1}, ..., x_{d+h})$, $1 \leq h < n$, is selected from input sequence of measured angles. $P \in \Gamma(x_{d+1}) \times ... \times \Gamma(x_{d+h})$ is added to *leaf* if $P|_{u(leaf)}$ is n-admissible. Dotted lines indicate paths with possibly several nodes.

Using the Proposition 3.4, $n$-admissibility is tested on sub-sequences only. The time complexity of the $n$-admissible tree algorithm increases exponentially with $n$ since the $n$-admissibility is tested on all candidate sequences of the window including at most $n-1$ angles. When the returned tree contains only one $n$-admissible sequence (as it is often the case in practice), it is more precisely $O(l \times k^n)$ where $k = max(|C_i|)$.

---

`n-admissible tree algorithm:`

1. `input:` $\Gamma$, $n$, $(x_1, ..., x_\ell)$ # *a sequence of measured divergence angles of length* $\ell$
2. `output: n-admissible tree`
3. # *with nodes labelled by both divergence angles* $\mu_i$ *and order indices* $u_i$
4. `Begin`
5. $T$`:=` $\{root\}$
6. $\mu(root)$`:=` 0
7. $u(root)$`:=` 0

8. while True:
9. nLeaves:= $nonterminal\_leaves(T)$ # *leaves of depth* $< \ell$
10. if nLeaves is empty: return ($T$, True)
11. for $leaf$ in nLeaves:
12. $d$:=depth($leaf$)
13. $m$:=min($n$, $\ell - d$)
14. $branches$:=an empty set
15. $added$:=False
16. for $k \in \{1, ..., m\}$:
17. for $P \in \Gamma(x_{d+1}) \times ... \times \Gamma(x_{d+h})$:
18. if $d$ < $n$:
19. $P2Test$ := labels of the path from root to $leaf$
20. $P2Test$ := $P \cdot P2Test$
21. else:
22. $P2Test$ := $P|_{u(leaf)}$ % *cf. 3.4, Proposition 3.4*
23. if n-admissible($P2Test$): % *also returns* $\mu$ *and* $u$ *values of nodes in* $P$
24. $added$ = True
25. add $P$ to $branches$
26. if not $added$:
27. return ($T$, False):
28. for $P$ in $branches$:
29. if $\nexists S \in branches$ such that $P$ is a subsequence of $S$:
30. append all nodes on $P$ to $leaf$
31. End

---

Since the length of $(x_d, \cdots, x_{d+h})$ varies $(1 \leq h < n)$ several $P \in \Gamma(x_{d+1}) \times ... \times \Gamma(x_{d+h})$ might be added to $T$ that are in fact sub-sequences of one another. This generates identical paths in $T$. The lines 28, 29 and 30 of the pseudo-code are to avoid such paths.

We implemented a version of this algorithm without taking advantage of the Proposition 3.4 (In the pseudo-code of the n-admissible algorithm the lines 18, 21 and 22 were omitted). In consequence, the n-admissible algorithm should be applied to the path from root to a leaf, see Figure 3.13. In consequence, the complexity of this algorithm becomes $O(l^2 \times k^n)$ where $k = max(|C_i|)$ (assuming that the returned tree contains only one $n$-admissible sequence). Therefore, the running time became more important for long sequences, see Figure 3.14. As one might see using the Proposition 3.4 the time is linear in regards to the length of sequence which

is not the case if this proposition is omitted.

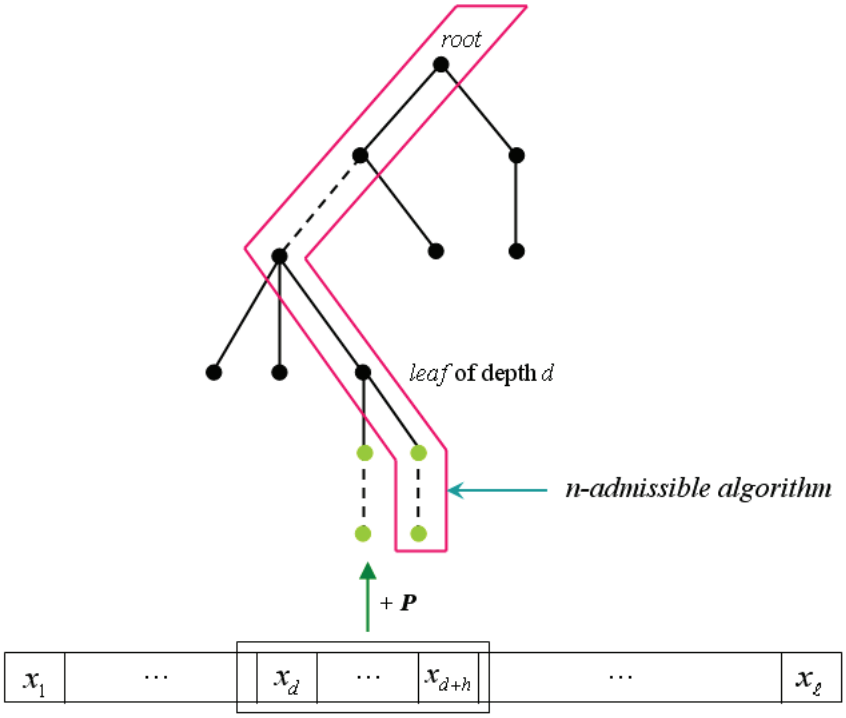



Figure 3.13: A simpler version of the `n-admissible tree algorithm`. The $n$-admissibility is tested on the path up to root.

**Proposition 3.8.** *Let $C = \Gamma(x_1) \times ... \times \Gamma(x_\ell)$ and $T$ be the tree returned by the* `n-admissible tree algorithm`*. Let us call $\mathcal{A}_n(C)$ the set of $n$-admissible sequences in $C$, and $\pi(T)$ the set of all (labels of) paths in $T$, from the root to the leaves. Then $\pi(T) = \mathcal{A}_n(C)$, i.e. the tree built in the* `n-admissible tree` *algorithm contains exactly the $n$-admissible sequences in $C$.*

*Proof.* The inclusion $\pi(T) \subset \mathcal{A}_n(C)$ is clear, since in the 2nd `for` loop only paths which are $n$-admissible can be added, thanks to Proposition 3.4.
To see that the converse holds, it suffices to remark that given an $n$-admissible sequence $(\mu_1, ..., \mu_\ell)$, one of the $n$ sub-sequences

$$(\mu_1, ..., \mu_{\ell-1}), \quad (\mu_1, ..., \mu_{\ell-2}) \quad \cdots \quad (\mu_1, ..., \mu_{\ell-n})$$

must be $n$-admissible as well, as follows from the definition. Because all these sub-sequences are tested in the `for` loop of the line 15, there cannot be an $n$-admissible sequence that is not detected by the algorithm, and thus $\pi(T) \supset \mathcal{A}_n(C)$. □

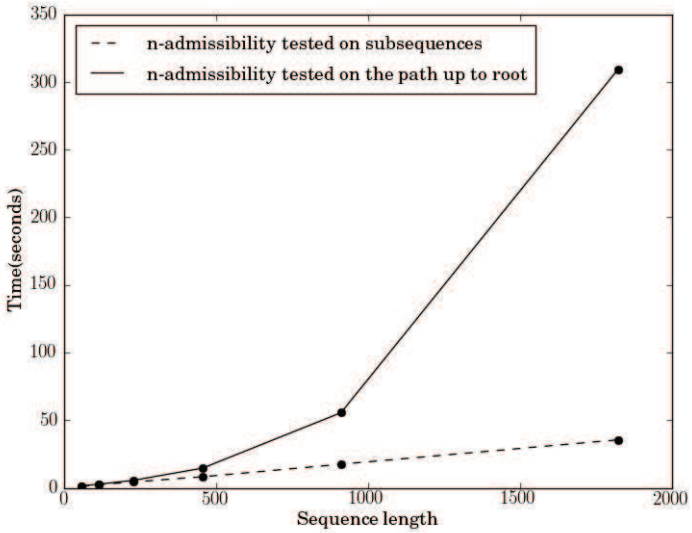



Figure 3.14: Comparison of time complexity of testing n-admissibility on sub-sequences and on the path up to root. $n = 3$.

**Further Pruning**

Let $T$ be an $n$-admissible tree returned by the `n-admissible tree` algorithm for a measured sequence $x = (x_1, ..., x_\ell)$. In the case where $T$ has several leaves, one may use the weights $\omega$ (see Equation (3.7)), to sort the paths from root to leaves. Actually, to a given path with labels $(\mu_1, ..., \mu_\ell)$ in the computed $n$-admissible tree, one may naturally assign the weight $\prod_{i=1}^{\ell} \omega(x_i, \mu_i)$. Then, all paths in the tree can be ordered according to their weights.

These weights can also be used to limit the size of the constructed tree, by ruling out all candidate paths whose weight is below a certain threshold. The algorithm can prune not only the non admissible paths, but also those having a weight below a threshold. This is how we have actually implemented the algorithm, using posterior probabilities for weights, and adding a threshold as an input to the algorithm.

Sorting the paths from root to leaves allow us to identify the most probable path. This path in an $n$-admissible tree corresponds to the most probable $n$-admissible sequence for a sequence of measured angles.

## 3.6 Results

In the previous sections, we developed a combinatorial model to study the permutation hypothesis put forward by exploratory analysis and statistical modelling, see Sections 3.2 and 3.3. In the following, we use the combinatorial model to analyse our data set as explained in section 3.1 and

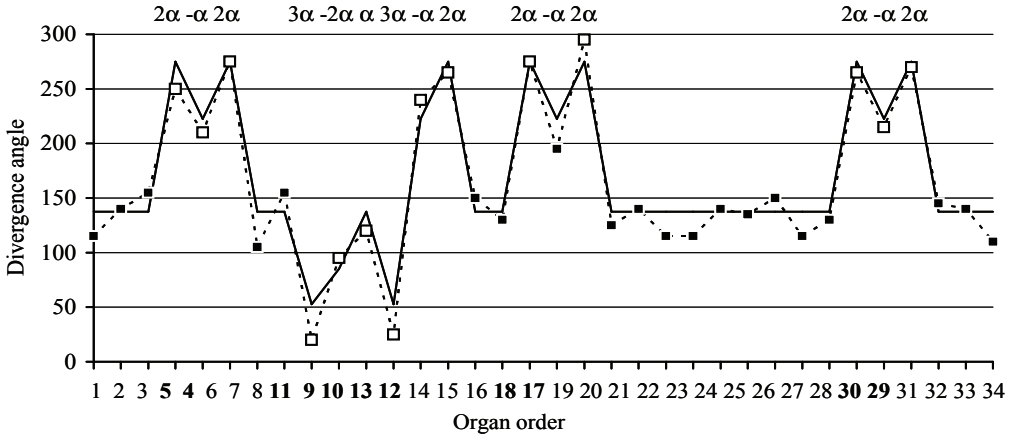



Figure 3.15: *ahp6* individual: prediction of the divergence angle sequence (continuous line) and labelling of the divergence angles within the permuted segments.

then we discuss the application of the algorithms and the propositions to interpret the data set.

### 3.6.1 Analysis of our dataset

We applied the `n-admissible tree` algorithm to analyse the data set. Nearly all segments of non-canonical angles, precisely more than 98% of measured angles for wild-type and and 95% of measured angles for the mutant, were explained by permutations in order of appearance of 2 or 3 consecutive organs. In other words, most of the perturbed sequences were identified as 3-admissible sequences, Figure 3.15 gives an example where the predicted divergence angle sequence is in close agreement with the measured divergence angle sequence.

For some sequences however, there was no $n$-admissible sequence in $C$ as defined in Problem 3.2. This was often due to either some error in the measurement of divergence angles, or to too large deviations between an observed angle and any predicted angle corresponding to a valid prediction. Let $x = (x_1, ..., x_\ell)$ be a sequence of measured angles. We analyse the sequence $x$ using `n-admissible tree algorithm`. Remind that we call angle $x_i$, $1 \leq i \leq \ell$, non-explained if the algorithm fails at the angles $x_i$. This means that there is no theoretical angle in $\Gamma(x_i)$ (the general form of $\Gamma(x_i)$ is defined in Equation 3.5) that can lead to an $n$-admissible sequence. More precisely:

**Definition 3.6.** *Let $x = (x_1, ..., x_\ell)$ be a sequence of measured divergence angles, $x_i$ is called non-explained iff $\exists S \in \prod_{k=1}^{i-1} \Gamma(x_k)$ such that $S$ is n-admissible and $\nexists S \in \prod_{k=1}^{i} \Gamma(x_k)$ such that $S$ is n-admissible.*

Due to the dependencies induced by the permutation patterns a non-explained angle, may in

fact result from an isolated error, earlier in the sequence. An isolated error in our data is an angle close to 275°, which is between sub-sequences of angles close to 137.5°. This can be due to a forgotten measured angle close to 137.5°. `n-admissible tree algorithm` tries to interpret 275° together with succeeding angles as a permutation pattern, this leads to a non-explained angle in succeeding angles. Since this non-explained angle and the permutation pattern is due to the isolated error, we consider the permutation pattern as invalid. Hence, permutations pattern (i.e. shuffled blocks) preceding a non-explained angle should be marked as not valid. To achieve this goal, we define splitting points. The notion of splitting point can be viewed as a deterministic analogue of a regeneration point for a stochastic process. A regeneration point is a time instant at which the future of the process depends only of its state at that instant and is thus independent of its past before that instant. The process is thus reborn at a regeneration point.

**Definition 3.7.** *Let $S$ be a sequence of divergence angles and $U(S)$ be the order index series of $S$, $u_i$ is a splitting point iff $u_i = i$ and $u_i$ is not in a shuffled block.*

Using the notion of splitting point, we implemented a procedure which was applied after the `n-admissible tree` algorithm. It consisted of a backtracking starting at the non-explained angle and progressing towards a splitting point. The permutation pattern(s) between the non-explained angle and the leftmost splitting point are invalidated. Note that there can be more than one permutation pattern between non-explained angle and the leftmost splitting point, the procedure invalidated all of them.

To refine this analysis, we used reversibility (Proposition 3.1), and applied the whole procedure to both measured sequences and their reverse. Then, the intersection of angles invalidated on a sequence and its reverse, was often reduced to a single angle. These angles were likely due to measurement errors, typically the omission of one angle in the series, leading to an isolated $2\alpha$ in a sequence of canonical angles $\alpha$, see Figure 3.16.

These steps can be summarized as follows, let $x = (x_1, ..., x_\ell)$ be a measured sequence of divergence angle:

1. The sequence $x$ is analysed using $n$-`admissible tree algorithm`. Let $S_1$ be the sequence of theoretical divergence angles returned by the algorithm. In case of multiple candidate sequences, the most probable one is selected using associated posterior probabilities. Using the procedure described previously, we invalidate blocks of the sequence $S_1$. We denote the set of invalidated angles by $B_1$. Note that $B_1$ might be empty.

2. The sequence $x' = (x_\ell, ..., x_1)$ is analysed by $n$-`admissible tree algorithm`. We denote $S_2$ the reverse sequence of theoretical angles returned by the algorithm. The set of invalidated angles is denoted by $B_2$.

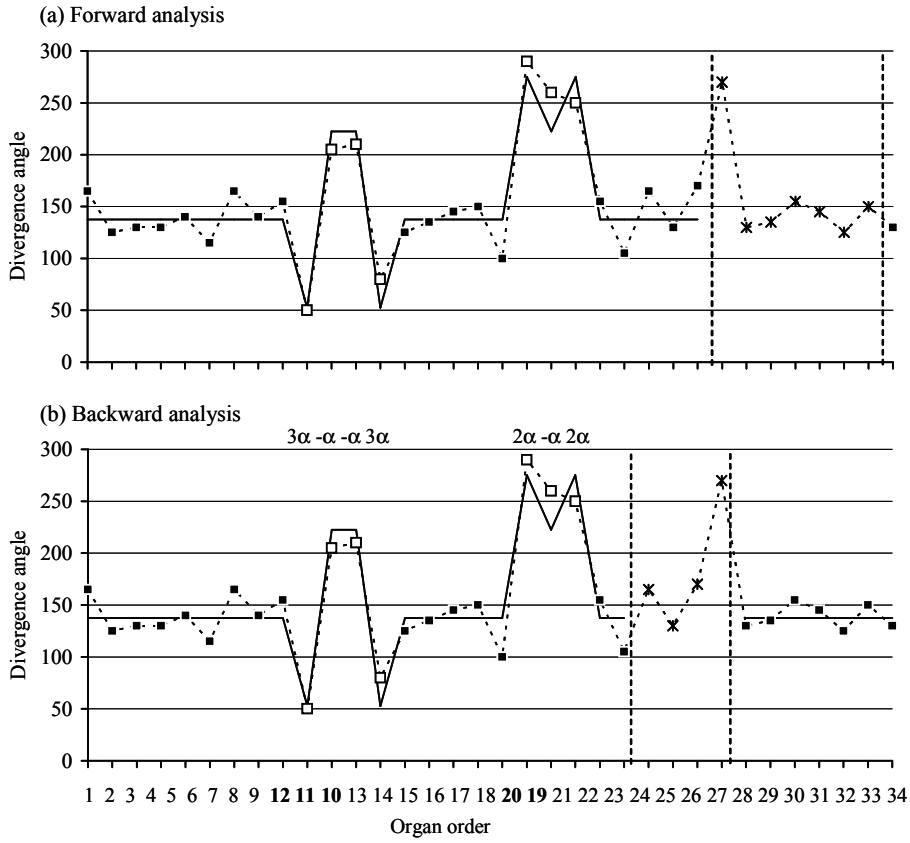



Figure 3.16: Analysis of the sequence in both directions to detect segments that are invalid with respect to the permutation assumption. The invalid sub-sequences are delimited by dashed lines. The continuous line corresponds to the predicted divergence angles. Backward analysis corresponds to output of `n-admissible tree algorithm` on reverse sequence of measured angles. Note that the intersection of invalidated sub-sequences is the single angle 27, an isolated angle close to 275°. This can be due to an omission of measurement of an angle close to 137.5°

3. By removing the corresponding angles belonging to $B_1 \cap B_2$ from $x$, we analyse the sub-sequences using $n$-`admissible algorithm`.

4. An expert analysis of the concatenated sub-sequences determines the interpretation for the sequence $x$.

The proposed modelling approach allows to explain a very large proportion of the non-canonical divergence angles despite the relatively high standard deviation (approx. 18.5°) of the estimated von Mises distributions. We compared the predictions of this approach with those obtained using the variable-order hidden Markov chain, see Section 3.3. When the predictions did not coincide, an expert examination of the two predictions was needed to choose the most likely one. The final results were thus a consensus deduced from divergence angles sequence optimally labelled by the hidden variable-order Markov chains and the combinatorial model. This is illustrated in Table 3.2 by counts of matches and mismatches between the model and final consensus.

This indicates that the proposed model correctly describes the phyllotactic patterns of *Arabidopsis thaliana*. Wild-type plants were characterized by relatively frequent occurrences of 2-permutations generally isolated while mutants were characterized by the frequent occurrences of both 2- and 3-permutations whose succession generates highly complex motifs, see Figure 3.15 for an example. A summary of the results is shown in Table 3.1; the proportion of non-canonical angles increases from 15% to 37% between the wild type and the mutant. This corresponds to a strong increase in the occurrence of 2-permutations and a burst in 3-permutations. The term Lucas phyllotaxis refers to a spiral phyllotaxis were the canonical divergence angle $\alpha$ is 99.5° (see Chapter 1). Although rarer than the Fibonacci spiral ($\alpha = 137.5°$), it is known to occur in nature, and was able to explain two wild-type and two *ahp6* sequences, for which 137.5° failed.

The permuted segments detected in the measured divergence angles are listed in Table 3.3 and Table 3.3. Tables 3.5 and 3.6 shows the counting of divergence angles and histogram of the proportions of non-canonical angles per sequence respectively.

### Generative spiral orientation

39 out of 82 wild-type individuals and 45 out of 89 mutant individuals had a clockwise orientation. Hence, clockwise and counterclockwise spirals were in roughly equal proportion in both wild-type and mutant samples. Several reports in literature indicate an equal number of clockwise and counterclockwise spirals [Beal, 1873, Allard *et al.*, 1951]. The equal number of clockwise and counterclockwise spirals is explained by the fact the spiral originates from an initial symmetry breaking event corresponding the transition between the decussate phyllotaxis of

| | Wild-type plant | *ahp6* |
|---|---|---|
| # sequences/# organs | 82/2405 | 89/ 2815 |
| % of non-canonical angles | 15% | 37% |
| % of explained angles | 98% | 95% |
| # individuals, Lucas phyllotaxis | 2 | 2 |
| # individuals without permutations | 18 | 1 |
| # 2-permutations | 123 | 297 |
| # 3-permutations | 3 | 53 |

Table 3.1: Results of application of the model to the data set.

| | match | mismatch | consensus |
|---|---|---|---|
| $\alpha$ | 3579 | 62 | 3641 |
| $2\alpha$ | 675 | 25 | 700 |
| $3\alpha$ | 120 | 5 | 125 |
| $4\alpha$ | 10 | 0 | 10 |
| $5\alpha$ | 1 | 0 | 1 |
| $-\alpha$ | 433 | 13 | 446 |
| $-2\alpha$ | 33 | 0 | 33 |
| ? | 99 | 50 | 149 |
| | 4950 | 155 | 5105 |

Table 3.2: Counts of match and mismatch between the automatic analysis and the final consensus. The expert analysis let us to fix the final consensus. ? stands for non-explained and invalidated angles. The four Lucas phyllotaxis individuals are excluded from sample

| | organ order | divergence angles | wild type | mutant |
|---|---|---|---|---|
| 2-permutation | 2 1 3 | 2 -1 2 | 90 | 193 |
| | 3 2 1 4 | 3 -1 -1 3 | 1 | 11 |
| 3-permutation | 3 1 2 4 | 3 -2 1 2 | 1 | 9 |
| | 2 3 1 4 | 2 1 -2 3 | | 13 |
| total | | | 2 | 33 |
| 2 2-permutations | 2 1 4 3 5 | 2 -1 3 -1 2 | 16 | 32 |
| | 2 1 5 3 4 6 | 2 -1 4 -2 1 2 | 1 | 1 |
| | 2 3 1 5 4 6 | 2 1 -2 4 -1 2 | | 2 |
| 2-permutation and | 2 1 5 4 3 6 | 2 -1 4 -1 -1 3 | | |
| 3-permutation | 3 2 1 5 4 6 | 3 -1 -1 4 -1 2 | | 1 |
| | 2 1 4 5 3 6 | 2 -1 3 1 -2 3 | | 1 |
| | 3 1 2 5 4 6 | 3 -2 1 3 -1 2 | | 3 |
| total | | | 1 | 8 |

Table 3.3: Permuted segments up to length 6. These segments are delimited by two splitting points. By convention, the origin of the organ sequence is 0 (not indicated). Palindromes (a subsequence and its reverse in terms of divergence angles are the same) are given at the beginning of each category. A subsequence and its reverse are then given on two successive rows. Frequencies of sub-sequences are given or the wild type and the mutant.

| | organ order | divergence angles | mutant |
|---|---|---|---|
| 3 2-permutations | 2 1 4 3 6 5 7 | 2 -1 3 -1 3 -1 2 | 5 |
| | 3 2 1 6 5 4 7 | 3 -1 -1 5 -1 -1 3 | 1 |
| 2 3-permutations | 2 3 1 6 4 5 7 | 2 1 -2 5 -2 1 2 | 1 |
| | 3 2 1 6 4 5 7 | 3 -1 -1 5 -2 1 2 | 1 |
| | 2 1 5 3 4 7 6 8 | 2 -1 4 -2 1 3 -1 2 | 1 |
| 2 2-permutations and | 3 1 2 5 4 7 6 8 | 3 -2 1 3 -1 3 -1 2 | 1 |
| a 3-permutation | 2 1 4 3 7 5 6 8 | 2 -1 3 -1 4 -2 1 2 | 2 |
| | 2 3 1 5 4 7 6 8 | 2 1 -2 4 -1 3 -1 2 | 1 |
| 4 2-permutations | 2 1 4 3 6 5 8 7 9 | 2 -1 3 -1 3 -1 3 -1 2 | 1 |
| 3 2-permutations and a 3-permutation | 2 1 4 3 7 6 5 9 8 10 | 2 -1 3 -1 4 -1 -1 4 -1 2 | 1 |

Table 3.4: Permuted segments of lengths from 7 to 10 (only the observed permuted segments are given). These segments are delimited by two splitting points. By convention, the origin of the organ sequence is 0 (not indicated). Frequencies of segments are given (all these long permuted segments were detected in mutant individuals).

| | Wild-type | *ahp6* |
|---|---|---|
| $\alpha$ | 2005 | 1692 |
| $2\alpha$ | 214 | 499 |
| $-\alpha$ | 126 | 330 |
| $3\alpha$ | 20 | 111 |
| $-2\alpha$ | 2 | 37 |
| $4\alpha$ | 1 | 10 |
| $5\alpha$ | 0 | 3 |

Table 3.5: Counting of divergence angles

| Upper limit | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|
| Wild-type | 31 | 24 | 16 | 4 | 2 | 4 | | 1 | |
| *ahp6* mutant | 3 | 6 | 23 | 17 | 19 | 10 | 9 | 1 | 1 |

Table 3.6: Histograms of the proportions of non-canonical angles per sequence.

the seedling/*embryo* and the Fibonacci phyllotaxis of the adult plant. Hence the AHP6 gene does not seem to affect this initial symmetry breaking event.

### 3.6.2 Sensitivity analysis

In order to see how the variations of the parameters of the model effect the results, a sensitivity analysis was carried out. We varied $\rho$ (introduced in subsection 3.5.2). We did not use the posterior probability defined in Equation (3.7) to prune the tree in `n-admissible tree algorithm`. They were rather used to choose the most probable candidate sequence. The prediction of the automatic analysis is compared to the consensus and the proportion of the matches to number of measured angles is calculated, see Table 3.7. As we can see in the table, the model is robust to these variations.

| % \ $\rho$ | $\rho = 0.95$ | $\rho = 0.99$ | $\rho = 0.995$ | $\rho = 0.999$ | $\rho = 0.9995$ |
|---|---|---|---|---|---|
| % of explained explained angles | 84 % | 95 % | 96 % | 97 % | 97 % |

Table 3.7: To investigate the robustness of the model predictions, we varied the $\rho$. For each parameter value the proportion of the matches to number of measured angles is calculated. The four Lucas phyllotaxis individuals are excluded from the sample.

### 3.6.3 Estimating the canonical divergence angle of a measured sequence

The canonical divergence angle for the Fibonacci spiral phyllotaxis is supposed to be $\alpha = 137.5°$. However, in measured divergence angles we observed angles fluctuating around $\alpha$. In our mea-

surement protocol, stem of plants are not always centred perfectly in the protractor, this could be to some extent the origin of fluctuations. Another source of noise can be due to division of meristem surfaces in discrete cells as small shift of one cell during organ initiation can produce a deviation up to 15° in mature organs [Kuhlemeier, 2007].

In order to quantify this fluctuation we first computed the weighted mean absolute deviation between the average baseline segment level for an individual and the canonical angle. Let $x^j = (x_1^j, ..., x_\ell^j)$ be the sequence of measured angles of the individual $j$ and $S^j = (m_1^j\alpha, ..., m_\ell^j\alpha)$ be the corresponding sequence of divergence angles where $m_i^j \in D_n$, $1 \leq i \leq \ell$ (see Property 3.2). The sample weighted mean absolute deviations given by

$$\sum_j \omega_j |\bar{x}_j - 137.5| / \sum_j \omega_j,$$

where $\bar{x}_j$ is the mean of divergence angles within baseline segments and $\omega_j$ is the cumulative length of the baseline segments for the $j$th individual. We obtain a mean absolute deviation of 8.1° for the wild-type, 5.7° for the mutant and 7° for the pooled samples (mutant and wild-type). We then estimated canonical divergence angle for each sequence of measured divergence angles. Let $\tilde{\alpha}$ be the canonical divergence angle of $x^j$. The distance between $x^j$ and $S^j$ should be minimal, in other words:

$$\tilde{\alpha} = \arg\min_\beta \frac{1}{\ell} \sum_{i=1}^{\ell} |x_i^j - m_i^j\beta|,$$

where $\beta \in [0, 360)$. We estimated $\tilde{\alpha}$ for the wild type and *ahp6* mutant. The estimated canonical divergence angles for wild type (and *ahp6* mutant respectively) were $137.5 \pm 4$ ( and $137.5 \pm 6$ respectively).

## 3.7 Software implementation

The algorithms are implemented using the Python[3] progamming language. We have used Numpy[4] and Scipy[5] which are packages for scientific computing with Python to calculate the posterior probabilities. Matplotlib [6] is used for data visualisation. A graphical user interface is implemented using PyQT4 [7] package see Figure 3.17 for a snapshot of the graphical user interface. Using this graphical interface, one can set the parameters of the models and test different divergence angles and analyse the sequences see Figure 3.18.

[3] http://www.python.org/
[4] http://numpy.scipy.org/
[5] http://www.scipy.org/
[6] http://matplotlib.sourceforge.net/
[7] http://www.riverbankcomputing.co.uk/software/pyqt/intro

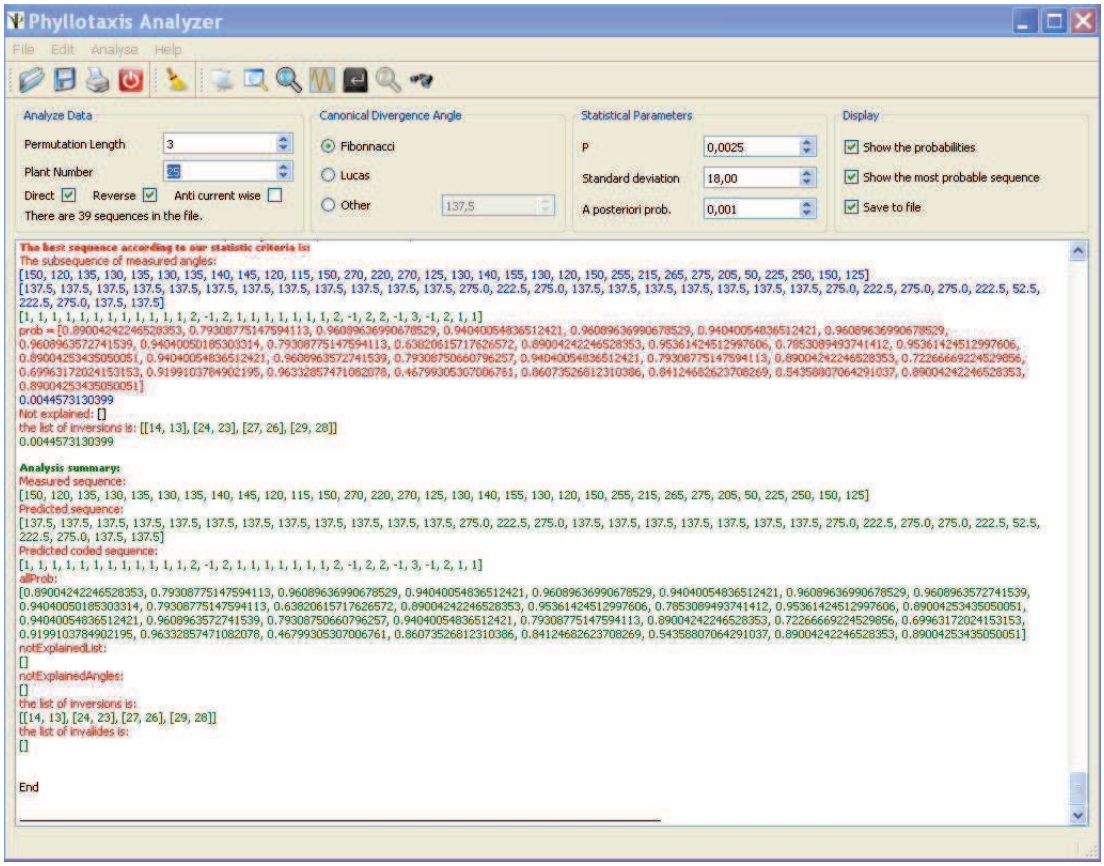



Figure 3.17: Snapshot of the graphical user interface.

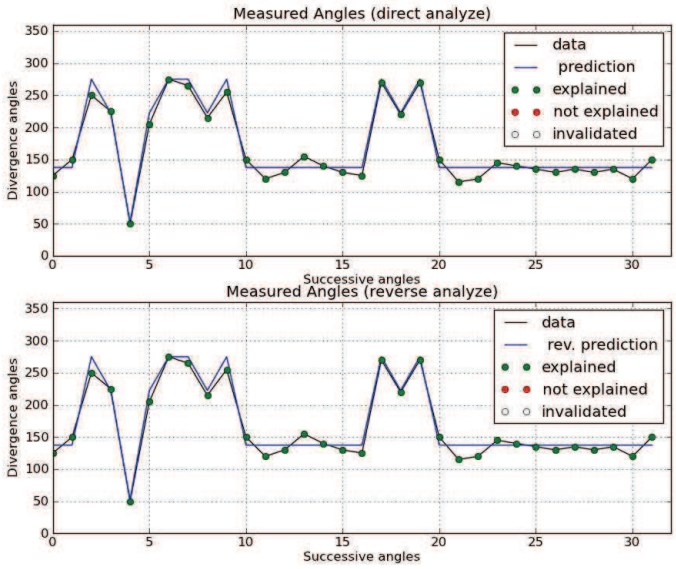



Figure 3.18: Snapshot of the graphical outputs of the implemented software.

The source code contains about 1400 lines of code and is available under Cecill-C licence in OpenAlea[8]. A tutorial is available online [openalea webpage] and was created using Sphinx[9].

## 3.8 Discussion

While nearly all perturbed sub-sequences are explained by the model, the quality of prediction of the model is difficult to assess quantitatively. As we mentioned before although each individual plant can be in general given a well defined canonical divergence angle, it varies from 137.5° with a standard deviation of 8°. This deviation from 137.5° should be taken into account in any effort to estimate prediction error. Therefore, a comparison prediction error between individuals will be biased by this individual effect.

This remark indicates that the usage of residue computation to estimate the accuracy of our predictions is not sufficient for our data set. However, several arguments give credit to these predictions.

- First, the statistical analysis that was performed to identify patterns was not relying on any permutation assumption. Being Markov based, it only relied on the natural hypothesis that divergence angles are dependent on their immediate predecessors in the spiral. It is then an *estimation* procedure that returned a transition structure like the one shown in Figure 3.5, which exactly encompasses the set of 2-admissible sequences (which are statistically more significant than 3-admissible sequences).

- A second argument, is the fact that the setting up of the final results, based on a consensus between several automated predictions, was performed via a scrutinous examination of all the measured sequences by two experts. Besides establishing the consensus, these experts paid attention to the agreement between the predictions and the data, correcting visually the individual effect mentioned above. Although less formal than the previous point, we believe that human judgement cannot be ignored when dealing with biological data, with all the complexity that this implies.

Altogether, the results strongly indicate that wild-type *Arabidopsis thaliana* in Columbia ecotype naturally generates permutations in the order of organs along the stem. The permutations are more frequent in *ahp6* mutant. Therefore, it is mandatory to take into account this type of perturbation in the analysis of either other ecotype or in the analysis of other mutants.

[8] http://openalea.gforge.inria.fr/dokuwiki/doku.php?id=openalea
[9] http://sphinx.pocoo.org/

## 3.9 Conclusion

In this chapter we proposed a model and associated algorithms to deal with perturbations observed in measured sequences of divergence angles [Refahi *et al.*, 2011]. The model is based on the hypothesis that the perturbations are due to permutations of successive organs. The application of the proposed algorithms confirms the plausibility of the permutation hypothesis, and characterizes the patterns observed in *ahp6* mutant. The model is credited by the results of the statistical model as more arguable since the statistical approach shows highly regular patterns in the measured sequences. These patterns turn out to be generated by permutations involving at most 3 successive angles.

Permutations in the order of organs along the stems of plants had been predicted by numerical simulations based on theoretical abstract models [Douady *et al.*, 1996(a), Douady *et al.*, 1996(b)]. Here we report such permutations in spiral phyllotaxis of wild-type *Arabidopsis thaliana*. The frequency of permutations suggests that permutations are very common. Couder [Douady *et al.*, 1996(a)] only reported permutations in plants with a Lucas spiral arrangement, which concerns a very small proportion of population. Here we showed that permutations can also occur in Fibonacci spiral phyllotaxis. We identified *aph6* as a mutation that specifically increases the number of permutations demonstrating a genetic control of this phyllotactic instability.

This study helps us to understand the perturbations occurring in the phyllotaxis of the mutant and to reveal intrinsic properties of phyllotaxis. The mechanisms controlling phyllotaxis are likely to be strongly deterministic as suggested by biological experiments [Besnard *et al.*, 2011] and the models.

In the future, this kind of approach could be used to analyse other mutants. In mutants, as shown for the *ahp6* mutant, such an analysis allows a precise characterisation and quantification of the phenotype. The analysis can be done in different species with different phyllotactic arrangement.

The model can be combined with other phyllotaxis models to study different hypothesis. Though our approach in the current chapter can be considered as geometrical, the model can be unified with a dynamical/physiological approach. It is now well-known that the macroscopic aspects of plants are controlled by the regulatory networks involving genes and hormones. In the next chapter the proposed algorithms are used to evaluate the output of a lower scale model to shed light on the role the meristem zones. Moving to lower scale in the plant modelling, we will try to understand more about the origin of the perturbations.

# Chapter 4

# Phyllotaxis at meristem scale: assessment of a model of perturbation

## Introduction

In Chapter 3, we modelled and analysed the perturbations observed in the phyllotaxis of *Arabidopsis thaliana*. We proposed a combinatorial model, and developed algorithms that were used to show that the perturbations are well explained by permutations of successive organs along the stem. Notably, we have seen that local permutations of organs along the generative spiral of *Arabidopsis* are far from being rare events.

As a consequence, dynamic models of phyllotaxis should be able to reproduce that particular type of perturbations. In this chapter, we show how the tools and algorithms introduced in Chapter 3 can be used to assess the extent to which models achieve that goal. To illustrate this, we analyse the outputs of simulations based on a model implemented by Vincent Mirabet at ENS Lyon to study different perturbation factors on phyllotaxis. This implementation is meant to extend a model proposed by Douady and Couder [Douady *et al.*, 1996(b)] presented in Chapter 1, by adding noise in a simple way to the main geometric parameters of the original model. Namely, the noise was added to sensibility to the inhibitory fields, or to the size of the central zone of the shoot apical meristem.

In the following, we first recall the models proposed by Douady and Couder presented in Chapter 1. Then we describe and formalise the modifications that Vincent Mirabet has added to introduce perturbations in the simulations. The tools developed in the previous chapter are applied to these simulated data. An exploratory analysis shows that permutations by themselves are not enough to explain all the perturbations. Besides the permutations, the model induced

another type of perturbations which seems to be due to inversions in the orientation of the generative spiral. We propose a method to study these new perturbations, which relies on the tools of Chapter 3 to evaluate how well-defined is the spiral orientation in a sequence of divergence angles. This method is then applied to the data from the previous chapter and to the simulated sequences, and the similarities are discussed.

## 4.1 Models proposed by Douady and Couder: recall

In the Chapter 1, we presented the models proposed by Douady and Couder. In this chapter we study a modified version of their second model [Douady *et al.*, 1996(b)] in which they revisited the conditions of apparition of organs put forward by Snow and Snow (see Chapter 1). Here we recall this model. In this model, the pre-existing organs create fields that prevent nearby formation of a new organ. More precisely, in the model the effect of fields generated by $l$ pre-existing organs, $P_1, ..., P_l$, on a sampling point $S$ on the meristem is given by:

$$h(S) = \sum_{i=1}^{l} f(d(P_i, S)),$$

where $d(P_i, S)$ is the distance between organ $P_i$, $1 \leq i \leq l$ and the sampling point $S$. The function $f$ is an inhibition function that represents the inhibitory contribution of organs and is defined as:

$$f(d) = \frac{-1 + \left(\tanh\ \alpha \frac{d}{d_0}\right)^{-1}}{-1 + (\tanh\ \alpha)^{-1}}, \tag{4.1}$$

where tanh is the hyperbolic tangent and $\alpha$ represents the stiffness of the interaction. An organ appears on the flanks of the meristem where and when the function $h(S)$ is below a predefined threshold $E_s$. Once an organ is formed there is no angular reorganization outside a circular region of radius $R_0$. This can be considered as the size of central zone. The authors showed that the phyllotactic pattern that can be generated by the model is characterized by a geometrical parameter

$$\Gamma = \frac{d_0}{R_0}, \tag{4.2}$$

where $d_0$ is the radius of the organs.

### 4.1.1 Introducing noise in the second model

In Chapter 3, we showed that the measured divergence angles of both *Arabidopsis thaliana* wild-type and *ahp6* mutant were subject to noise and variations. The measured divergence angles were distributed around the golden angle and its multiples. This observation raises the question of how such variations can happen. In particular, we have shown that very structured patterns

of permuted organs arise in both the wild type and the mutant, with more permutations in the latter case.

Although permutations of two organs were observed in the model proposed by Douady and Couder [Douady *et al.*, 1996(a)], they seemed to occur only for a restricted range of parameters, corresponding to higher order phyllotaxis. In other words, this model does not seem to generate permutation patterns for the parameter ranges consistent with the phyllotaxis of *Arabidopsis thaliana*.

As a first attempt to include perturbations in a dynamic model of phyllotaxis, V. Mirabet (ENS Lyon) has chosen to introduce some stochastic parameters in the second model proposed by Douady and Couder. More complex models, like those which consider active auxin transport, have a large number of parameters. It it thus difficult to investigate the role of stochasticity in such models [Mirabet *et al.*, 2011]. The model proposed by Douady and Couder has the ability to generate most of the known phyllotactic patterns. In a preliminary study, Gaussian noise terms were added to three parameters: the size $d_0$ of primordia, the size $R_0$ of the central zone and the organ initiation threshold $E_s$ for the inhibitory field (1.1). In the first case, only very perturbed sequences could be found, with no satisfactory resemblance to experimental data. This led to the implementation of two models to study the role of noise in phyllotaxis, to which we give names now to ease the discussion. In the first model, a source of noise affects $E_s$, the threshold of minimum energy for organ initiation and is called "primordium noise", denoted shortly by *P-noise*. The second source of noise affects the size of the central zone (i.e. $R_0$). The latter is called "central zone noise" and is denoted by *CZ-noise*. Hereafter the model in which the central zone size varies is denoted by $M_{CZ}$ and $M_P$ denotes the model in which the $E_s$ is affected by the noise.

In practice to calculate the position of a new organ in model $M_{CZ}$ a random parameter $\Delta r$ is drawn at each time step, using a Gaussian distribution. The new radius of the central zone is $R_0 \pm \Delta r$, see Figure 4.1. Similarly, in $M_P$ model, to set the new threshold for each new primordium a random parameter $\Delta E$ is drawn, again using a Gaussian distribution. The new threshold of organ initiation is $E_s \pm \Delta E$.

V. Mirabet performed a series of simulations for both the $M_P$ and $M_{CZ}$ models. He explored different parameter values for both models: both of the parameters $\alpha$ and $\Gamma$ (see Equations (4.1) and (4.2)) were varied, as well as the parameters of the Gaussian distributions affecting $R_0$ and $E_s$. The simulations were time-consuming therefore he tested different parameter values that generated sequences that seemed the closest to real biological. In the end he selected ranges of parameters $(\alpha, \Gamma)$ and for each of these he varied the amplitude of the noise. This resulted in a dataset of 600 sequences (all having 157 angles), denoted $M_{CZ}^{sim}$ and $M_P^{sim}$ respectively.

These sets were doubled, according to the following observation. Due to the noise, some or-

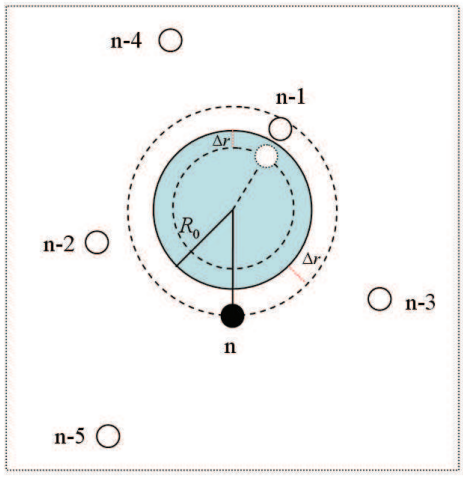



Figure 4.1: Variation of the central zone size. The new radius of the central zone is $R_0 + \Delta r$. The newly forming organ number $n$ is represented in black. The dotted circle is where the organ number $n-1$ was first formed. As we can see in the figure the organ number $n$ is forming further from the central zone than the organ number $n-1$. Hence in $M_{CZ}^{order}$ simulations the organ number $n-1$ appears before the organ number $n$.

gans may emerge further from central zone than the previous ones in the course of a simulation, see Figure 4.1. In this case, two choices are possible: either the order of organs is left unchanged, or they are sorted according to their distance from the centre. In the second case, these simulations are referred to as *re-ordered simulations* and are denoted by $M_{CZ}^{order}$ and $M_P^{order}$ according to the noise source, see Figure 4.1.

More precisely, for either of the models $M_{CZ}$ and $M_P$, let $(P_1, ..., P_l)$ be the generated sequence of organs (primordia) such that $P_i$ is initiated before $P_{i+1}$, $1 \leq i < l$, $(D_1, ..., D_l)$ be the sequences of distances of organs from the centre (i.e. $D_i$ is the distance of organ $P_i$ from the center). In addition let $x_i$ be the divergence angle between organs $P_i$ and $P_{i+1}$. The sequence returned by $M_{CZ}$ or $M_P$ is $(x_1, ..., x_{l-1})$. We define $(P'_1, ..., P'_l)$, as a permutation of $(P_1, ..., P_l)$, such that $D'_i \leq D'_{i+1}$ ($D'_i$ is the distance of the $P'_i$ from the centre). We call the corresponding sequence of divergence angles a re-ordered simulated sequence.

In real plants, the apparition of new organs at the meristem is followed by a series of events of growth and differentiation. Only little is known about the precise mechanisms through which a young organ leaves the meristem to become a separate organ on the stem. Events that follow this transition are called post-meristematic.

The re-ordering of organs amounts to the fact that the final position of organs on the stem is entirely determined by the position of primordia on the meristem. In other words there is no

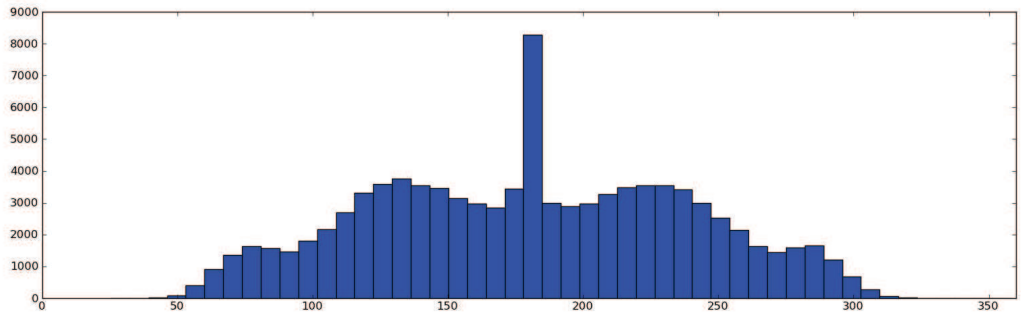



Figure 4.2: Histogram of angles for a typical subset of the simulations generated by $M_{CZ}$ model. One peak close to 180° can be observed. In addition four modes close to $\alpha = 137.5°$, $-\alpha = 222.5°$, $2\alpha = 275°$, and $-2\alpha = 85°$ cab be observed.

post-meristematic re-arrangement in $M_{CZ}^{order}$ and $M_P^{order}$. The difference between the order of organs on the stem and the order of primordia initiation on the meristem in $M_{CZ}^{sim}$ and $M_P^{sim}$ can be interpreted as a post-meristematic phenomenon.

## 4.2 Quantification of spiral inversions

Introducing two sources of noise and reordering the sequences led us to four sets of simulations $M_{CZ}^{order}$, $M_{CZ}^{sim}$, $M_P^{order}$, and $M_p^{sim}$. We applied the algorithms introduced in Chapter 3 to the sets of simulations, to check whether they contained non-trivial $n$-admissible sequences.

From a first inspection of the outputs, it appeared that the simulated sequences did not seem to have a unique predefined generative spiral orientation. This implies notably that the distribution of divergence angles have two modes close to 137.5° and 1360°−137.5°, see Figure 4.2. Therefore, we developed an exploratory tool to confirm that impression.

In order to detect potential changes in spiral direction, we compared the theoretical predictions of the algorithms of Chapter 3, when applied to the sequences interpreted as clockwise and counter-clockwise spirals, respectively. Because the algorithms of are able to qualify some angles as "non-explained", the idea of this method was to explore the sub-sequences of angles that could be explained for one spiral orientation only. Let us now describe this procedure in details.

Let $X = (x_1, ..., x_\ell)$ be a simulated sequence (of one of the models $M_{CZ}^{sim}$, $M_{CZ}^{order}$, $M_P^{sim}$, or $M_P^{order}$). We denote the sequence with opposite spiral orientation as $\bar{X} = (\bar{x_1}, \bar{x_2}, ..., \bar{x_\ell})$ where $\bar{x_i} = (360° - x_i)$, $1 \leq i \leq \ell$.

We used the combinatorial tools of the preceding chapter as follows:

1. The $n$-`admissible tree algorithm` was applied to both $X$ and $\bar{X}$. In case of multiple candidate sequences, the most probable was selected using associated posterior probabilities. Let $S(X)$, $S(\bar{X})$ denote the returned sequences of theoretical divergence angles.

2. For non-explained angles, we applied the backtracking procedure described in section 3.6.1, to invalidate all the preceding angles whose theoretical prediction could not be trusted. Invalidated angles as well as the non-explained angles are replaced by "?" in the sequences $S(X)$ and $S(\bar{X})$.

Let us denote $S(X) = (\mu_1, ..., \mu_\ell)$ and $S(\bar{X}) = (\bar{\mu_1}, ..., \bar{\mu_\ell})$, where all $\mu_i$ and $\bar{\mu}_i$ belong to $D_n \cup \{?\}$, and $D_n$ is the set of theoretical angles (see Property 3.2 in Chapter 3). At this stage, every angle $\mu_i$ (or $\bar{\mu}_i$) was either in $D_n$, and thus considered as explained by the procedure, or equal to "?". We defined a unique sequence $W = (w_1, ..., w_\ell) \in \{R, L, B, 0\}^\ell$ as follows:

- $\underline{w_i = L}$ if $\mu_i \in D_n$ and $\bar{\mu}_i =$ "?", *i.e.* $\mu_i$ is only explained in clockwise orientation.
- $\underline{w_i = R}$ if $\mu_i =$ "?" and $\bar{\mu}_i \in D_n$, *i.e.* $\mu_i$ is only explained in counter-clockwise orientation.
- $\underline{w_i = B}$ if $\mu_i \in D_n$ and $\bar{\mu}_i \in D_n$, *i.e.* $\mu_i$ is explained for both orientations.
- $\underline{w_i = 0}$ if $\mu_i =$ "?" and $\bar{\mu}_i =$ "?", *i.e.* $\mu_i$ is neither explained in clockwise nor in counter-clockwise orientation.

It may happen that a clockwise oriented spiral be explained by complex permutation patterns in the counterclockwise orientation (or vice versa). In consequence, a well-defined spiral without any inversions is expected to be coded by a word $W$ involving only $B$s and $L$s, or only $B$s and $R$. A sequence involved in generative spiral inversions should contain both $R$ and $L$ symbols, and thus the above coding can be used as a tool to track sequences involving spiral inversions.

The occurrence of 0's was expected to be correlated with those portions of sequences for which the algorithms fail, in general due to the fact that $n$-admissibility does not encompass the totality of perturbations present in phyllotactic sequences (see also Chapter 3).

To quantify to what extent a sequence follows a well defined spiral, we counted the symbols $L$, $R$, $B$ and 0 in the corresponding $W$ sequence. Formally we define the function $f$ as follows:

$$f(X) = (\frac{N_L}{\ell}, \frac{N_R}{\ell}, \frac{N_B}{\ell}, \frac{N_0}{\ell})$$

where $N_L$, $N_R$, $N_B$ and $N_0$ are respectively the number of $L$, $R$, $B$ and 0 in the corresponding $W$ sequence of length $\ell$. To represent the quadruplet on a plane we reduce it to $(\frac{N_L}{\ell - N_0}, \frac{N_R}{\ell - N_0}, \frac{N_B}{\ell - N_0})$, where the denominators ensure that these numbers sum to 1, and can thus be interpreted as

barycentric coordinates of a point in a triangle. The vertices of this triangle are marked by **B**, **L** and **R**. The color of the point will be specified by $\frac{N_0}{\ell}$. Let $(x_L, y_L)$, $(x_R, y_R)$ and $(x_B, y_B)$ be respectively the coordinates of **L**, **R** and **B** points that represent the vertices of the triangle. The barycentric coordinates of the point $P$ corresponding to each simulation are:

$$x_P = \frac{N_L}{\ell - N_0} x_L + \frac{N_R}{\ell - N_0} x_R + \frac{N_B}{\ell - N_0} x_B$$
$$y_P = \frac{N_L}{\ell - N_0} y_L + \frac{N_R}{\ell - N_0} y_R + \frac{N_B}{\ell - N_0} y_B.$$

The closer the point is to **L** (respectively **R**) the more the sequence contains $L$ ($R$ respectively) symbols. A point situated at equal distance of **L** and **R** contains the same number of $L$ and $R$ symbols. More generally, any point situated in the interior of the triangle potentially represents a sequence containing spiral inversions. This graphical representation will thus be used as a heuristics to detect spiral inversions in the simulated sequences.

## 4.3 Results

### 4.3.1 Investigating spiral inversions in wild-type and ahp6 mutant

In the preceding chapter, we analysed observed perturbations in wild-type and in *ahp6* mutant. The perturbations were mostly explained by permutations of successive organs on the stem. We did not suspect any obvious inversion of the generative spiral orientation. To further explore this aspect we applied the method introduced in Section 4.2 to data set of Chapter 3, see Figure 4.3. We present here the results for the whole data set, as there was no significant difference between the wild-type and mutant data sets.

In the Figure 4.3 the color corresponds to the proportion of non explained angles to the size of sequence. As one might notice most of the points have dark blue colors. Thus most of the sequences are explained (and are 3-admissible) for at least one orientation. It also appears that the points are mostly on the edge between **L** and **B**. In other words, most sequences are explained either as clockwise oriented, or for both orientations, with apparently all possible proportions of $L$ and $B$ (there are points on the whole edge, from the vertex **L** to **B**). The bias toward $L$, rather than $R$, is due to the fact that prior to application of the algorithms, the orientation of the sequences had been manually set to clockwise. There are also some points in middle of the triangle, which thus contain both $L$'s and $R$'s. The corresponding sequences will be described later, but let us first discuss the occurrence of $B$'s.

First, we plotted the histogram of angles that were explained in both orientations (i.e. coded by $B$), see Figure 4.4(c). There is a peak close to 137.5°, but other multiples of 137.5° also seem to be present. This is reminiscent of the whole distribution of measured angles (see Chapter

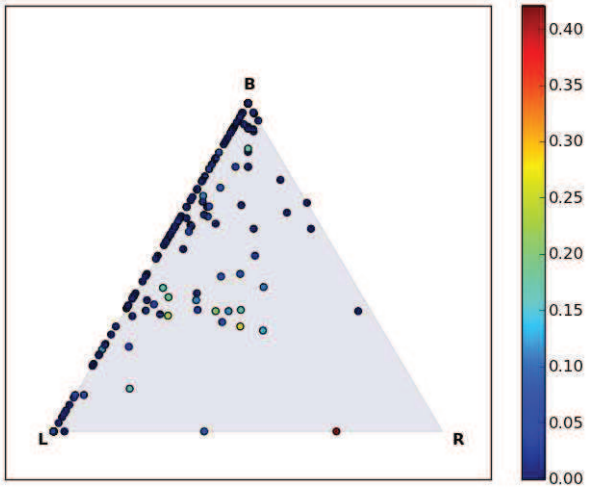
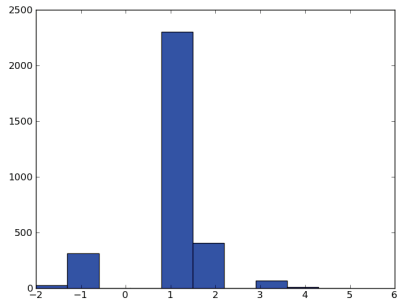



Figure 4.3: A geometric representation of $W$ sequence for the data set of Chapter 3. The cloud of points is mostly on the line between **L** and **B**. The color corresponds to the proportion of angles explained for neither orientation.

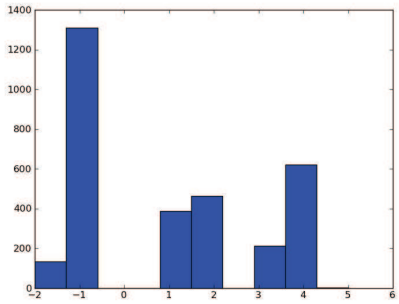
3), although with fewer elements. In Figures 4.4(a) and 4.4(b), the distributions of predicted theoretical angles are shown, for both the clockwise and counter-clockwise orientations. It turns out that most angles are seen as $\alpha = 137.5°$ in clockwise orientation, whereas the counter-clockwise prediction requires various multiples of $\alpha$, the most represented being $-\alpha$, $4\alpha$ and $2\alpha$, in that order.

A closer look at the sequences themselves led to the following explanation. Consider a subsequence of angles close to $\alpha = 137.5°$. In the opposite spiral orientation, and with the standard deviation we used (based on the estimation made in Chapter 3), an angle close to $360° - \alpha = 222.5°$ can be coded as $4\alpha = 190°$ or $2\alpha = 275°$. These angles can lead to long permutation patterns like for instance $(2\alpha, -\alpha)^*, 2\alpha$ and $(4\alpha, -\alpha, -\alpha, -\alpha)^*$, where the $^*$ indicates that the subsequence can be repeated arbitrarily: both these forms of patterns correspond to 3-admissible sequences (in fact, 2-admissible for the first one). It follows that sub-sequences of angles close to $\alpha = 137.5°$ can be interpreted in the opposite orientation as patterns like the ones just described leading to segments of $B$s in the sequence $W$.

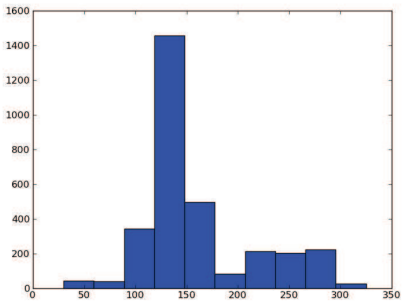
A manual inspection confirmed this interpretation: most angles coded by $B$ that were close to 137.5° had received an explanation involving repeated permutations like the ones above, in the counter-clockwise orientation. Some others, far from 137.5°, were explained as permutations in the clockwise orientation, and as baseline segments in counter-clockwise. After manual inspection




(a) Histogram of angle codes in clockwise analysis explained in both orientations.




(b) Histogram of angle codes in counterclockwise analysis explained in both orientations.




(c) Histogram of angles coded by $B$

Figure 4.4: Histograms of angles coded by $B$ of the data set of the Chapter 3, including both wild-type sequences and the mutant sequences. In (a) and (b) the distributions of predicted angles in $D_3 = \{-2, -1, 1, 2, 3, 4, 5\} \times \alpha$ ($\alpha$ omitted on abscissa) for clockwise and counter-clockwise orientations, respectively. (c) Histogram of angles explained in both clockwise and counter-clockwise orientations.

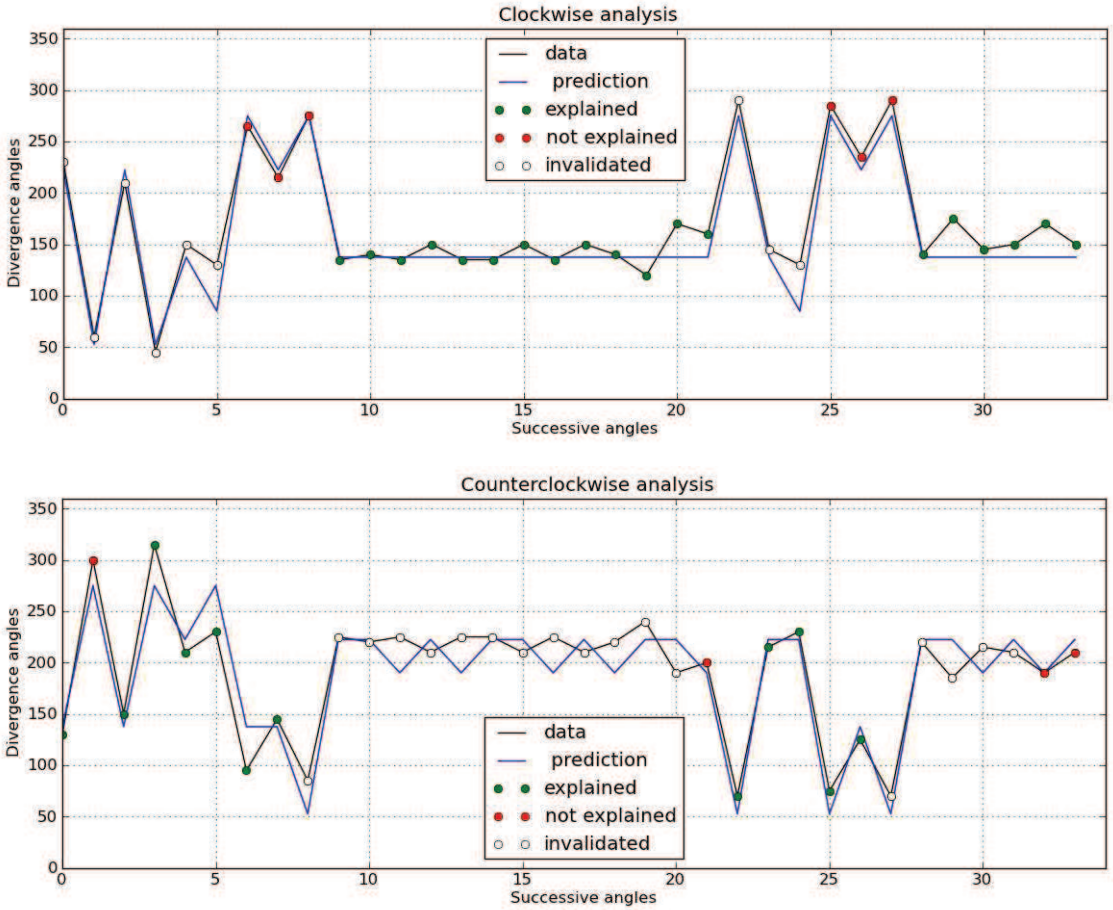



Figure 4.5: Output of the algorithms from Chapter 3, for a wild-type sequence whose coding contains both $L$'s and $R$'s. Both clockwise (top) and counter-clockwise (bottom) orientations are shown.

again, in none of these instances the counter-clockwise interpretation seemed plausible.

A few number of $B$'s were in fact not explained by these permutation motifs, but were often isolated angles, surrounded by non-explained sub-sequences, and rather close to 180°, which can be interpreted as 137.5° by the algorithms of the Chapter 3 for both orientations. Note however that this case concerned a minority of $B$'s.

Before concluding this section, let us discuss the sequences corresponding to points inside the triangle in Figure 4.3, which thus contain both $L$'s and $R$'s. We manually checked these sequences, by selecting all sequences having at least 20% of $R$'s, *i.e.* with barycentric coordinate associated with the vertex **R** greater than 0.2. Among these sequences there were all the most perturbed ones of the dataset, including the two Lucas phyllotaxis from the mutant dataset (see Chapter 3). These sequences were typically those for which the expert verification mentioned at the end of the previous chapter did not lead to a complete explanation.

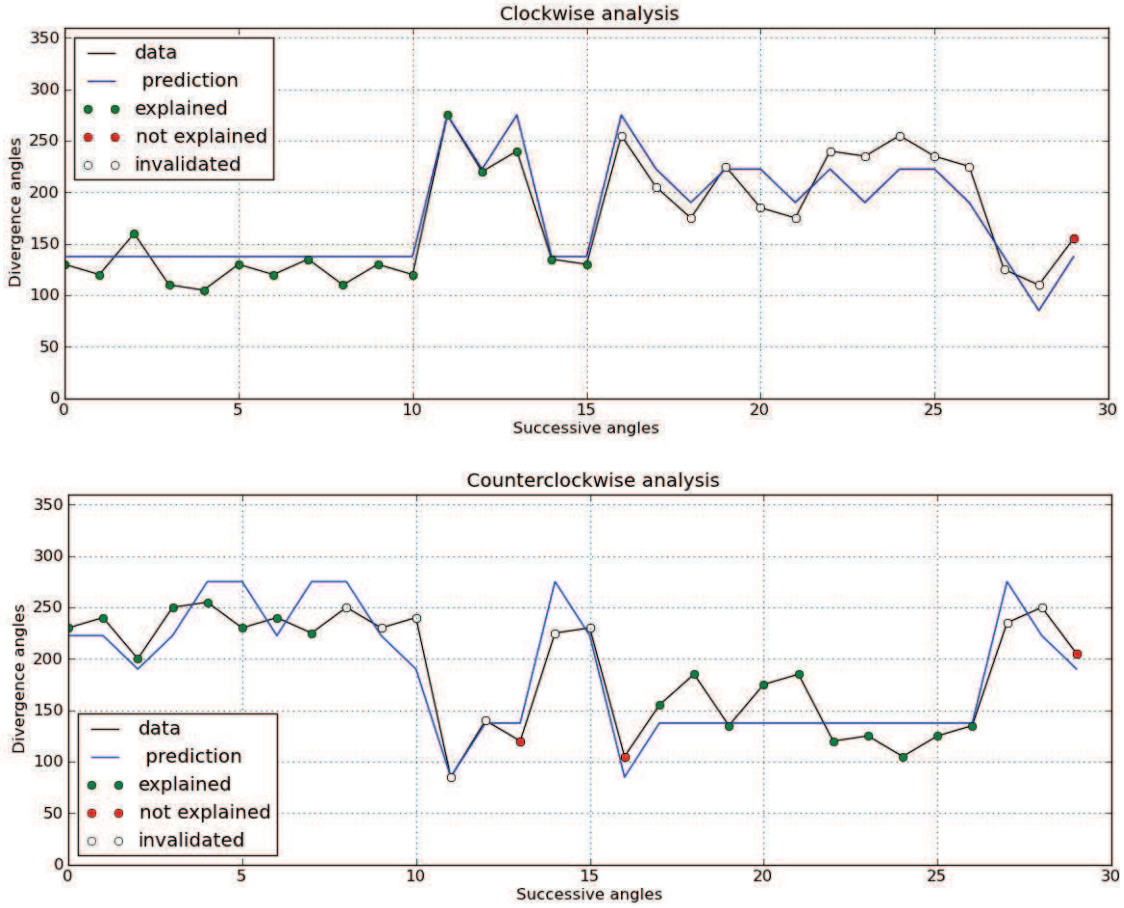



Figure 4.6: A mutant sequence, whose coding contains both $L$'s and $R$'s. This is the only example where a change of orientation might be considered plausible. Both clockwise (top) and counter-clockwise (bottom) orientations are shown.

As an example, see Figure 4.5, the beginning of the sequence is highly perturbed but can be explained by a non-plausible permutation in counter-clockwise orientation. Also, an isolated angle close to 250° (at abscissa 22) induces an invalidation of the obvious two-permutation that follows ("M-shape"), whereas an unlikely permutation is found, counter-clockwise.

Finally, let us show the unique example we have found in these sequences, were a reversion of the spiral orientation was plausible, in Figure 4.6.

In brief, a high proportion of $B$ were in fact resulting from angles close to 137.5°: a simple explanation (repeated $\alpha$) was found in clockwise orientation, which by parsimony we would prefer to the chained permutation patterns found in counter-clockwise orientation. In other words, the $B$'s that are due to this artefact would in fact better be recoded as $L$'s, and there is no apparent change of spiral orientation in the data.

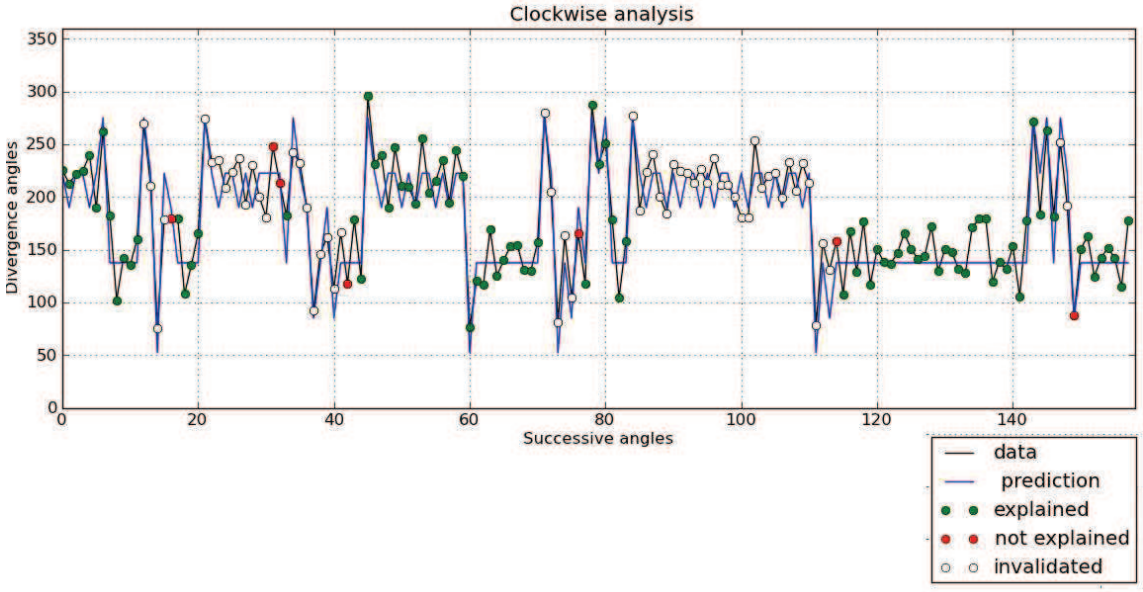



Figure 4.7: A simulated sequence of divergence angles of the model $M_{CZ}^{sim}$. The sequence is analysed using the tools introduced in Chapter 3. Two sub sequences whose baselines are $360° - 137.5° = 222.5°$ and $137.5°$ can be seen after the angle 84. This can indicate a generative spiral inversion.

### 4.3.2 Analysis of the simulations

Let us now discuss the outcome of the same procedure, applied to the simulated sequences from the models described in section 4.1.1. Due to spiral inversions, the mixture between different modes in the distribution of angles was even more drastic than the data set of Chapter 3, see Figure 4.2. This made the estimation of parameters of our algorithms (standard deviation, posterior probability) difficult for the simulated sequences. Therefore, we set these parameters to the values estimated in the previous chapter for wild-type plants and *ahp6* mutants.

The models $M_{CZ}^{sim}$ and $M_{CZ}^{order}$ generated sequences that were highly perturbed. A considerable number of angles close to 180° were observed in simulations. The algorithms presented in Chapter 3, detected two-permutations and some three-permutations in the simulated sequences. The detected three-permutations of organs cannot be validated since they always occurred in very perturbed sub-sequences. Besides the permutation patterns, another type of perturbation was frequently observed, that can be interpreted as sustained inversions of the spiral orientation, see Figure 4.7. Based on previous section, where we have seen that such events did not occur in the biological dataset, we consider that these sequences were not "realistic", at least as far as the phyllotaxis of *Arabidopsis thaliana* wild-type and *ahp6* mutant are concerned.

| | Initial order simulation set | Re-ordered simulation set |
|---|---|---|
| CZ-noise | spiral inversions<br>two-permutations<br>unrealistic sequences | spiral inversions<br>two-permutations<br>unrealistic sequences |
| P-noise | few spiral inversions<br>two-permutations<br>almost realistic sequences | few spiral inversions<br>two-permutations<br>wild-type realistic sequences |

Table 4.1: Summary of analysis of simulations. Initial order simulation set refers to the simulations $M_{CZ}^{sim}$ and $M_{P}^{sim}$ and re-ordered simulation set refers to the simulations $M_{CZ}^{order}$ and $M_{P}^{order}$.

The models $M_{p}^{sim}$ and $M_{p}^{order}$ generated sequences similar to the measured wild-type sequences of *Arabidopsis thaliana* studied in the previous chapter. However, the models generated significantly more angles close to 180°. Although changes of the spiral orientation are observed in the simulations, they are much less frequent than in the sequences generated by CZ-noise based model. We detected only 2-admissible sequences and no higher order $n$-admissibility was observed in sequences.

The Table 4.1 summarizes the analysis. In the following, two sub sections we give more details about the simulations $M_{CZ}$ and $M_{P}$, respectively.

**CZ-noise**

We analysed the simulations of the model $M_{CZ}^{sim}$ and $M_{CZ}^{order}$. Figure 4.2 shows a typical histogram of $M_{CZ}$ simulations. There is one peak close to 180°. Four large modes close to $\alpha = 137.5°$, $-\alpha = 222.5°$, $2\alpha = 275°$, and $-2\alpha = 85°$ can be observed. The histogram is almost symmetric. An exploratory analysis taking into account these peaks and the symmetry suggests that inversions of spiral should be frequent in these simulations. We represented each set of simulations by the triangle described earlier, see Figure 4.8. Only a minority of points are not coloured in blue, and thus most angles were explained by our algorithms, for at least one orientation of the generative spiral. On the other hand, the cloud of points is centred on the center of the triangle, showing that the sequences are far from following well defined spirals. A manual inspection of these sequences showed us that a non-negligible proportion were perturbed sequences for which the algorithms failed, just like discussed for the biological dataset. However, we also found a high number of sequences presenting apparent change of the spiral orientation, see Figure 4.7.

The inversion of generative spiral was not observed in measured divergence angles of *Arabidopsis thaliana*. A simple comparison between Figures 4.8 and 4.3, suggests that the model $M_{CZ}$ generates sequences that are completely different from the dataset of measured wild-type and *ahp6* mutant sequences in terms of generative spiral orientation. Therefore, we consider

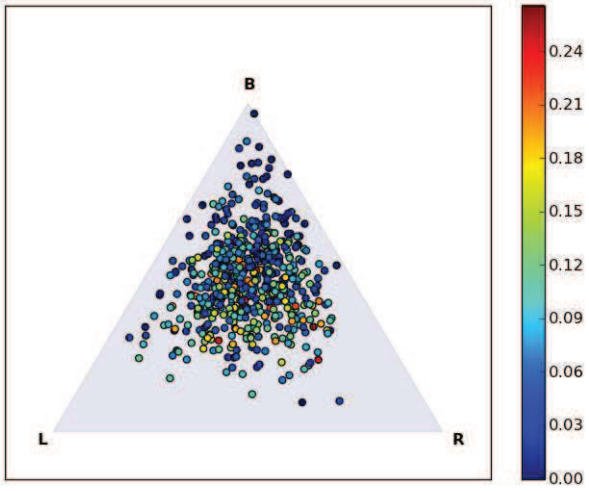



Figure 4.8: A geometric representation of $W$ sequence for a typical set of simulations of $M_{CZ}$. The cloud of points are in the center of the triangle. Few point are closer to **L** (**R** respectively) than to **R** (**L** repectively). The color corresponds to the proportion of non explained angles to the size of sequence. This diagram indicates that most sequences cannot be given a well-defined spiral.

these simulations "unrealistic". Despite the instability of the generative spiral, permutations of organs are detected. Among the detected permutations only permutations involving two successive organs can be validated, because three-permutations are always surrounded by too much noise and non-explained angles. The only sensible difference we found between $M_{CZ}^{sim}$ simulations and $M_{CZ}^{order}$ simulations is the number of permutations: the latter has more two-permutations. The inversions of generative spiral persists in both models.

Precisely because the model $M_{CZ}$ was not able to reproduce "realistic" sequences, it shows the importance of a robust central zone for phyllotaxis. Indeed, these simulations and our analysis show that variations of the central zone size during the process of phyllotaxis may induce strong perturbations of the phyllotactic spiral, which are qualitatively and quantitatively very distinct from our observations from *A. thaliana*.

### P-noise

For the $M_P$ simulated sets, the spiral inversions were much less frequent. In Figure 4.9, we represent the $M_P$ simulations by the triangle introduced earlier. Here again, most point are explained by our algorithms for at least one orientation of the generative spiral (they are coloured

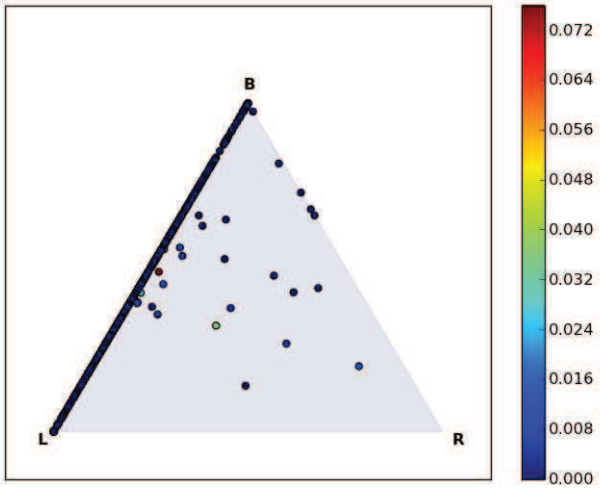



Figure 4.9: A geometric representation of $W$ sequence for a typical set of simulations of $M_P$. The cloud of points is mostly on the line between **L** and **B**. This indicates that the sequences are mostly in clockwise orientation. The sequences whose represented as points in the center of the triangle are almost exactly those with spiral inversion, see an example Figure 4.10.

in blue). The points are mostly on the line between **L** and **B**, in good agreement with the biological dataset, Figure 4.3. There are some points in the center of the triangle. We had already identified all sequences having generative spiral inversions. The points in the center correspond to these sequences.

Similarly to the data set of Chapter 3, the implementation of V. Mirabet was designed to ensure that the sequences begin with clockwise spiral, which explains the bias towards **L** instead of **R**.

Like for the real data, a significant number of points are near **B**. We observed the distribution of angles and theoretical predictions for the angle coded by $B$, see Figures 4.11(c), 4.11(a) and 4.11(b). Similar to the biological dataset, this could often be explained by chained permutations like $(2\alpha, -\alpha)^*, 2\alpha$ or $(4\alpha, -\alpha, -\alpha, -\alpha)^*$ in counter-clockwise orientation. By parsimony, we would again suggest that $B$ would in fact correspond to clockwise oriented baseline segments.

Importantly, some of the $B$ symbols found in our analysis had an other origin than the patterns we just discussed, as they were in fact isolated in their sequence. This can be explained by the already mentioned occurrence in the model of a considerable number of angles around 180°, see Figure 4.12(c). Since 180° can be coded by the algorithms to 137.5° or 222.5°, and an isolated angle of 137.5° (or 222.5° for the counterclockwise analysis) is always explained as a

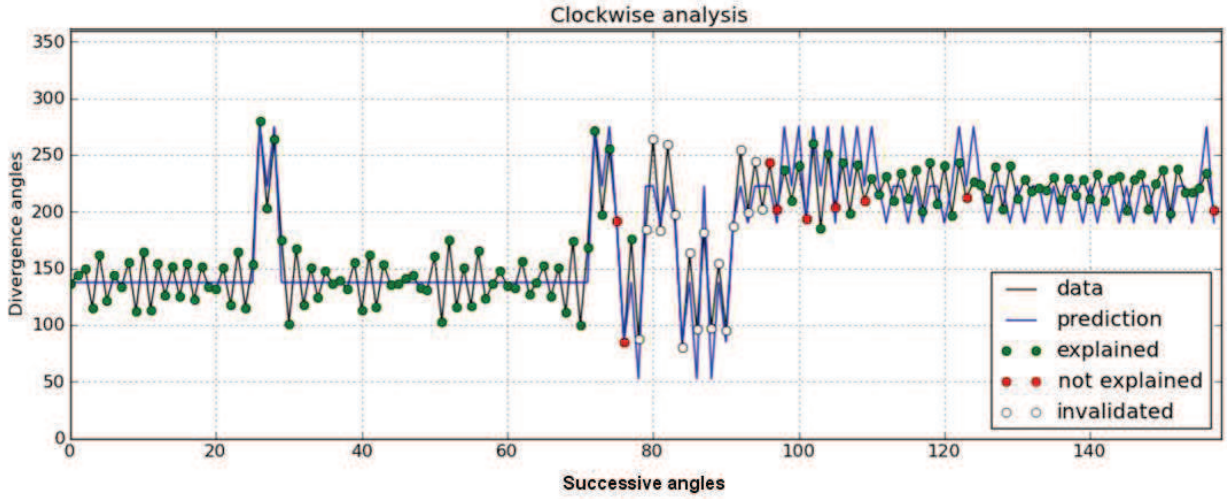


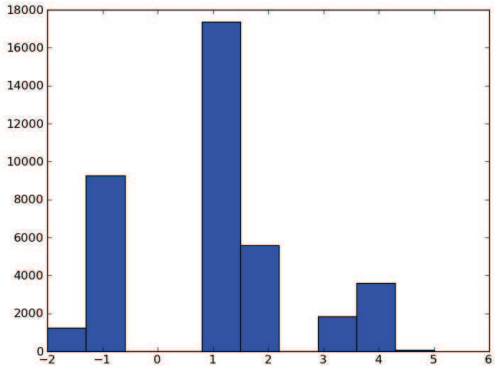

Figure 4.10: A sequence corresponding to one the points in the middle of the triangle of Figure 4.9. Spiral inversion can be observed at the subsequence at the end of the sequence.

occurring between non-permuted pair of successive organs, isolated $B$ followed from angles close to 180°, see Figures 4.12(a) and 4.12(a). Although this phenomenon was also observed with biological data, it is by far more widespread in the $M_P$ dataset.

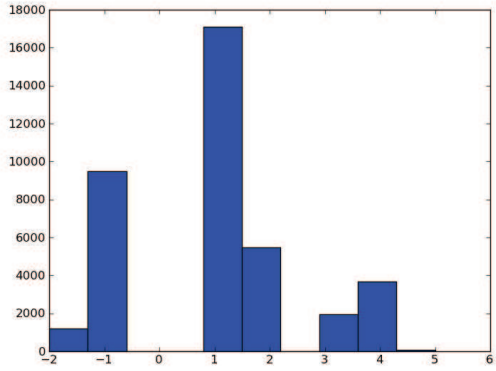
Since the points situated inside the triangle in Figure 4.9 were not too many, we investigated the associated sequences one by one. They turned out to all involve a change of spiral orientation, see an example on Figure 4.10. To go further, we manually inspected all the 600 sequences from this simulation set. The result of this inspection was that no other (visually significant) spiral inversion could be found in the data. In other words, the representation in Figure 4.9 provided here a way to exactly detect abnormal sequences.

On the basis of angle distributions, it appeared that the re-ordered set of simulations, $M_p^{order}$ creates more realistic sequences. This might be an indication that the final permutations of organs are likely to be mostly determined at the meristem scale, and that post-meristematic rearrangements are not required for such permutations to occur.

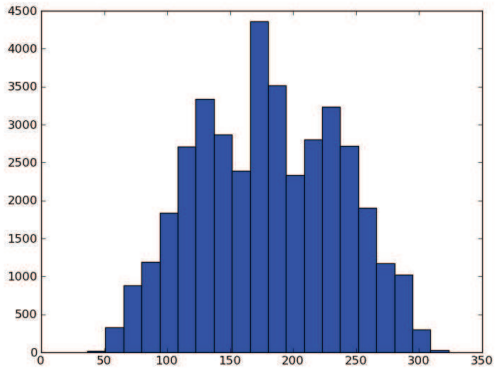
In conclusion of this section, the simulation sets $M_P$ were in good agreement with biological data for some parameter values. The main difference was the more widespread occurrence of angles close to 180°. On the other hand, it failed at reproducing sequences similar to the *ahp6* mutant, since no three-permutations were observed.




(a) Histogram of angle codes in clockwise analysis explained in both orientations.

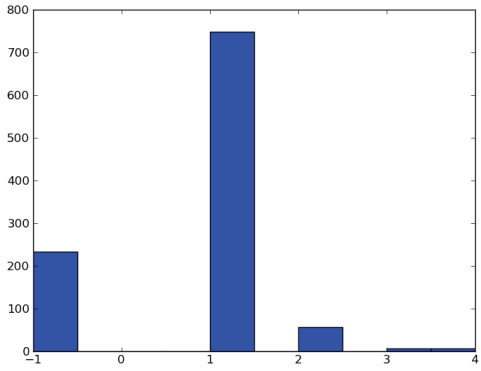



(b) Histogram of angle codes in counetrclockwise analysis explained in both orientations..

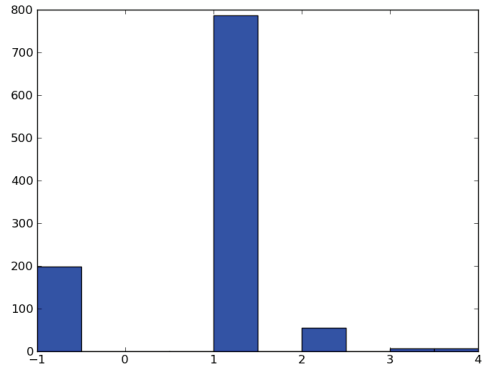



(c) Histogram of angles coded by $B$

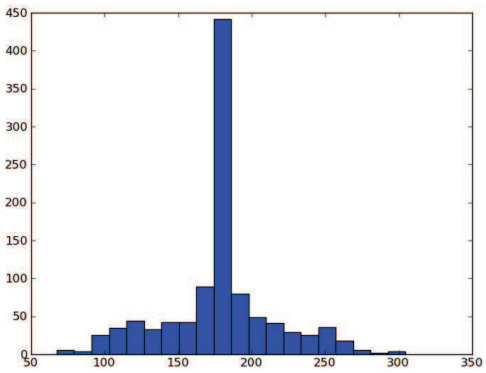
Figure 4.11: Analysis of the angles coded by $B$ of the model $M_P$. (a) Histogram of angle codes in clockwise analysis explained in clockwise and counterclockwise orientation. (b) Histogram of angle codes in counterclockwise analysis explained in clockwise and counterclockwise orientations. (c) Histogram of angles explained for both clockwise and counter-clockwise orientations, *i.e.* coded by $B$. In (a) and (b) $\alpha$ is omitted on abscissa.




(a) Histogram of isolated angle codes in clockwise analysis explained in both orientations.




(b) Histogram of isolated angle codes in counetrclockwise analysis explained in both orientations.




(c) Histogram of isolated angles coded by $B$.

Figure 4.12: Analysis of the isolated angles coded by $B$ of the model $M_P$. (a) Histogram of angle codes in clockwise analysis explained in clockwise and counterclockwise orientation. (b) Histogram of angle codes in counterclockwise analysis explained in clockwise and counterclockwise orientations. (c) Histogram of angles explained for both clockwise and counter-clockwise orientations, *i.e.* coded by $B$. In (a) and (b) $\alpha$ is omitted on abscissa.

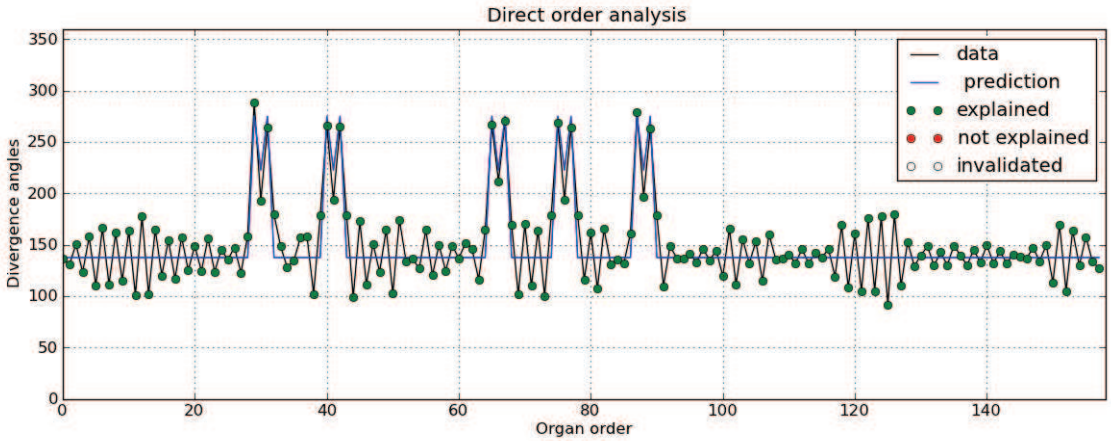



Figure 4.13: A simulated sequences if the model $M_P$. The sequence is similar to the measured sequences of wild-type. The M-shaped two-permutations are clearly visible.

## 4.4 Discussion

The variations of the central zone size induced generative spiral inversions. In case of the model plant *Arabidopsis thaliana* spiral inversions were considered unrealistic, although this phenomenon cannot be excluded in general; inversions have been said to occur in citrus for instance [ Schroeder, 1953]. This led us to consider the sequences with more care. It is noteworthy that the number of plants with clockwise and counter-clockwise generative spiral are equal in the data used in Chapter 3, and in the literature [Smyth *et al.*, 1990]. This indicates at least that both orientations are potentially occurring in nature, though the "choice" of an orientation seems to be irreversible.

The model $M_p$ generates sequences involving two-permutations in parameter ranges where the original model of Douady et Couder [Douady *et al.*, 1996(a), Douady *et al.*, 1996(b), Douady *et al.*, 1996(c)] did not generate double permutations.

In the simulations we observed a considerable number of angles close to 180°. This can be interpreted as change from spiral patterns to distichous patterns. This phenomenon has been recently observed in *plt3 plt5 plt7* multiple mutant of *Arabidopsis* [Prasad *et al.*, 2011]. In the model this can be explained by the effect of the sources of noise on the control parameter $\Gamma$. To calculate the position of a new organ random parameters are chosen according to a Gaussian distribution. Since $\Gamma$ depends on these parameters, and in fact is a bifurcation parameter that controls changes of phyllotaxis modes, the noise terms in both $M_P$ and $M_{CZ}$ can induce temporary changes in phyllotaxis. In particular, when $\Gamma$ changes the next divergence angle can change to 180° or to $360° - 137.5° = 222.5°$, see Figure 4.14. One possible way to add noise without

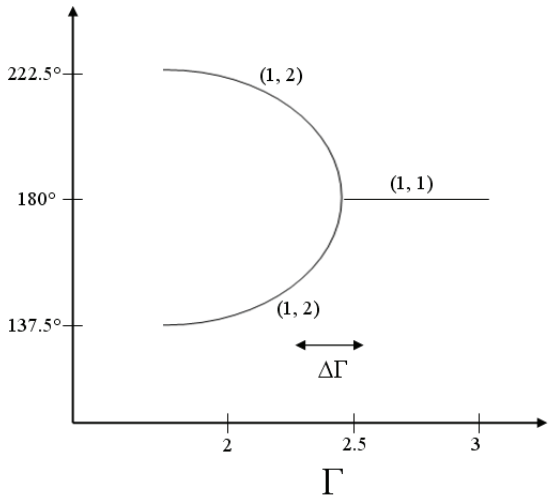



Figure 4.14: Diagram of the divergence angle as function of Γ. A change in the value of Γ induced by adding noise can change the value of divergence angles. (1,2) corresponds to the spiral pattern with 1 and 2 as its parastichy numbers. (1, 1) corresponds to distichous phyllotactic pattern.

inducing such angles can be adding noise such that the value of Γ does not change with noise.

The main conclusion of the simulations is that the variations in the central zone size results in drastic perturbations in phyllotaxis. Moreover this conclusion is supported by biological facts, as observed in clavata3 mutant [Reinhardt *et al.*, 2003(b), Pilkington, 1929] and justifies a lower scale study of robustness of central zone.

As for the model including variations in the inhibitory field threshold, it accounts rather well with our previous data set for some parameter values, as far as permutation patterns are concerned. However, it also induces changes in spiral orientation, though to a much lesser extent than the $M_{CZ}$ model, as well as a high population of divergence angles close to 180°.

## 4.5 Conclusion

In this chapter we used the tools for analysis of phyllotaxis perturbations proposed in Chapter 3 to assess a dynamical phyllotaxis model. The model investigated the impact of noise on the phyllotaxis. Two sources of noise were added to the model proposed by Douady and Couder [Douady *et al.*, 1996(b)]. The first noise influenced the central zone size and the second one influenced the threshold of organ initiation. The algorithms developed in the previous chapter have allowed us to verify that the noise added to the size of central zone generated unrealistic sequences, with generative spiral inversions. On the other hand, our algorithms also allowed us

to select some sequences in the second model that were similar to the wild-type sequences of Chapter 3. These sequences involved permutations of two consecutive organs along stem.

A future direction to improve such models would be to study other potential function of inhibitory fields to produce more realistic simulations. Adding sources of noise to the model can also be reconsidered. To reproduce realistic sequences, adding noise should not induce changes in phyllotactic modes. In other words, the control parameter of the model $\Gamma$ should stay in a limited, bifurcation free range. The noise can also be added to the velocity of organs moving further from the meristem center.

More importantly, this study shows how the tools and the model proposed in the preceding chapter can be used to evaluate other models. These tools are designed to recognize permutation patterns in perturbed sequences of divergence angles. Thus, they have to be combined with other methods detect other type of perturbations. For example, methods to detect changes of the canonical angle could encompass both inversions in the spiral orientation, and changes of phyllotaxis known to occur in nature: a change from $137.5°$ to $99.5°$, for instance, would correspond to a transition from Fibonacci to Lucas phyllotaxis, whereas a change from $137.5°$ to $-137.5°$ is more likely related to inversion in the spiral orientation.

In this chapter we showed that the variations in the central zone size results in unrealistic phyllotactic patterns. This highlights the importance of the central zone size in regulation of phyllotaxis. In the next chapter, we study the negative feedback between WUSCHEL, CLAVATA1, and CLAVATA3 genes, which controls the size of the central zone. Therefore, moving to a lower scale, we investigate the role of genes in controlling the size of the central zone.

# Chapter 5

# Modelling the regulation of the central zone on a digitized 4D meristem

## Introduction

In Chapter 1 and 4, we have seen that phyllotactic patterns, and their robustness, are controlled by the ratio of two geometric meristem parameters: the diameter of primordia, and the diameter of the central zone. The value of these parameters emerges from regulatory processes occurring in the meristem at molecular and cellular scales. As explained in Chapter 1, the formation of new primordia is essentially triggered at locations where auxin accumulation spots are formed, at the periphery of the SAM [Reinhardt *et al.*, 2000, Reinhardt *et al.*, 2003(b)]. The central zone, on the other hand, is controlled by a gene regulatory network, consisting in a negative feedback loop between WUS and CLV genes [Schoof *et al.*, 2000].

We have seen in Chapter 4 that the diameter of the central zone was particularly critical for the stability of phyllotaxic patterns, which are drastically disrupted when this parameter is perturbed. This is further supported by biological observations, as CLV loss-of-function mutants have a large central zone [Clark *et al.*, 1993, Lindsay *et al.*, 2006] and a highly perturbed phyllotaxis [Hobe *et al.*, 2003].

In the present chapter, we thus focus on models of the regulatory network controlling the formation and maintenance of the central zone. This follows previous studies, which were all done on fixed, 2D digital meristems [Hohm *et al.*, 2010, Jönsson *et al.*, 2003]. These models concerned CZ maintenance, and not its apparition when a newly formed meristem emerges. The absence of 3D model can be explained by the lack of data concerning meristem geometry, at the time when these studies were published. The apparition of a CZ in a new meristem, on the other hand,

has hardly been considered in the biological literature on stem cell population in the meristem, where maintenance is the main concern. In this chapter, we take advantage of recent 4D data (*i.e.* 3D with tracking over several time steps) [Fernandez *et al.*, 2010] on shoot apical meristem at cellular resolution. This allows us to revisit existing models, with a particular focus on the information provided by these new data sets, which is of geometrical nature.

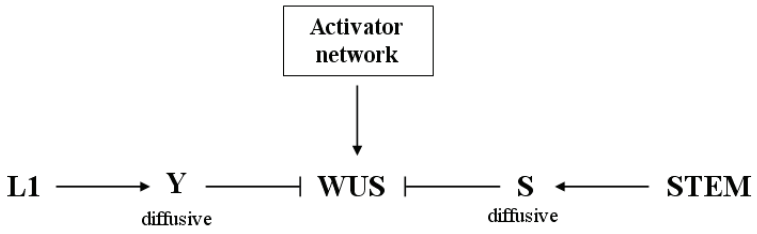
In Chapter 1 we presented a overview on biological facts and the current state of art of modelling the gene regulatory networks involving WUS within the SAM. In this chapter we present the models proposed by Jönsson *et al.* [Jönsson *et al.*, 2005]. Then we present the methods that allow us to have precise spatial information about a growing meristem over a period of time. Next, we describe a method to optimize parameter values of ordinary differential equation models, based on a specified equilibrium pattern. We then motivate the choice of one of the published model, which is based on signals diffusing from external layers of the meristem. We then investigate the mathematical properties of this model, using a 1D tissue, namely a row of cells. Finally we consider this model on a 4D digitized meristem and use the tools introduced in this chapter to investigate the ability of the model to generate realistic central zone patterns.

## 5.1 Models proposed by Jönsson *et al.*

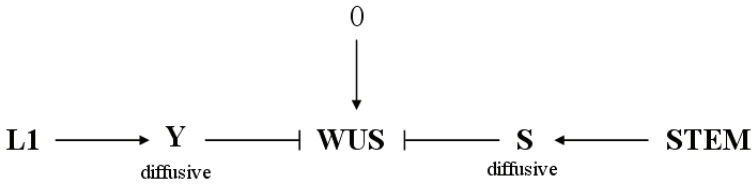
It is important to note here that most of the work presented hereafter will in fact concern the WUS expression domain, rather than the entire CZ itself. In other words, we will sometimes mention the CZ by abuse of language, in situation where our models concern in fact the organising center (OC).

Jönsson *et al.* [Jönsson *et al.*, 2005] proposed two models for organization of WUS expression domain. The models are called respectively *activator* and *repressor*. In both models, the dynamics is supposed to be controlled by pattern-forming reaction-diffusion mechanisms (see Chapter 2). The models are implemented on a two-dimensional tissue, representing a transversal cut of the SAM. They both include a putative signal diffusing from L1 layer that represses the expression of WUS. In both models CLV3 is not explicitly modelled, it is supposed to be a part of the L1 originating repressive signal. Another putative signal diffusing from cells situated on the stem, below the OC, represses the expression of WUS. Though the latter signal is not applicable in their two-dimensional system, which represents a disk-shaped horizontal slice of the SAM (see Figure 5.2), it is necessary in a three-dimensional implementation.

The main difference between the repressor and activator models is how WUS is induced. In repressor model the gene WUS is supposed to be expressed in all cells unless a repressive signal is present, whereas in activator model an activator network induces WUS expression, see Figure 5.1.

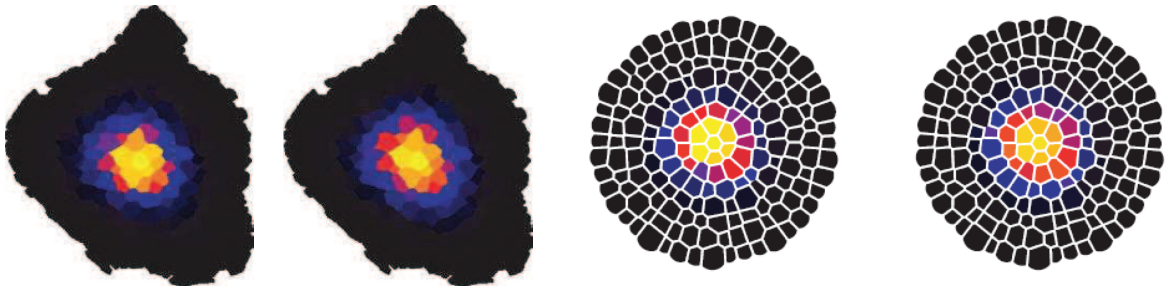



(a) Activator model




(b) Repressor model

Figure 5.1: Illustration of activator and repressor models. (a) In activator model the activator network induces WUS. Two signals emanating from L1 layer and stem repress the expression of WUS. (b) In repressor model WUS is in all cells unless one of the repressive signals originating from L1 layer or stem is present in the cell. However the model is three-dimensional, the signal $S$ is not included in the simulations of the models proposed by Jönsson [Jönsson *et al.*, 2005] since their models are restricted to a two-dimensional system which in fact is a two dimensional transversal cut of meristem, see Figure 5.2



(a) Activator model (b) Repressor model (c) Activator model (d) Repressor model

Figure 5.2: (a) Equilibrium concentration of WUS in activator model (b) Equilibrium concentration of WUS in repressor model (c) Equilibrium concentration of WUS in activator model in lattice simulation (d) Equilibrium concentration of WUS in repressor model in lattice simulation. The color code is defined as black (min) - blue - red - yellow (max). Figures from [Jönsson *et al.*, 2005].

Dynamics of both activator model and repressor model are described by differential equations. The three key components shared by the two models are:

1. Mass action kinetics is used to model molecular reactions. According to this law the rate of a reaction is proportional to the product of concentration of reactants.

2. Sigmoidal functions are used to model interaction of elements of the models. The regulation of a gene $i$ is modelled by

$$\frac{dv_i}{dt} = \frac{1}{\tau_i} g\Big(\sum\nolimits_{j\in J_i} T_{ij} v_j + h_j\Big),$$

where $v_i$ is the concentration of the corresponding protein and $g(x)$ is a sigmoidal function of form:

$$g(x) = \frac{1}{2}\Big(1 + \frac{x}{\sqrt{1+x^2}}\Big),$$

where $\tau_i$ and $h_i$ are the inverse of the maximum rate and the minimum expression level of element $i$ respectively. $J_i$ is the set of elements regulating the element $i$. $T_{ij}$ indicates the regulation of gene $i$ by gene $j$, positive $T_{ij}$ defines an activation of gene $i$ by gene $j$ while a negative $T_{ij}$ represents repression. This form of sigmoid function is different from the Hill functions discussed in Chapter 2. It is in fact reminiscent of neural network models, on which the authors of Jönsson *et al.* [Jönsson *et al.*, 2005] had been working previously. Note however that the dynamics of such systems is not strongly dependent on the precise type of sigmoid functions that are used [Plahte, *et al.* 1998].

3. Molecules move between neighbouring cells by diffusion. The diffusion of a molecule $x$ between cells $i$ and $j$ is modelled as follows:

$$\frac{dx_i}{dt} = D(x_j - x_i),$$

where $D$ is the diffusion constant. If we have precise spatial data about the SAM, the above equation can be replaced by:

$$\frac{dx_i}{dt} = D' \frac{A_{ij}}{d_{ij} V_i}(x_j - x_i), \tag{5.1}$$

where $A_{ij}$ is the contact surface between the cells $i$ and $j$ and $d_{ij}$ is the distance between the cell centres and $V_i$ is the volume of cell $i$, and $D'$ is a constant identical for all cells.

**Activator model**

The assumption of activator model is that WUS expression is induced by an activator. This activator is produced by a pattern-forming reaction-diffusion model, Brusselator model [Prigogine *et al.*, 1968] (see Chapter 2). The model is defined by a set of differential equations.

$$\begin{cases} \dfrac{dW}{dt} = \dfrac{1}{\tau_w} g(h_w + T_{wa}A + T_{wy}Y) - \delta_w W, & (5.2) \\ \dfrac{dY}{dt} = k_y \mathbf{1}_{L_1} - d_y Y + D_y \nabla^2 Y & (5.3) \\ \dfrac{dA}{dt} = a - (b+\beta)A + cA^2B - dY + D_a \nabla^2 A, & (5.4) \\ \dfrac{dB}{dt} = bA - cA^2B + D_b \nabla^2 B, & (5.5) \end{cases}$$

where $\nabla^2$ denotes a discrete diffusion operator. WUS concentration is denoted by $W$ and $\delta_w W$ is its degradation term. Activator $A$ induces $W$ and $Y$ represses $W$. $Y$ corresponds to the signal coming from L1 layer. It has also has a simple degradation term, $d_y Y$. $\mathbf{1}_{L1}$ is the indicator function of L1 layer which is one in L1 cells and 0 in other cells.

The Equations (5.4) and (5.5) are standard equations of Brusselator model except the degradation term $-dY$ in Equation (5.4) which represents repression from $Y$ (see Chapter 2, for Brusselator model where $a = U$, $V = b$, $\beta = 1$, $c = 1$).

Due to the discrete, cell-based geometry that is used, $\nabla^2$ is in fact the sum of terms as in Equation (5.1) over all neighbours of cell $i$, denoted by $\mathcal{N}_i$,

$$\left(\nabla^2 x\right)_i = \sum_{j \in \mathcal{N}_i} D' \frac{A_{ij}}{d_{ij} V_i} (x_j - x_i), \tag{5.6}$$

where $\left(\nabla^2 x\right)_i$ is the $i$th component of $\nabla^2 x$.

**Repressor model**

The simple idea behind this model is that WUS is expressed in a cell when a repressive signal is not present. In consequence, the model does not include any activator. Similar to activator model a signal emanating from L1 layer represses WUS. The system is described by the following set of equations:

$$\begin{cases} \dfrac{dW}{dt} = \dfrac{1}{\tau_w} g(h_w + T_{wy}Y) - \delta_w W, & (5.7) \\ \dfrac{dY}{dt} = k_y \mathbf{1}_{L_1} - d_y Y + D_y \nabla^2 Y, & (5.8) \end{cases}$$

The above equations are similar to the Equations (5.2) and (5.3), except the term $T_{wa}$ in Equation (5.2), which is related to the activator A, therefore not applicable in repressor model.

According to Jönsson *et al.* [Jönsson *et al.*, 2005], the models are both able to regenerate the wild-type WUS expression domain. They also predict the spatial expansion of WUS expression zone after removing or weakening the CLV3 signal. The removal or weakening of CLV3 in the simulations is implemented by weakening of the L1 originating signal. In addition the activator model is able to mimic the reorganization of the meristem after the ablation of the organizing center [Reinhardt *et al.*, 2003(b)], a behaviour that the repressor model fails to reproduce properly. Accordingly, they conclude that an underlying activating network is more likely to generate the WUS domain in response to a perturbation as strong as partial ablation of the meristem.

However, note that in the repressor model, the only non-linearity is the sigmoidal production rate of WUS, in Equation (5.7). This term depends only on the signal $Y$, which itself is entirely determined by the geometry of the L1 layer, since it is produced through diffusion of a substance created in this layer (and degraded everywhere as all proteins). Even though it is not able to reproduce WUS maintenance under strong perturbations, Jönsson *et al.* show that it is able to generate a plausible WUS domain. Because of this ability and its essentially geometrical nature, this model is a good candidate to investigate the link between meristem geometry and the shape of the WUS domain.

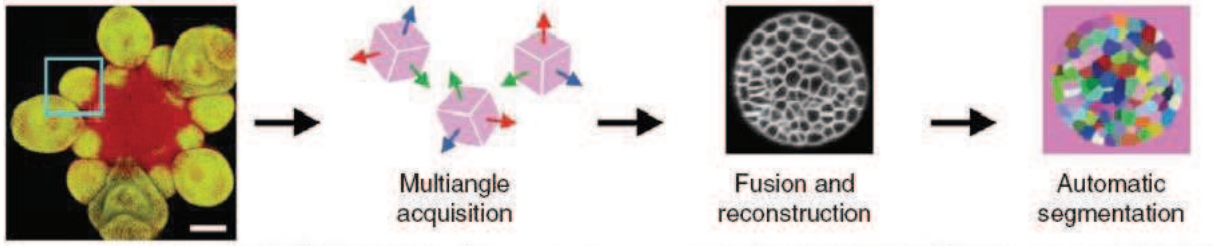
## 5.2 Construction of a realistic 4D meristem

In this section, we explain how we construct a digitized meristem from images obtained using confocal microscopy. We first explain the methods proposed by Fernandez *et al.* [Fernandez *et al.*, 2010] and how we extract the spatial information about cells. Fernandez *et al.* applied these methods to a floral meristem to construct a 4D digitized meristem, which we will use in this chapter. We define a target pattern on the digitized meristem.

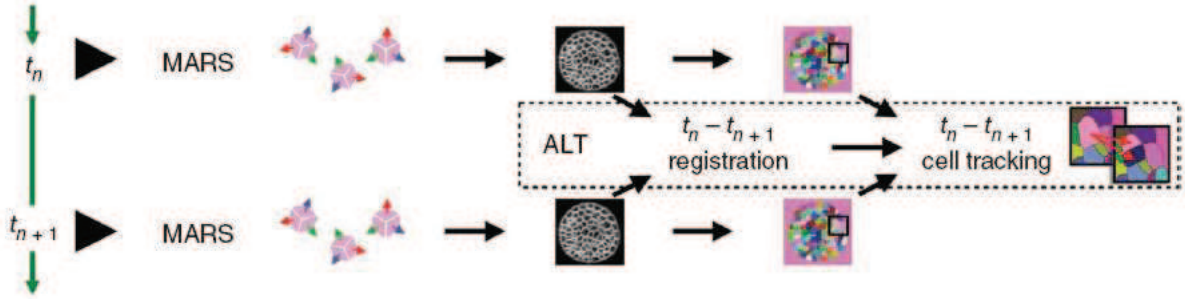
### 5.2.1 Methods proposed by Fernandez *et al.*

Fernandez *et al.* [Fernandez *et al.*, 2010] proposed a pipeline of processes to obtain quantitative information about growing meristems at cellular resolution. This method creates three-dimensional digitized tissues at cellular resolution and automatically extracts the cell lineage over a period of time. The method consists of two main components called MARS and ALT:

- MARS (multiangle image acquisition, 3D reconstruction and cell segmentation): first images are obtained from meristem using confocal microscopy under different orientations. Then the images are fused to create a single sample. The fusion improves the quality of cell outlines. Then the merged 3D image is segmented using a 3D watershed algorithm [Soille *et al.*, 1999], see Figure 5.3(a).




(a) Illustration of MARS.




(b) Illustration of ALT.

Figure 5.3: (a) MARS pipeline for reconstruction of a 3D tissue, description in the text. (b) ALT pipeline. Segmented images from two successive time points ($t_n$ and $t_{n+1}$) are compared to extract cell lineages. The registration refers to a non-linear intensity-based algorithm that matches the voxel intensities of two images.

- ALT (automated lineage tracking): first MARS is used to segment confocal images obtained at different time points $t_1, ..., t_m$ of a growing meristem. This generates a series of segmented images $(I_1, ..., I_m)$. Then ALT is used to track the cell lineage between each pair of segmented images. ALT treats this problem as a flow problem and calculates a series of lineages $(L_1, ..., L_{m-1})$ where $L_i$ is a list that contains pairs, where the first element is a mother cell in $I_i$ and the second element is the list of its daughter cells in $I_{i+1}$, $1 \leq i \leq m-1$.

### 5.2.2 Constructing 4D digitized meristem

The output of MARS-ALT is a series of three-dimensional segmented images and cell lineages. A three-dimensional segmented image can be viewed as a three-dimensional image in which the voxels (volumetric pixels) belonging to a given cell are marked by the same label. The labels can be considered as cell identifiers. From each segmented image we calculate cell volumes, cell centres, contact surfaces between neighbouring cells, as well as adjacency graph of cells. Adjacency graph is a graph in which vertices correspond to cells. Edges correspond to cell walls and connect cells to their neighbours.

| | $T_1$ | $T_2$ | $T_3$ | $T_4$ |
|---|---|---|---|---|
| # cells ($N_i$) | 358 | 445 | 583 | 1237 |
| # cell walls | 2262 | 2865 | 3887 | 8230 |

Table 5.1: The number of cells and number of cell walls in the 4D digitized meristem. Our 4D digitized meristem consists of four tissues tracking the evolution of meristem over 70 hours, each tissue $T_i$ has $N_i$ cells, $1 \leq i \leq 4$.

The volume of a cell can be simply calculated by counting voxels labelled by the cell identifier, while center of a cell corresponds to the center of mass of the voxels. Adjacent voxels in a segmented image with different labels indicate neighbouring cells and are used to construct the adjacency graph and to calculate the contact surfaces. We call all this information put together a *tissue*.

More precisely, from a segmented image $I_i$, we extract a tissue $T_i = (G_i, V_i, C_i, A_i), 1 \leq i \leq m$, where $m$ is the number of segmented images, $G_i$, $V_i$, $C_i$ and $A_i$ are adjacency graph, list of cell volumes, list of cell centres and list of contact surfaces between neighbouring cells respectively. $G_i = (S_i, E_i)$ where $S_i = \{s_{i1}, ..., s_{iN_i}\}$ is the set of cells, $N_i = |S_i|$, $|S_i|$ is the number of cells of tissue $T_i$. $e = (s_{ij}, s_{ih}) \in E_i$ if $s_{ij}$ and $s_{ih}$ are two neighbouring cells, $1 \leq j, h \leq N_i$.

Hereafter the ordered pair $((T_1, ..., T_m), (L_1, ..., L_{m-1}))$, where $T_i$ is the $i$th tissue and $L_i$ is the lineage between $T_i$ and $T_{i+1}$ is called 4D digitized meristem.

### 5.2.3 Application to a floral meristem

Fernandez *et al.* used MARS on a floral meristem at stage 3 (see Chapter 1). Then from this segmented image we constructed a tissue structure of the form described in the previous section. It includes 2124 cells and in the following it will be referred to as *static* tissue.

They also applied the whole pipeline, MARS-ALT, to a young floral meristem. The meristem was imaged from different angles every 24 hours up to 4 days [Fernandez *et al.*, 2010]. Then the images were segmented and the cell lineages were tracked by ALT. This resulted in four segmented images and cell lineages $(L_1, L_2, L_3)$. Again from the segmented images we extracted the tissues $T_1, T_2, T_3, T_4$. This series of tissues together with the lineages (i.e. $((T_1, T_2, T_3, T_4), (L_1, L_2, L_3))$) is the 4D digitized meristem that we use in this chapter, see Table 5.1 for more information about the number of cells and number of cell walls in our 4D meristem.

### 5.2.4 Target pattern on 4D digitized meristem

In principle, 4D digitized meristem can be obtained from meristems bearing reporter genes (e.g. GFP) hence providing us with gene expression patterns. Such patterns can also be defined manually by biologists on digitized tissues. However, measuring gene expression levels remains

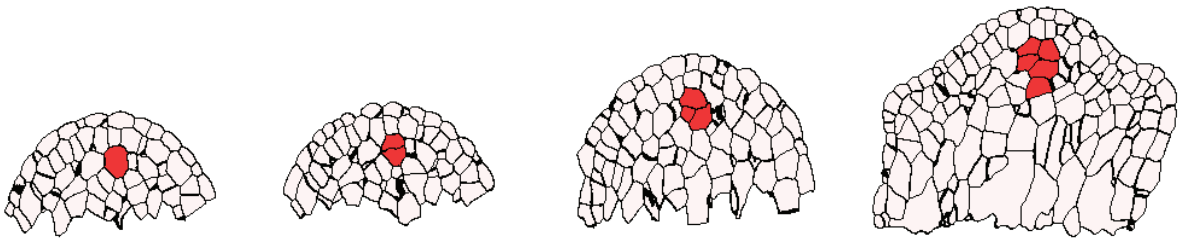


Figure 5.4: Slice images of the WUS target pattern defined by an expert on a cross section of the tissues $T_1, T_2, T_3, T_4$ (from left to right). Cells expressing WUS are marked by red color. The sets of cell indexes in WUS expression zone are $\{242\}, \{133, 286\}, \{173, 234, 558\}, \{130, 217, 221, 498, 612, 674, 993\}$. Note that the illustration is two dimensional, therefore all the cells in WUS pattern are not always visible.

very difficult. Therefore, our data about gene expression patterns is qualitative. Given such a qualitative data, one might try to use it to develop and explore hypotheses. For example, we can test whether a given model of gene regulatory network can generate a qualitative target pattern.

In this chapter we focus on WUS expression pattern. As our 4D digitized meristem did not include WUS pattern, this zone was defined by a biologist expert (Pradeep Das, ENS Lyon), on the last tissue. Then using the lineages $L_i$, $1 \leq i \leq 4$, WUS patterns were defined on the preceding tissues (see supplementary materials of [Fernandez *et al.*, 2010]). The WUS pattern on each tissue corresponds to a set of cell indexes. Each set is defined for a tissue and consists of cell indexes that are in the WUS expression zone of the tissue, see Figure 5.4 for an illustration of the target patterns.

Let $Q_i$ be the set of cell indexes which are in the WUS target pattern of tissue $T_i$, $1 \leq i \leq 4$. We define the target pattern $\hat{W}^i \in \{0,1\}^{N_i}$ for tissue $T_i$ as follows:

$$\begin{cases} \hat{W}^i_k = 1 \text{ if } k \in Q_i, \\ \hat{W}^i_k = 0 \text{ otherwise } . \end{cases}$$

We define the global, or 4D, target pattern $\hat{W}$ as $\hat{W} = (\hat{W}^1, \hat{W}^2, \hat{W}^3, \hat{W}^4)$.

In addition, for each tissue $T_i$ we define two sets of cells at the outermost layer of meristem: L1 layer and set of cells on the stem. To do this, we automatically detected all the cells that were in contact with the exterior of the meristem, using the adjacency graph. Among these cells, we then selected those whose centres were above (respectively below) a manually defined altitude. Finally, the identified cells were manually checked by an expert to remove potential errors, see Figures 5.5 and 5.6.

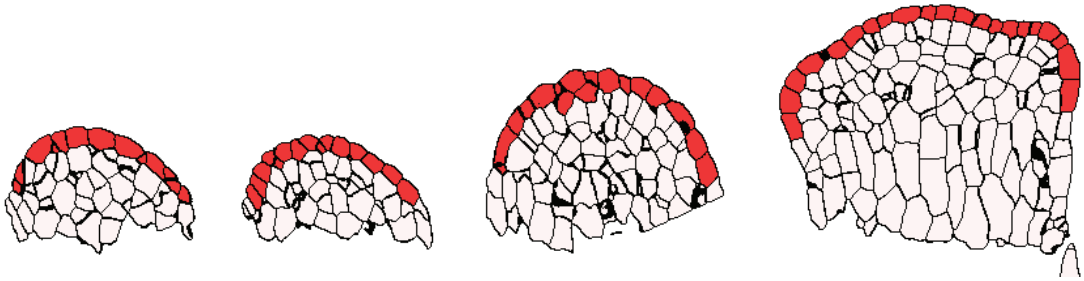


Figure 5.5: Illustration of L1 layers on a cross section of the tissues $T_1, T_2, T_3, T_4$ (from left to right). Cells belonging to L1 layer are marked by red color.

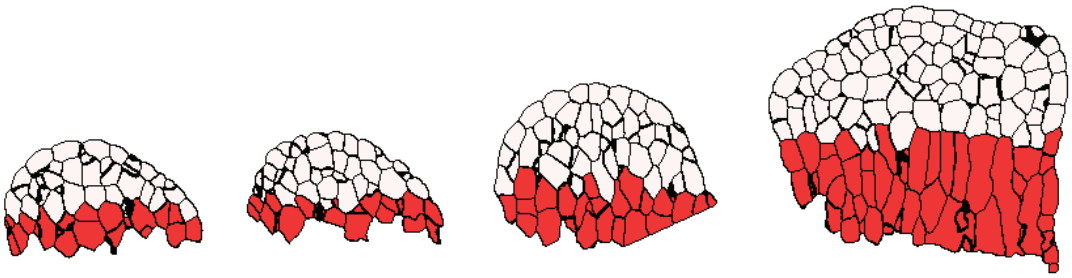


Figure 5.6: Illustration of cells on the stem on a cross section of the tissues $T_1, T_2, T_3, T_4$ (from left to right). Cells on the stem are marked by red color.

## 5.3 Parameter values optimisation of ordinary differential equation-based gene regulatory networks

Given a gene regulatory network described by a set of differential equations and a spatial pattern of gene expression domain, i.e. a target pattern, the question is whether or not there exists a set of parameters that can reproduce the target pattern. This can be formalised as an optimisation problem.

Here we present a general parameter values optimisation method for a model described by ordinary differential equations and a target pattern defined on a 4D digitized meristem. We try to identify parameter values that generate the target pattern, or the closest pattern to the target pattern found numerically.

Let $((T_1, ..., T_m), (L_1, ..., L_{m-1}))$ be a 4D digitized tissue and the tissue $T_i$ be composed of $N_i$ cells. For tissue $T_i$, $1 \leq i \leq m$, let the dynamics be described by:

$$\begin{cases} \dfrac{dX^i}{dt} = F^i(X^i; P^i), & (5.9) \\ X^i(0; P^i) = X_0^i, & (5.10) \end{cases}$$

where $X^i \in \mathbb{R}^{nN_i}$ represents the concentration of elements in cells, $n$ being the number of

elements (e.g. genes, hormones) in the gene regulatory network. Since the set of column vectors $\mathbb{R}^{nN_i}$ is isomorphic to the set of matrices $\mathbb{R}^{n\times N_i}$, we will use the notation $X^i_{u,k}$ to denote the concentration of $u$th element in cell $k$, for $1 \leq u \leq n$ and $1 \leq k \leq N_i$.

$X^i(0) = X^i(0; P^i)$ sets the initial condition and $P^i$ is the set of parameters of the model for tissue $T_i$.

Let us assume that for a set of parameters $P^i$ and initial condition $X^i_0$, the system reaches a stable steady state which we denote by $\tilde{X}^i(P^i)$.

Different approaches can be used to set the initial condition in Equation (5.10). A simple way to set initial conditions is to omit dilution by assuming that the initial concentration in daughter cells is the same as the steady state concentration in the mother cell, in the previous tissue.

However, cell volume change can be an important factor in chemical kinetics in growing and dividing cells [Pawlowski *et al.*, 2004] and therefore should be taken into account. One way to do this, is to suppose that when a mother cell divides the sum of molecules in daughter cells is equal to the molecules of the mother cell before division, and molecules are homogeneously distributed among daughter cells. This implies that all daughter cells have the same initial concentration. Formally, let $j$ be a cell index in tissue $T_i$, $1 \leq i \leq m-1$, the initial concentration of $u$th element of the daughter cell of index $k$ in tissue $T_{i+1}$ can be calculated by:

$$X^{i+1}_{u,k}(0; P^i) = \frac{\tilde{X}^i_j V^i_j}{\sum_{h\in L^j_i} V^{i+1}_h}, \forall k \in L^j_i, \tag{5.11}$$

where $V^i_j$ is volume of the cell of index $j$ in tissue $T_i$ and $L^j_i$ is the daughter cell indexes of the mother cell of index $j$. The Equation (5.11) can be obtained by supposing that the concentration of a gene in mother cell multiplied by its volume is equal to the sum of concentrations of the gene in daughter cells multiplied by their volumes:

$$\tilde{X}^i_{u,j} \times V^i_j = \sum_{h\in L^j_i} X^{i+1}_{u,h}(0; P^i) \times V^{i+1}_h,$$

where $\tilde{X}^i_{u,j}$ is the expression level of gene $u$ in mother cell $j$ in tissue $T_i$ at equilibrium, $V^i_j$ is the mother cell volume in tissue $T_i$, $X^{i+1}_{u,h}(0; P^i)$ is the initial concentration of daughter cell $h$ in tissue $T_{i+1}$ and $V^{i+1}_h$ is the daughter cell volume in tissue $T_{i+1}$.

We also suppose that the dynamics of each tissue is described by the same set of differential equations except the set of parameters and cell numbers that can change from one tissue to another. We rely on the assumption that growth dynamics is much slower than the interactions of a gene regulatory network such as diffusion or gene regulation process. This is supported by the fact that cell cycles in the SAM occur at a period approximately 12 to 24 hours whereas gene

regulation and diffusion occur on a time scale varying from few minutes to an hour [Alon, 2007, Reddy *et al.*, 2004].

At this point we presented the general form of models that we consider in this chapter, and we mentioned the underlying assumptions.

**Problem 5.1.** *Given a 4D digitized meristem* $((T_1, ..., T_m), (L_1, ..., L_{m-1}))$*, a model describing the dynamics by ordinary differential equations as in Equations (5.9) and (5.10), and* $\hat{X} = (\hat{X}^1, ..., \hat{X}^m)$ *as the 4D target pattern, find parameters such that* $\hat{X}$ *is an equilibrium pattern.*

Note that generally, it is possible that the model cannot reproduce $\hat{X}$ as an equilibrium pattern. Therefore, one might need to find an equilibrium pattern close to $\hat{X}$. To find the closest pattern, we first define the distance between $\hat{X}^i$ and $\tilde{X}^i(P^i)$ as the fitness function. The optimal parameter values generating closest pattern should minimise the fitness function:

$$P_i^{opt} = \arg\min_{P^i}(|\hat{X}^i - \tilde{X}^i(P^i)|), \tag{5.12}$$

where $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^{nN_i}$.

The above equation leads us to optimal parameter values for tissue $T_i$ to reproduce the target pattern, $\hat{X}^i$. It thus corresponds to a situation where tissue $T_i$ is considered on its own, independently from growth and other tissues.

The system can be optimised for all the tissues put together, called the 4D digitized meristem. Let $\hat{X} = (\hat{X}_1, ..., \hat{X}_m)$, $\hat{X} \in \mathbb{R}^{n(\sum_i N_i)}$, be the 4D target pattern, and $\tilde{X}(P)$ be the equilibrium of the system below:

$$\begin{cases} \dfrac{dX^1}{dt} = F^1(X^1; P), & (5.13) \\ \quad \vdots \qquad\qquad \vdots & \\ \dfrac{dX^m}{dt} = F^m(X^m; P), & (5.14) \\ (X^1, ..., X^m)(0) = (X_0^1, ..., X_0^m). & \end{cases}$$

The optimal parameters can be obtained for this system by:

$$P^{opt} = \arg\min_{P}(|\hat{X} - \tilde{X}(P)|), \tag{5.15}$$

where $|\cdot|$ denotes now the Euclidean norm in $\mathbb{R}^{n(\sum_i N_i)}$.

The difference between the Equations (5.15) and (5.12) is that in Equation (5.12) we fit the parameters for each tissue whereas in Equation (5.15) the parameters are fitted for all tissues put together. In other words Equation (5.12) corresponds to a situation where parameters are allowed to vary among the tissues which can be due to an unknown process. On the other hand,

the Equation (5.15) relies on the assumption that an identical set of parameters $P$ should be applied for all tissues.

In the following, for a given WUS target pattern and a WUS model whose dynamics is described by differential equations as in Equations (5.9) and (5.10) we will try to fit the parameters to generate the target pattern. We will consider both Equations (5.12) and (5.15) and we will try to find the optimal parameters that generate a pattern as close as possible to the WUS target pattern. The Equation (5.12) yields a set of parameters $P_i^{opt}$ for the tissue $T_i$, $1 \leq i \leq m$ while the Equation (5.15) yields a set of globally optimal parameters $P^{opt}$. Studying the set of optimal parameters for each individual tissue, i.e. $P_i^{opt}$, and the relation between optimal sets of successive tissues, i.e. $P_i^{opt}$ and $P_{i+1}^{opt}$, $1 \leq i < m$, can help us understand the evolution of the dynamics of the system.

## 5.4 Adimensionalizing the repressor model

As discussed at the end of sections 1.3.3 and 5.1, we focus in this chapter on the repressor model proposed by Jönsson *et al.* In this model, the expression of WUS is confined by two repressive signals diffusing from L1 layer and from the stem, represented by $Y$ and $S$ in Figure 5.2. The diffusive repressive signals can be considered as a simple way to study the role of meristem shape in WUS patterning. The gradients of the diffusive signals control the WUS expression domain and themselves are controlled by the geometry of L1 layer and the positions of cells on the stem. The gradients can represent the geometry of the meristem in its internal layers and inform the internal cells about their distance from L1 layer and the stem. For this reason, we chose the three dimensional repressor model to study the role of geometry of meristem in the apparition of WUS pattern on our 4D meristem.

Our initial system is a three-dimensional variant of the repressor model as described by Equations (5.7), and (5.8). Since we will consider realistic 3D tissues, we take also into account the repressing signal $S$ diffusing from the stem, as shown in Figure 5.1. In this simplified model we assume that the two signals $Y$ and $S$ repress the expression of WUS in the same way, thus in function $h$ a comparison of the sum of the signals $Y$ and $S$ with threshold $\theta$ defines the value of $h$ function:

$$\begin{cases} \dfrac{dW}{dt} = k_w h(S+Y,\theta) - \delta_w W, & (5.16) \\ \dfrac{dY}{dt} = k_y \mathbf{1}_{L1} - \delta_y Y + D_y \nabla^2 Y, & (5.17) \\ \dfrac{dS}{dt} = k_S \mathbf{1}_{stem} - \delta_S S + D_S \nabla^2 S, & (5.18) \end{cases}$$

where $h$ is a Heaviside function $h : \mathbb{R} \to [0,1]$:

$$h(x,\theta) = \begin{cases} 1, & x < \theta, \\ 0, & \text{otherwise.} \end{cases}$$

The target WUS pattern that we have is a vector of zeros and ones where ones indicate the cells within WUS expression domain. Therefore, we chose Heaviside function to model the interaction of WUS with the repressive signals.

To reduce the number of parameters, we adimensionalize the above model. Later in this chapter the method explained in Section 5.3 is used to optimize the parameter values for the WUS target pattern as explained in Section 5.2.4. The fewer the number of parameters is the shorter is the optimisation time.

By change of variable $\bar{W} = \frac{\delta_w}{k_w} W$, we obtain:

$$\frac{d\bar{W}}{dt} = \frac{\delta_w}{k_w}\frac{dW}{dt} = \frac{\delta_w}{k_w}(k_w h(Y+S,\theta) - \delta_w W) = \delta_w(h(Y+S,\theta) - \bar{W}).$$

Let $\tilde{W}$ be the value of $\bar{W}$ at equilibrium, from the above equations we know that $\tilde{W} = h(S+Y)$. Hence $\tilde{W}$ is a boolean vector.

In the same way, by change of variables $\bar{Y} = \frac{\delta_y}{k_y} Y$ and $\bar{S} = \frac{\delta_S}{k_S} S$ , we obtain:

$$\frac{d\bar{Y}}{dt} = \delta_y[\mathbf{1}_{L1} - Y] + \frac{\delta_y}{k_y} D_y \nabla^2 Y,$$

$$\frac{d\bar{S}}{dt} = \delta_S[\mathbf{1}_{stem} - S] + \frac{\delta_S}{k_s} D_S \nabla^2 S.$$

Since $\nabla^2$ is linear, see Equation (5.6), we have:

$$\frac{d\bar{Y}}{dt} = \delta_y[\mathbf{1}_{L1} - \bar{Y}] + D_y \nabla^2 \bar{Y},$$
$$\frac{d\bar{S}}{dt} = \delta_S[\mathbf{1}_{stem} - \bar{S}] + D_S \nabla^2 \bar{S}.$$

We can further simplify the system by change of time scale $\tau = \delta_y t$:

$$\frac{d\bar{Y}}{d\tau} = \frac{1}{\delta_y}\frac{d\bar{Y}}{dt} = \mathbf{1}_{L1} - \bar{Y} + \frac{D_y}{\delta_y}\nabla^2 Y,$$

$$\frac{d\bar{S}}{d\tau} = \frac{1}{\delta_y}\frac{d\bar{S}}{dt} = \frac{\delta_S}{\delta_y}(\mathbf{1}_{stem} - S) + \frac{D_S}{\delta_y}\nabla^2 S,$$

which leads us to define the following parameters:

$$\bar{D}_y = \frac{D_y}{\delta_y}, \bar{\delta_S} = \frac{\delta_S}{\delta_y}, \bar{D_S} = \frac{D_S}{\delta_y}.$$

By omitting the bars for simplicity and noting $t$ instead of $\tau$ we obtain

$$\begin{cases} \dfrac{dY}{dt} = \mathbf{1}_{L1} - Y + D_y \nabla^2 Y, & (5.19) \\ \dfrac{dS}{dt} = \delta_S(\mathbf{1}_{stem} - S) + D_S \nabla^2 S. & (5.20) \end{cases}$$

Let $L \doteq mat(\nabla^2)$ (i.e. the matrix defined by $\nabla^2$), for Equation (5.20), we have:

$$\frac{dS}{dt} = \delta_S \mathbf{1}_{stem} + (D_S L - \delta_S Id)S,$$

where $Id$ is the identity matrix. Note that by construction, the matrix $L$ has nonnegative extra-diagonal elements, and all its rows sum to zero. From this, we can deduce that for any positive $\delta_S$, the matrix $(-D_S L + \delta_S Id)$ is strictly diagonally dominant, and thus invertible and the system has a unique steady state. At steady state we can write:

$$\begin{aligned}\tilde{S} &= -(D_S L - \delta_S Id)^{-1}(\delta_S \mathbf{1}_{stem}),\\ &= \frac{1}{\delta_S}(Id - \frac{D_S L}{\delta_S})^{-1}(\delta_S \mathbf{1}_{stem}),\\ &= (Id - \frac{D_S L}{\delta_S})^{-1}\mathbf{1}_{stem},\end{aligned}$$

where $Id$ is the identity matrix of size of the number of cells in tissue. We define $\bar{D_S} = \frac{D_s}{\delta_S}$. Since the decay rate is 1 in Equation (5.19), we also have:

$$\tilde{Y} = (Id - \bar{D}_y L)^{-1}\mathbf{1}_{L1}.$$

By omitting the bars again, the system at steady state can be described by:

$$\begin{cases} \tilde{W} = h(\tilde{S} + \tilde{Y}, \theta), & (5.21)\\ \tilde{Y} = (Id - D_y L)^{-1}\mathbf{1}_{L1}, & (5.22)\\ \tilde{S} = (Id - D_S L)^{-1}\mathbf{1}_{stem}. & (5.23)\end{cases}$$

This steady state is independent of the initial conditions. Hence we do not have to consider Equation (5.11). An other convenient aspect is that the steady state of the adimensionalized model is controlled by only three parameters: $(D_y, D_S, \theta)$. They can be interpreted in intuitive terms as diffusive strengths of the signals from the L1 layer $(D_y)$ or the stem $(D_S)$, and as sensitivity of WUS to the repressive action of these signals $(\theta)$.

This steady state is also dependent on the shape of the zones where $Y$ and $S$ are produced, as materialized by the indicator functions, as well as the shape of the meristem itself, materialized by the Laplacian matrix $L$. These regions can be considered as geometric parameters. They can be given either as biological data, or set up manually on a tissue. Both cases will be considered in the following, in order to investigate both the theoretical properties of this model, and its biological implications when confronted with actual biological data.

## 5.5 Results and discussion

In this section we use the tissues described in section 5.2.3 as a support for simulations. Then to have a better intuition about the adimensionalized repressor model we study it on a row of

identical cells. We further study the WUS expression zone on a 4D digitized meristem. This allows us to take into account precise spatial information about the SAM in investigating the apparition of WUS expression zone.

### 5.5.1 Simulating three dimensional activator and repressor model on the static tissue

We first performed simulations on the static tissue described at the beginning of section 5.2.3, see Figures 5.7 and 5.8. The goal was to evaluate the feasibility of simulating gene regulatory networks on three-dimensional digitized tissue. As we explained in Section 5.2, our static tissue is extracted from segmented image obtained using MARS method. Thus spatial compartmentalization was at cellular scale and cellular growth and proliferation were discarded.

We implemented the repressor model using Equations (5.16)-(5.18). To test the numerical routines on a non-linear model, we also implemented the activator model using Equations (5.2)-(5.5) with an additional repressing signal from the stem.

The tissue consisted of 2124 cells leading to 10620 and 6372 equations differential equations to solve for repressor and activator models respectively. The code was developed in Python and we used a fifth-order Runge-Kutta solver with adaptive step size from Scipy library for the numerical integration of the ODEs.

Though in principle simulating the activator model on a three-dimensional meristem is possible, in practice this model was difficult to study; computation time varied according to the parameter values and adjusting the parameter values was difficult. The Brusselator model was initially chosen in [Jönsson *et al.*, 2005] because of its well-known ability to generate spot like patterns in 2D (see Chapter 2). However, it is also known that in 3D this model can generate much more complex patterns, looking like intermingled tubes (see also Chapter 2). Such patterns were an additional source of difficulty in the search for parameters, since they lead to lengthy computations and are by no means adequate to reproduce a plausible WUS domain.

For the repressor model the simulation time to reach equilibrium state was less than 1 minute. This short simulation time allowed us to adjust manually the parameters of the model. An optimisation of the parameter values has not been carried out for these simulations. One of the difficulties of adjusting the parameter values was to generate a connected and unique WUS pattern. As one might see in the first snapshots of Figure 5.7 the model generates a small WUS pattern in the right primordium and this zone disappears in the last snapshots, i.e. later in time. It was not easy to adjust the parameters to make this zone disappear at equilibrium. Almost all of the parameters created this zone at the equilibrium, see Figure 5.9.

As we explained before, in repressor model two repressive signals diffuse downward from L1 layer and upwards from stem. Their gradients inform the internal layers of meristem about the

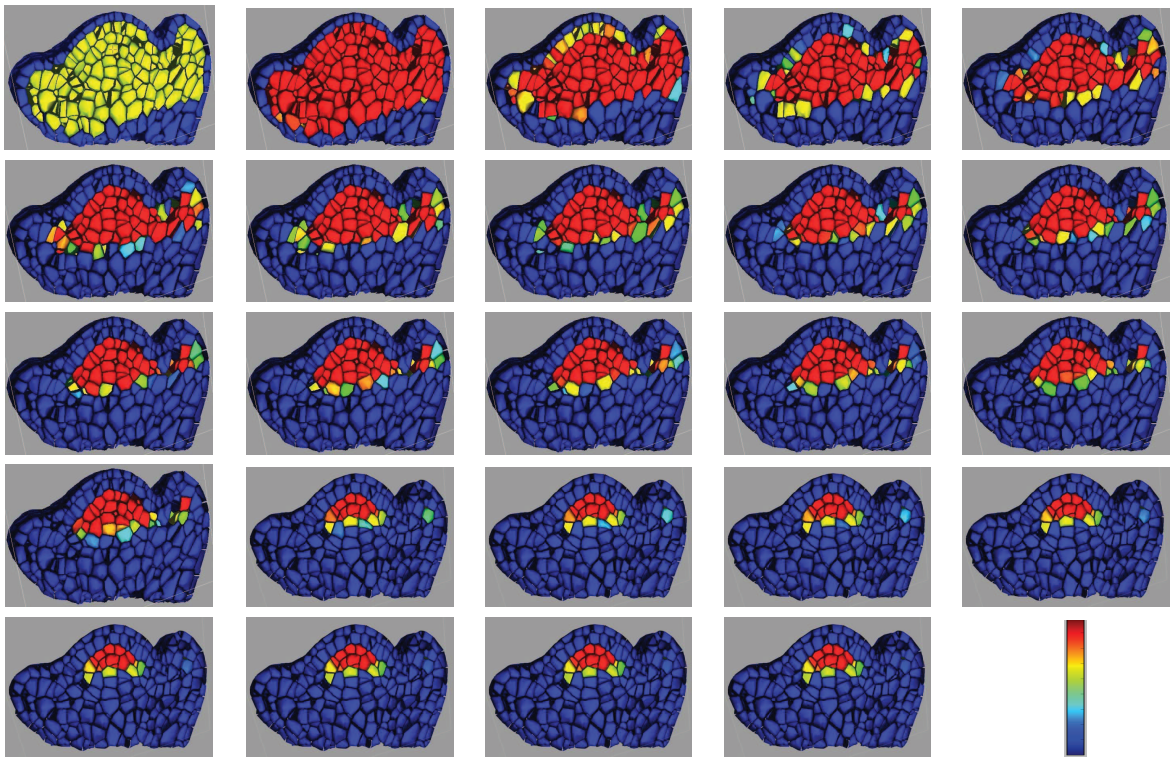


Figure 5.7: Time course simulation of the repressor model on a three-dimensional digitized meristem. The figures illustrate the relative concentration of WUS on a cross section of meristem. Simulation proceeded until equilibrium. The color coding is defined as blue(min), green, yellow, red(max). Parameter values: $k_y = 0.1$, $d_y = 0.008, D_y = 0.009, k_S = 0.1$, $d_S = 0.002, D_S = 0.052, T_w = 10.0, d_w = 0.1, h_w = 2.0, T_w y = -10.0$

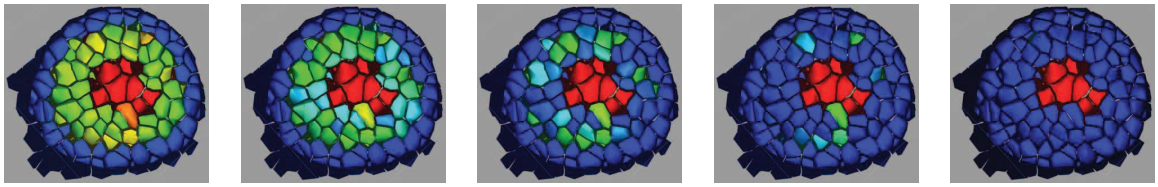


Figure 5.8: Time course simulation of repressor model, a horizontal view of meristem. From left to right simulations were developed until the equilibrium was reached. The color code is the same as in Figure 5.7.

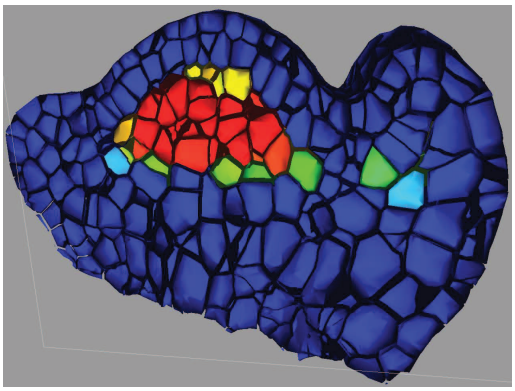


Figure 5.9: The effect of geometric shape in the zone of expression of WUS. The figure illustrates the equilibrium state. Most of the parameter values created a small group of cells within right primordium with relatively high WUS concentration (in this figure green coloured cell). As one might notice the L1 originating signal represses the WUS expression in L1 layer and in the cells just below the L1 layer (L2 layer). The size and shape of L1 and L2 cells in the right primordium induce a zone that is further from the periphery than the central zone of the main meristem. The color code is the same as in Figure 5.7.

geometry of the external layers. Therefore, this model can be used as a simple way to study the role of meristem geometry in WUS expression domain. This property of repressor model as well as the WUS pattern generated in simulations inside right primordium led us to the hypothesis that *geometrical shape of meristem can control the WUS expression domain.* Indeed, even though the right bump in Figure 5.9 actually corresponds to a sepal, it is geometrically very similar to the first stages of an emerging meristem.

During the last decade the role of regulatory networks in controlling morphogenesis and shape has been largely studied. Here we suggest that shape can also have a decisive role in controlling the WUS pattern without ruling out the role of interactions between WUS and other genes and plant hormones. This hypothesis is further explored in the following by taking into account precise geometrical information about the SAM using the 4D digitized meristem.

### 5.5.2 WUS pattern on a row of cells

In Section 5.4 we adimensionalized the repressor model to reduce the number of parameters. To have a better intuition about this model, we first carry out a mathematical study of the model at steady state on a simple system consisting of a row of $N$ identical cells.

#### Description of the system

In this configuration, the leftmost and rightmost cells of the row are supposed to be on L1 layer

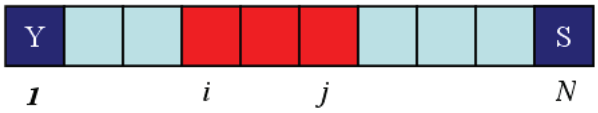



Figure 5.10: The adimentionalized model on a row of cells. The two repressive signals $Y$ and $S$ emanate from the leftmost and rightmost cells. Cells marked by red color create the expression zone of WUS. $\bar{W}_{i,j}$ denotes the WUS pattern, $i$ and $j$ being the indexes of the leftmost and right most cell indexes in WUS pattern respectively.

and on the stem respectively. Thus the two repressive signals $Y$ and $S$ originate from the leftmost and rightmost cells respectively, see Figure 5.10.

The adimensionalized repressor model at steady state as presented in Section 5.4 is described by:

$$\begin{cases} \tilde{W} = h(\tilde{S} + \tilde{Y}, \theta), & (5.24) \\ \tilde{Y} = (Id - D_y L)^{-1} \mathbf{1}_{L1}, & (5.25) \\ \tilde{S} = (Id - D_S L)^{-1} \mathbf{1}_{stem}. & (5.26) \end{cases}$$

For this system of cells in row, the functions $\mathbf{1}_{L1}$ and $\mathbf{1}_{stem}$ in above equations are:

$$\begin{cases} \mathbf{1}_{L1}(i) = 1 \text{ if } i = 1, \\ \mathbf{1}_{L1}(i) = 0 \text{ otherwise.} \end{cases}$$

$$\begin{cases} \mathbf{1}_{stem}(i) = 1 \text{ if } i = N, \\ \mathbf{1}_{stem}(i) = 0 \text{ otherwise,} \end{cases}$$

where $i$ is cell index.

We assume that for this configuration:

$$\begin{cases} \dfrac{A_{ij}}{d_{ij} V_i} = 1, \ 1 \leq i \leq N, j \in \mathcal{N}_i, \\ \dfrac{A_{ij}}{d_{ij} V_i} = 0, \ 1 \leq i \leq N, j \notin \mathcal{N}_i, \end{cases}$$

where $\mathcal{N}_i = \{i-1, i+1\}$, $2 \leq i \leq N-1$, $\mathcal{N}_1 = \{2\}$, $\mathcal{N}_N = \{N-1\}$, $\frac{A_{ij}}{d_{ij} V_i}$ being defined in Equation (5.1). Thus the matrix $L_{N \times N}$ in Equations (5.25) and (5.26) is:

$$L_{N \times N} = \begin{pmatrix} -1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & -2 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \ldots & 0 \\ \vdots & & & & & & \vdots \\ 0 & & \ldots & & 0 & 1 & -1 \end{pmatrix}.$$

**Studying WUS pattern**

Let $\Psi$ be a function that associates the WUS pattern at equilibrium to each set of parameter values,

$$\Psi : \mathbb{R}^3 \to \{0,1\}^N$$
$$(D_y, D_S, \theta) \mapsto \tilde{W},$$

and $\tilde{Y}_i$, $\tilde{S}_i$, and $\tilde{W}_i$ denote the concentration of $Y$, $S$, WUS in cell of index $i$, $1 \leq i \leq N$ at equilibrium respectively. The two repressive signals, $Y$ and $S$, are diffusing from cells of indexes 1 and $N$ respectively, therefore we have $\tilde{Y}_i < \tilde{Y}_j$ and $\tilde{S}_i > \tilde{S}_j$, $i < j$. Consequently only a subset of $\{0,1\}^N$ can be generated by the system as WUS pattern and the cells in WUS pattern must consist of consecutive and contiguous cells. As an example in this section, we considered a row of $N = 10$ cells, see Figure 5.11.

More precisely for a given $(D_y, D_S, \theta)$ the zone of expression of WUS can be delimited by the cells of indexes $i$ and $j$ such that:

$$\begin{cases} \tilde{W}_k = 1 \text{ if } i \leq k \leq j, \\ \tilde{W}_k = 0 \text{ otherwise }, \end{cases} \tag{5.27}$$

where $0 \leq i \leq j \leq N$. We denote this pattern by $\bar{W}_{i,j}$ , see Figure 5.10. $W_{0,0}$ notes the empty pattern, i.e. $\tilde{W} = (0, ..., 0)$.

From the above observation, $\forall (D_y, D_S, \theta) \in \mathbb{R}^3$, $\exists i, j$, $0 \leq i \leq j \leq N$, such that:

$$\Psi(D_y, D_S, \theta) = \bar{W}_{i,j}.$$

**Property 5.1.** *Let $M$ be the set of patterns that can be generated by the system on a row of $N$ identical cells for any value of $(D_y, D_S, \theta)$, we have:*

$$|M| = \frac{N(N+1)}{2} + 1.$$

*Proof.* Let $i$ be the leftmost cell index in a non empty WUS pattern, and $j$ be the rightmost cell index in WUS pattern. From Equation (5.27) we know that $j > i$. $j$ can take $N - i + 1$ values. The total number of such $i$ and $j$ defining the WUS pattern is obtained by summing up $N - i + 1$ for all values of $i$. We add 1 to take into account the empty WUS pattern, $W_{0,0}$ :

$$\sum_{i=1}^{N} (N + 1 - i) + 1 = \frac{N(N+1)}{2} + 1.$$

□

For $\bar{W}_{i,j} \in M$, $1 \leq i \leq j \leq N$, $\Psi^{-1}(\bar{W}_{i,j}) \subset \mathbb{R}^3$ is the set of all parameters that generate the pattern $\bar{W}_{i,j}$. From the Property 5.1 we know exactly the number of distinct patterns that can be reproduced. In particular this number is finite, whereas the parameter space is a continuum.

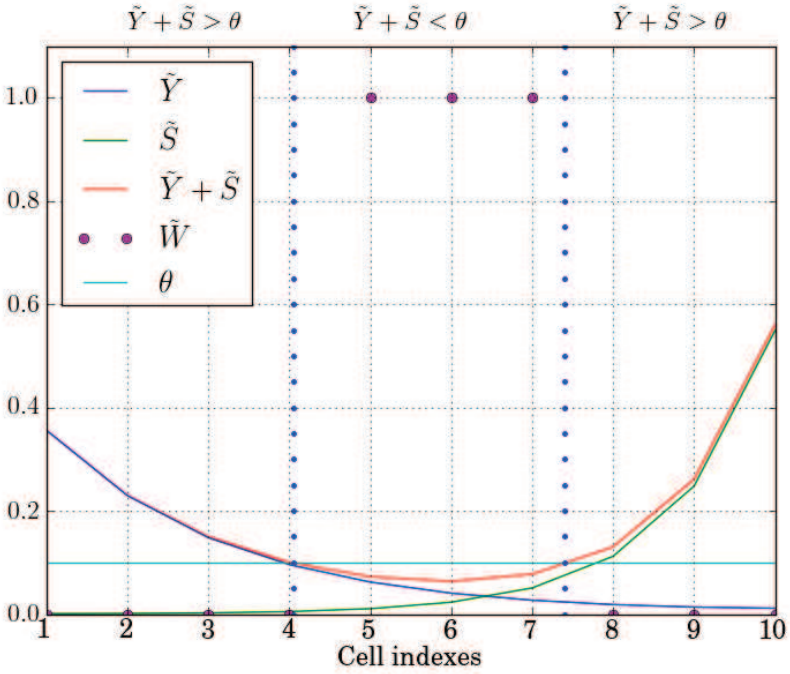



Figure 5.11: Simulation of adimensionalized model on a row of 10 cells. The concentrations of the repressive signals at equilibrium $\tilde{Y}$, $\tilde{S}$ and $\tilde{Y}+\tilde{S}$ are plotted. As we move forward from cell 1 to cell 10, the concentration of $Y$ reduces and as we move backwards from cell number 10 to cell number 1 the concentration of $S$ reduces. The WUS pattern is $\tilde{W}=(0,0,0,0,1,1,1,0,0,0)$. The parameter values are $D_y=5.1$, $D_S=1.5, \theta=0.1$.

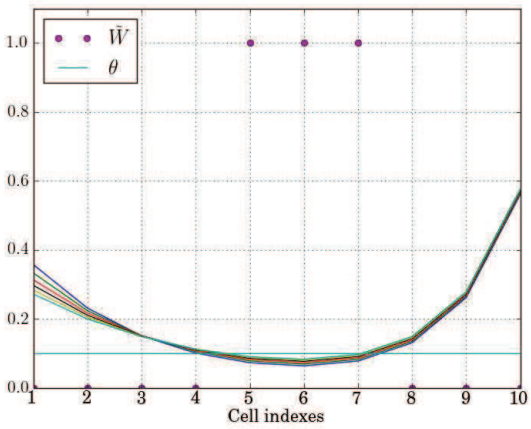



Figure 5.12: The adimentionalized model in cells in line. For $D_y \in \{5.1, 6.1, 7.1, 8.1, 9.1, 10.1\}$ $\tilde{Y} + \tilde{S}$ is plotted. Cells in the WUS expression zone are marked by small circular magenta particles. Different values of $D_y$ generate the same WUSCHEL pattern. $D_S = 1.5, \theta = 0.1$

In consequence, a given WUS pattern, $\bar{W}_{i,j}$, can be generated by different set of parameters, i.e. $|\Psi^{-1}(\bar{W}_{i,j})| \geq 1$. In fact, these inverse images are regions in parameter space that contain an infinite number of points. They form a partition of parameter space into domains of parameter values that lead to an identical WUS pattern.

For our 10 cell example, we took a point in parameter values space, and calculated the corresponding WUS pattern using the Equations (5.24), (5.25), and (5.26). We then increased the value of $D_y$ by fixing $\theta$ and $D_S$. At some critical values, the increase in $D_y$ generated new WUS patterns that were different from the previous pattern by one cell, see 5.14. For an other point of view, see also Figure 5.12, where the profile of the $\tilde{Y} + \tilde{S}$ repressing signal is modified by variations of $D_y$, without modification of $\tilde{W}$. Similarly, varying the threshold $\theta$ in Figure 5.13 illustrates the robustness of the $\tilde{W}$ with respect to this parameter. Hence for each parameter, when the two others are fixed, there is a whole interval of values that lead to the same $\tilde{W}$ pattern, and contiguous intervals correspond to patterns that typically differ by one cell only.

The intervals of parameters just discussed could have been anticipated on an intuitive basis. Moving to the full parameter space, the shape and organization of regions associated to the different $\tilde{W}$ patterns is more difficult to foresee. In particular, an analytical description would be difficult to achieve, since the steady state patterns $\tilde{Y}$ and $\tilde{S}$ themselves do not have a simple expression in terms of the parameters $(D_y, D_S, \theta)$. Indeed, the set of parameter values leading

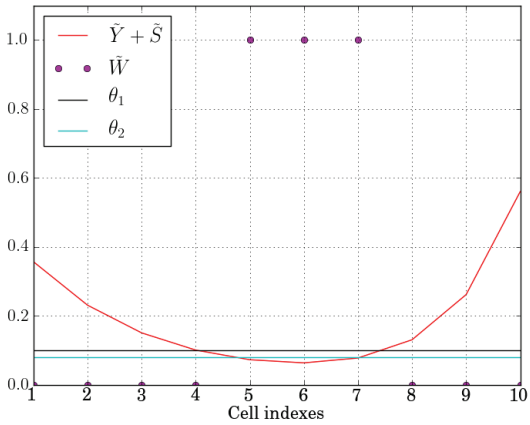



Figure 5.13: The adimentionalized model in cells in line. $\tilde{Y}+\tilde{S}$ and $\tilde{W}$ are plotted. Cells in the WUS expression zone are marked by small circular magenta particles. Different values of $\theta \in \{0.08, 0.1\}$ generate the same WUS pattern.

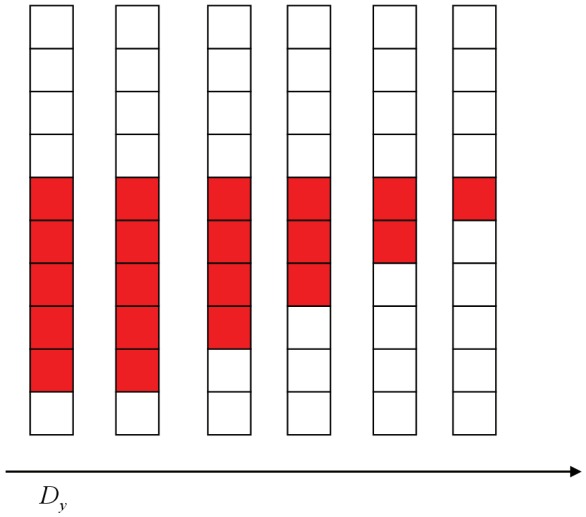


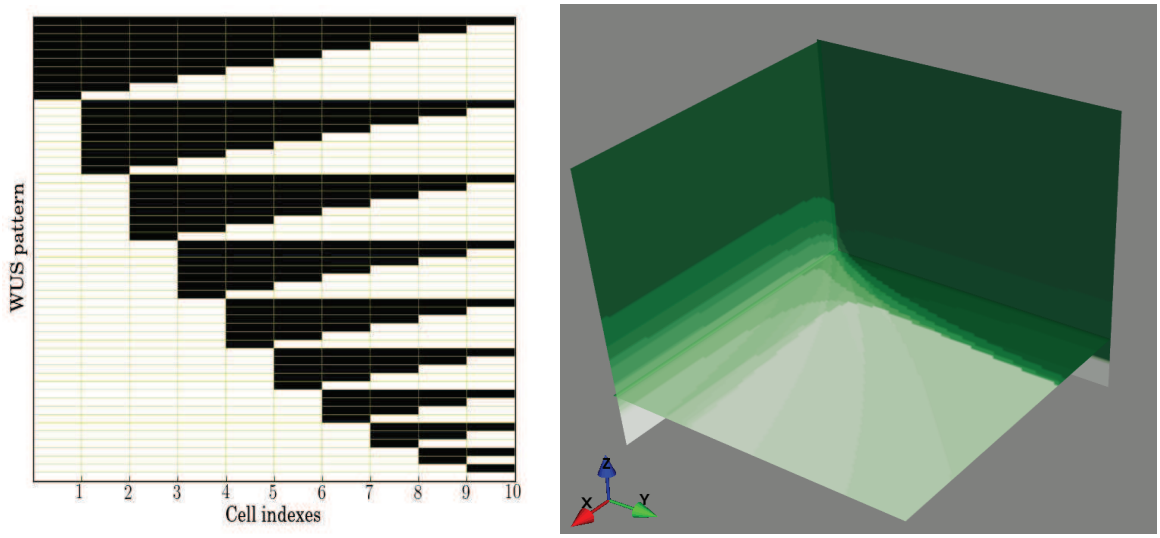

Figure 5.14: The patterns created by varying $D_y$. The $\theta$ and $D_S$ are fixed. As $D_y$ increases the WUS pattern changes to a pattern which is different from the previous one by one cell. $D_y \in \{0.08, 0.1, 0.4, 0.8, 2.8, 3.5\}$, $D_S = 7.0$, $\theta = 0.1$.

to a given $\bar{W}_{i,j}$ pattern is defined by the following list of $N$ inequalities:

$$\begin{cases} \tilde{Y}_k(D_y) + \tilde{S}_k(D_S) > \theta & \text{if } i \leq k \leq j \\ \tilde{Y}_k(D_y) + \tilde{S}_k(D_S) < \theta & \text{otherwise ,} \end{cases} \tag{5.28}$$

where from Equations (5.25) and (5.26) one can deduce that each term $\tilde{Y}_k(D_y,)$ (resp. $\tilde{S}_k(D_S)$) can be expressed as a rational function of $D_y$ (resp. $D_S$), whose precise form follows from the matrix inversion appearing in these equations, and is quite cumbersome.

This led us to a numerical exploration, where we considered a grid of parameter values $(D_y, D_S, \theta)$ that was fine enough to obtain all the $(10 \times 11)/2 + 1 = 56$ possible WUS patterns for a row of 10 cells. This exhaustive search allows us to have a visual representation of the decomposition of parameter space into regions of the form $\Psi^{-1}(\bar{W}_{i,j})$, for all admissible $\bar{W}_{i,j}$, see Figure 5.15.




(a) WUS patterns. (b) Planar cuts of parameter space

Figure 5.15: (a) All 56 WUS patterns generated for a row of 10 cells are illustrated. The WUS pattern is marked by black cells. (b) For each point in parameter values space, a colour is associated that corresponds to one of the 56 distinct WUS patterns. $x$-axis, $y$-axis, and $z$-axis correspond to $D_y$, $D_S$, and $\theta$ parameters, respectively.

This picture only gives a partial view of the parameter space partition induced by our model, which is uneasy to visualize because it is three dimensional. The two dimensional slice shown in Figure 5.16 illustrates the symmetry in the system, which is due geometrically to the symmetry of the row of cells we have used, and to the invariance of Equations (5.25) and (5.26) under permutation of $Y$ and $S$. In more physical terms, as $D_y$ increases the leftmost cell of the WUS

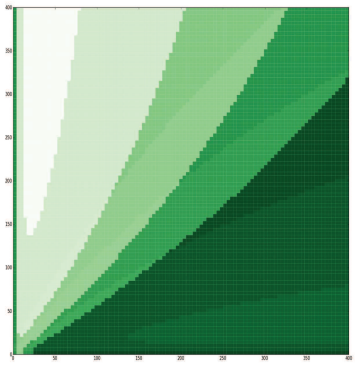


Figure 5.16: Slice of the parameter space partition shown in Figure 5.15 (b), for $\theta = 0.2$.

domain shifts to the right of the row of cell, whereas increasing $D_S$ shifts the rightmost cell to the left. Then, WUS patterns that are symmetrical with respect to the center of the row of cells are obtained by exchanging these to diffusion parameters. This question could lead to interesting considerations, of a more involved mathematical nature. However, even though the symmetry in the equations remains true on any tissue, the biological tissues that will be considered hereafter do not present the type of symmetry that is present in this artificial situation of a row of cells. Hence, we leave this as a remark, and maybe a suggestion for future work.

Overall, one main conclusion from this study is that the simple three parameter model we have chosen is able to produce a variety of WUS patterns, that reflect the geometric properties of the tissue. Each of these WUS patterns is in fact generated by a whole set of parameter values, in such a way that the parameter space is partitioned into "iso-WUS" regions, where neighbouring regions of this form induce close WUS patterns. In a continuous dynamical systems framework this would have come as no surprise, but here $\tilde{W}$ is in fact boolean valued, and thus cannot be continuous (a function with finite valued range is continuous if and only if it is constant). Although intuitive, it was thus worth confirming this fact on a simple example: to create similar close patterns to a given pattern $\tilde{W}$, one can explore the points around the parameter values in $\Psi^{-1}(\tilde{W})$. Note however that this remark can be used in a heuristic manner, but cannot be generalized straightforwardly to a series of tissues with different numbers of cells, as the one to be dealt with in the next section.

### 5.5.3 Simulating the adimensionalized model on a 4D digitized meristem

To further explore the hypothesis that the geometry of the SAM determines the WUS expression zone, we study the steady state of adimensionalized repressor model on the 4D digitized meristem that we explained in Section 5.2. The 4D digitized meristem provides us precise spatial

information about the meristem. Thereby we can investigate whether by considering the exact geometrical information, the repressor model can reproduce the WUS target pattern explained in Section 5.2.4.

**Equilibrium state of the system**

Our 4D digitized meristem is composed of four tissues $(T_1, T_2, T_3, T_4)$ and the cell lineages $(L_1, L_2, L_3)$. Given a set of three parameter values $P = \{D_y, D_S, \theta\}$, we can calculate the equilibrium of the system in tissue $T_i$ using the Equations (5.21), (5.22) and (5.23):

$$\begin{cases} \tilde{W}^i = h(\tilde{S}^i + \tilde{Y}^i, \theta) \\ \tilde{Y}^i = (Id - D_y L^i)^{-1} \mathbf{1}^i_{L1} \\ \tilde{S}^i = (Id - D_S L^i)^{-1} \mathbf{1}^i_{stem}, \end{cases}$$

where $\mathbf{1}^i_{L1}$ and $\mathbf{1}^i_{stem}$ are the indicator functions defining cells in L1 layer and cells situated on the stem respectively for tissue $T_i$, $1 \leq i \leq 4$, see Figures 5.5 and 5.6. To indicate the set of parameters used in calculating the equilibrium, $P = \{D_y, D_S, \theta\}$, we denote the equilibrium state of tissue $T_i$ by $\tilde{W}^i(P)$, $\tilde{Y}^i(P)$ and $\tilde{S}^i(P)$.

The global equilibrium state of the system, i.e. for all tissues put together $(T_1, T_2, T_3, T_4)$ for a given set $P = \{D_y, D_S, \theta\}$ is defined by:

$$\begin{cases} \tilde{W}(P) = \left(\tilde{W}^1(P), \tilde{W}^2(P), \tilde{W}^3(P), \tilde{W}^4(P)\right), & (5.29) \\ \tilde{Y}(P) = \left(\tilde{Y}^1(P), \tilde{Y}^2(P), \tilde{Y}^3(P), \tilde{Y}^4(P)\right), & (5.30) \\ \tilde{S}(P) = \left(\tilde{S}^1(P), \tilde{S}^2(P), \tilde{S}^3(P), \tilde{S}^4(P)\right). & (5.31) \end{cases}$$

Note that in calculating the global equilibrium state of system as above, we impose the same set of parameter values $P = \{D_y, D_S, \theta\}$ for all tissues.

An exhaustive analytical study of equilibrium states, i.e. patterns that can be obtained for our 4D digitized meristem is not possible. The system on our 4D digitized meristem is much more complicated than the row of identical cells because of the complex geometrical aspects and the number of cells. Therefore, for such systems, computer simulations is often the only way to understand the behaviour of the model.

**Fitting the parameters to obtain target pattern**

To obtain the closest WUS pattern to the defined target pattern we used the method described in Section 5.3. We used both Equations (5.12) and (5.15) to fit the parameter values for each individual tissue and all tissues put together (i.e. global system):

$$P_i^{opt} = \arg \min_{P=\{D_y, D_S, \theta\}} (|\hat{W}^i - \tilde{W}^i(P)|) \quad (5.32)$$

| | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $(T_1, T_2, T_3, T_4)$ |
|---|---|---|---|---|---|
| # cells in $\hat{W}^i$ | 1 | 2 | 3 | 7 | 13 |
| $d_i$ | 0 | 1 | 0 | 1.732 | 3 |
| # different cells ($= {d_i}^2$) | 0 | 1 | 0 | 3 | 9 |
| # optimal parameters | 30 | 9 | 66 | 120 | 125 |
| # different WUS pattern | 1 | 1 | 1 | 2 | 2 |

Table 5.2: Parameter values optimisation on the 4D digitized meristem. $\{T_1, T_2, T_3, T_4\}$ indicates the global system. *# cells in* $\hat{W}^i$ indicates the number of WUS expressing cell in each target pattern. $d_i$ is the distance between the target pattern and the best solution. *# different cells* indicates the number of cells in target pattern and the closest pattern obtained by simulations with different WUS expression concentration. *# optimal parameters* indicates the number of different parameter values that created the closest pattern to the target pattern. *# different WUS pattern* is number of distinct WUS zones created by optimal parameters, see Figure 5.19.

$$P^{opt} = \arg \min_{P=\{D_y, D_S, \theta\}} (|\hat{W}^i - \tilde{W}(P)|) \tag{5.33}$$

using the above equations we obtain four sets of optimal parameters, $P_1^{opt}, P_2^{opt}, P_3^{opt}, P_4^{opt}$ for the tissues $T_1, T_2, T_3, T_4$. For the global system using the Equation (5.33) we obtain the set of parameters $P^{opt}$ see Figures 5.17 and 5.18.

To investigate how well the model can generate target pattern, we define:

$$d_i = |\hat{W}^i - \tilde{W}^i(P)| \tag{5.34}$$

where $1 \leq i \leq 4$ and $P \in P_i^{opt}$. $d_i$ is the Euclidean distance between the target pattern and the best possible pattern obtained by optimising the parameters for each tissue. Since the patterns are boolean, ${d_i}^2$ is in fact the number of cells in $\hat{W}^i$ and $\tilde{W}^i(P^{opt})$ with different WUS expression concentration.

Similarly to Equation (5.34) we define:

$$d = |\hat{W} - \tilde{W}(P)|, \tag{5.35}$$

where $P \in P^{opt}$. $d$ is the Euclidean distance between the target pattern and the best possible pattern obtained for the global system.

To implement Equations (5.34) and (5.35), we tried different optimization routines, and finally chose the downhill simplex algorithm, which was the fastest and the most robust on our test sample. Namely, we used the implementation of this algorithm provided in the $fmin$ function from the Scipy library. To initialize the $fmin$ function we used a grid of 4913 points in parameter space $[1, 400] \times [1, 400] \times [0.1, 0.9]$ (17 values for each parameter) (i.e. 4913 initial

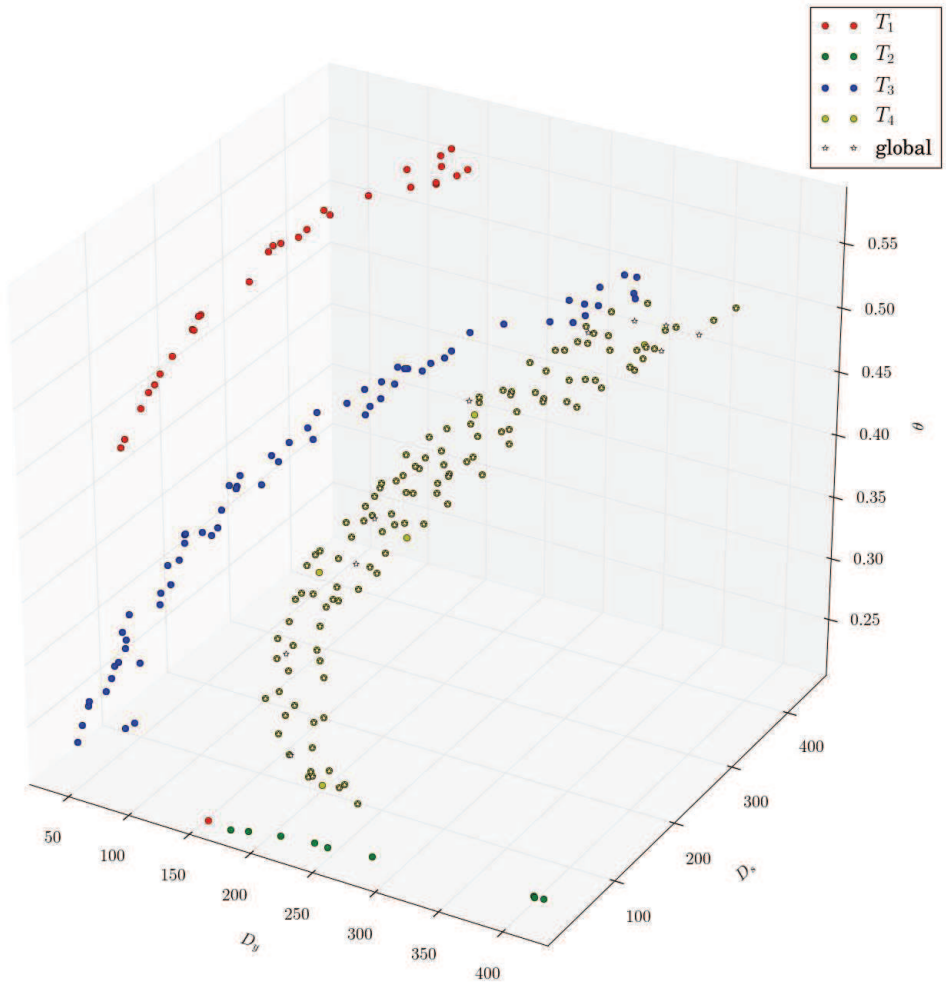



Figure 5.17: Optimal parameter values for the individual tissues and all tissues put together (i.e. *global* system). The global optimal parameter and optimal parameter values of $T_4$ are in the same zone and most of them are superposed, i.e. $P_4^{opt} \approx P^{opt}$ . As one might notice the parameter values for the tissues $T_1, T_3, T_4$ and *global* seem to create zones in space looking like bananas that somehow follow each other, except the parameter values of $T_2$. The shape of these objects could in principle be put in equations of the form (5.28).

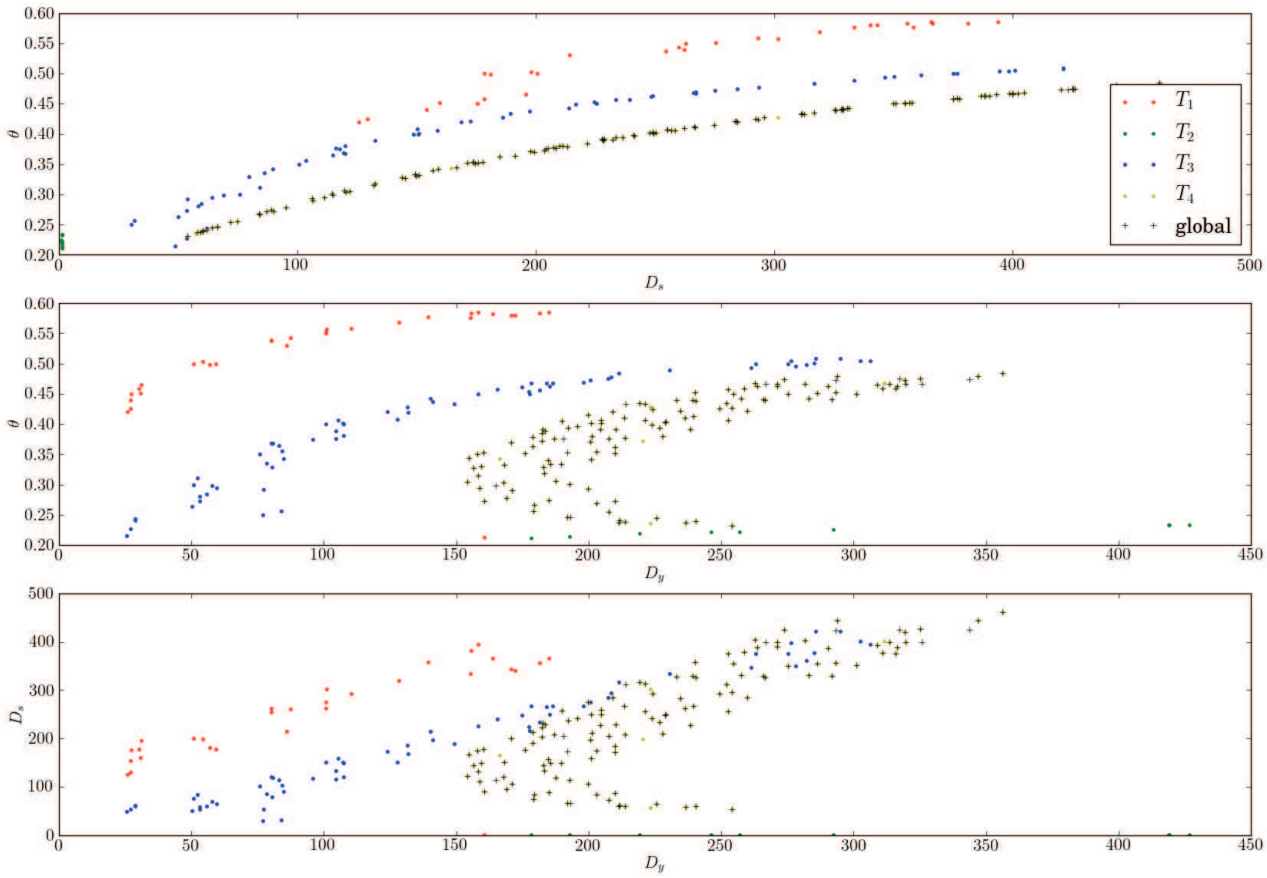



Figure 5.18: Optimal parameter values projected on 2D space. The optimal parameter values of the global system are superimposed on the optimal parameter values of tissue $T_4$, i.e. $P_4^{opt} \approx P^{opt}$. The zone of optimal parameter values of tissue $T_2$ does not look like the zones of optimal parameter values of $T_1, T_3, T_4$.

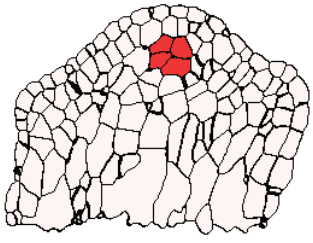


Figure 5.19: One of the two optimal patterns found by optimising the parameter values for tissues $T_3$. The cells in WUS pattern are marked by red color.

values of $(D_y, D_S, \theta) \in ([1,400] \times [1,400] \times [0.1, 0.9])$ are used to initialize the $fmin$ function). We used parallel computing on a computer with 8 processors. The complete optimization took about 48 hours.

Among the parameter values returned by $fmin$ function (4913 parameter values, one for each point in the grid), we selected the parameter values that created closest WUS expression zone to the target patterns. Note that different parameter values can generate the same WUS patterns see the Table 5.2 and Figure 5.15. As one might notice in the table, despite the simplicity of the model, for the tissues $T_1$ and $T_3$, we managed to accurately fit the parameter values, and reproduce the target pattern exactly. The distances $d_i$ $(1 \leq i \leq 4)$ and $d$ calculated by Equations (5.34) and (5.35) respectively show that the globally optimal parameter values, $P^{opt}$, do not create the perfect WUS target pattern for the individual tissues.

We propose now an hypothesis that would explain the distinct optimal parameters found for the successive tissues: the parameters could in fact be slowly evolving from one snapshot to the following. In principle, the slow evolution of parameters could be described by the following equations:

$$\dot{D_y} = \epsilon H_1(\rho_y),$$

$$\dot{D_S} = \epsilon H_2(\rho_S),$$

$$\dot{\theta} = \epsilon H_3(\rho_\theta),$$

where the functions $H_1$, $H_2$ and $H_3$ and their arguments $\rho_y$, $\rho_S$ and $\rho_\theta$ are unknown (they can be related to the cellular growth and proliferation) and $\epsilon$ is infinitely small quantity. Since $\epsilon \approx 0$, we can suppose that $D_y$, $D_S$ and $\theta$ are constant for each tissue $T_i$, $1 \leq i \leq 4$. However, since $\epsilon$ is in fact non-zero, these parameter values may change from tissue $T_i$ to tissue $T_{i+1}$, $T_i$, $1 \leq i \leq 3$.

To investigate the evolution of dynamics of the system, we studied the sets of optimal parameter values, $P_1^{opt}$, $P_2^{opt}$, $P_3^{opt}$, $P_4^{opt}$, $P^{opt}$. We noticed $P_i^{opt} \cap P_j^{opt} = \emptyset$ and $P_i^{opt} \cap P^{opt} = \emptyset$, $1 \leq i, j \leq 3$ and $P_4^{opt} \approx P^{opt}$, see Figures 5.17 and 5.18. The optimal parameter values for the tissues $T_1, T_2, T_3, T_4$ seem to create zones in space looking like arcs except the optimal parameter values of $T_2$. At least visually it seems that $\theta$ reduces from one tissue to the next, whereas $D_S$ and $D_y$ seem to increase. To be fully consistent with a continuous evolution of parameters from one tissue to the next, we were led to propose alternative WUS target patterns for $T_2$. In our exhaustive study of WUS patterns on a row of cells we showed that the close patterns differing from each other by one cell are generated by close parameter values. This inspired us to interpolate parameter values for tissue $T_2$ from $P_1^{opt}$ and $P_3^{opt}$.

Let $p = (D_y, D_S, \theta) \in P_1^{opt}$ we define the closest point in $P_3^{opt}$

$$P_{1,3}(p) = \arg \min_{\mathcal{P} \in P_3^{opt}} (|p - \mathcal{P}|).$$

For $p' = (D'_y, D'_S, \theta') \in P_{1,3}$, we define the intermediate point between $p$ and $p'$:

$$p''_{1,3} = \lambda p + (1-\lambda)p' = \lambda(D_y, D_S, \theta) + (1-\lambda)(D'_y, D'_S, \theta'),$$

where $0 \leq \lambda \leq 1$. In the same way for $p = (D_y, D_S, \theta) \in P_3^{opt}$ we define

$$P_{3,1}(p) = \arg \min_{\mathcal{P} \in P_1^{opt}} (|p - \mathcal{P}|).$$

For $p' = (D'_y, D'_s, \theta') \in P_{3,1}$, we define the intermediate point between $p$ and $p'$:

$$p''_{3,1} = \lambda p + (1-\lambda)p' = \lambda(D_y, D_S, \theta) + (1-\lambda)(D'_y, D'_S, \theta').$$

We calculated interpolated parameter values using above equations for $\lambda \in \{0.1, 0.2, 0.3, ..., 0.9\}$ see Figure 5.20. Most of the interpolated parameter values create WUS patterns that are considerably different from the target pattern. Particularly they often generate WUS expression domains considerably bigger than the target pattern, see Figure 5.21. However, among the new set of interpolated points we detected parameter values (black squares in Figure 5.20) that created WUS expression domains that were close to the target pattern see Figures 5.22 and 5.23.

To identify without ambiguity cells expressing WUS on a three-dimensional meristem, we need confocal images where WUS expression is marked by reporter genes e.g. GFP. This was not available at the time of simulating the model. Therefore, WUS target pattern was defined by hand. The method that we described for interpolating parameter values can be used to propose candidate target patterns, given a pair of patterns defined on the previous and next tissue. This can also be a way to detect possible errors in defining the target pattern of the intermediate tissue, in situations like the one we dealt with in this section.

We managed to re-generate rather accurate WUS patterns by optimising the parameter values of the adimensionalized system. This validates partially our hypothesis that the WUS pattern can be controlled by the geometrical shape of meristem. By interpolating the optimal parameter values of tissues $T_1$ and $T_3$ we proposed candidate WUS target pattern for tissue $T_2$.

## 5.6 Conclusion

In this chapter we treated the WUS expression domain in shoot apical meristem. We presented the methods that allow us to create three dimensional digitized meristems and track cell lineages. We treated a general method to estimate the parameters of a model based on differential equations for a given target pattern. To investigate the WUS pattern, we began by implementing three-dimensional versions of the models proposed by Jönsson *et al.* [Jönsson *et al.*, 2005]. The simulations led us to hypothesize that the geometrical shape of the shoot apical meristem can

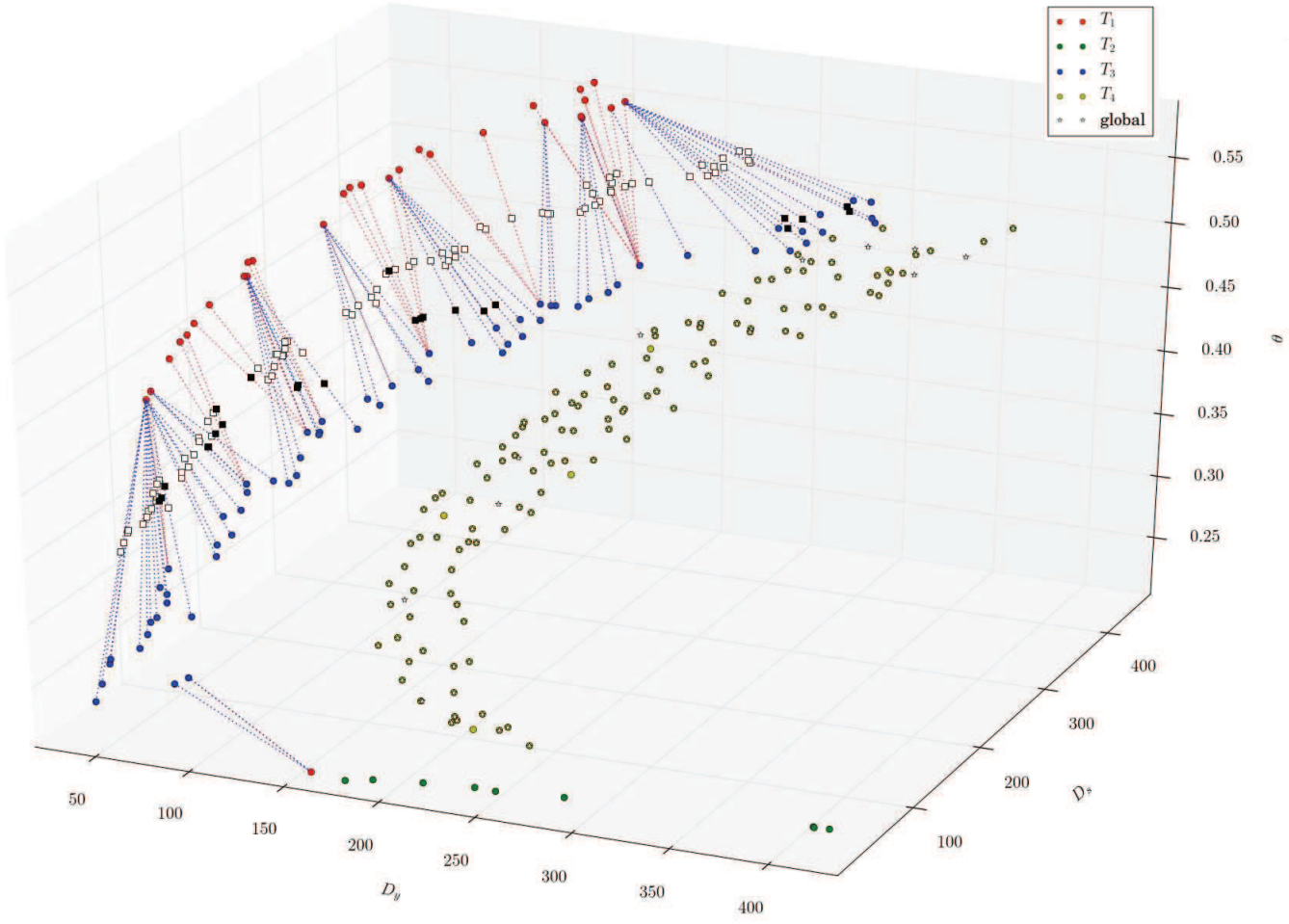



Figure 5.20: Interpolated parameter values for tissue $T_2$. Each point in optimal parameter values of $T_1$ is joined to its nearest point in optimal parameter values of $T_3$, whereas the blue dotted line joins optimal parameter values of tissue $T_3$ to its nearest point in optimal parameter values of $T_1$. The white squares indicate intermediate points for $\lambda = 0.5$ and black squares indicate intermediate points for $\lambda \in \{0.1, 0.2, 0.3, ..., 0.9\}$ and creating WUS patterns consisting of at most 5 cells.

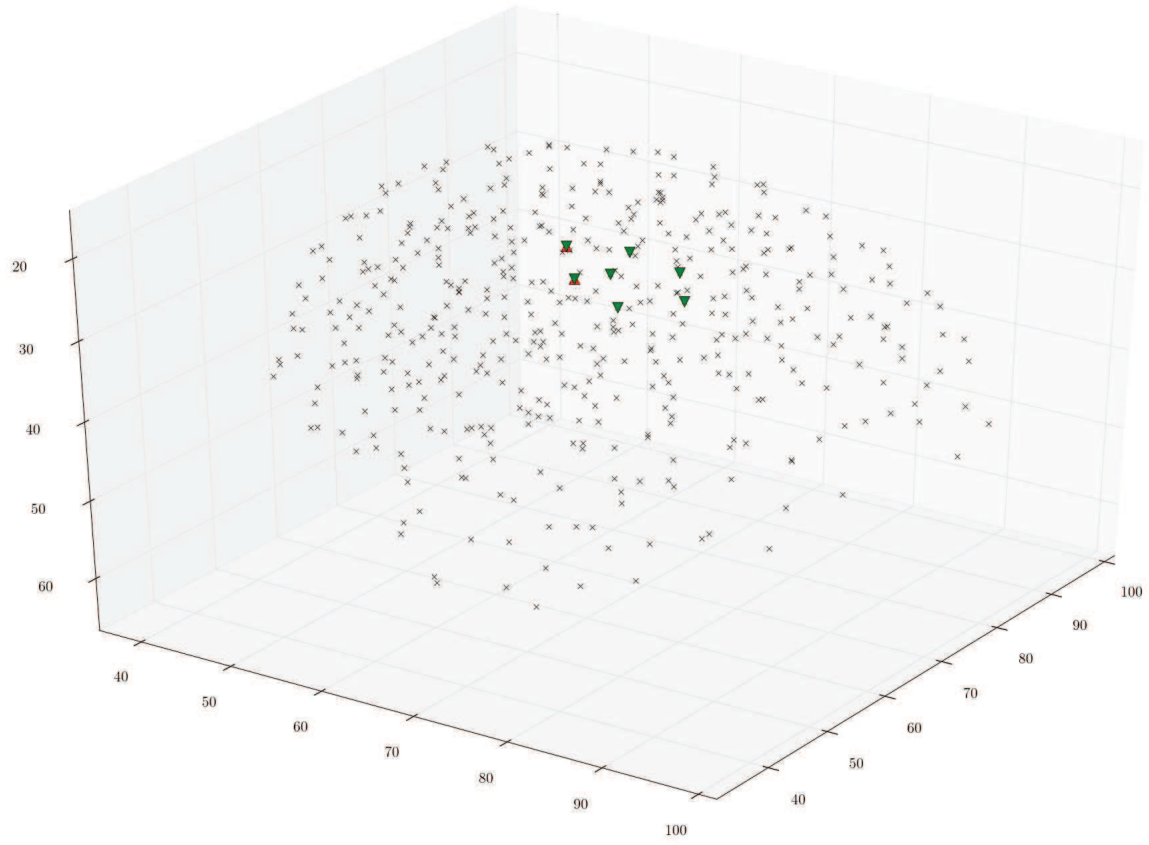



Figure 5.21: The WUS pattern created by interpolated parameter values for tissue $T_2$, $\lambda = 0.6$. Small $\times$ signs corresponds to cell centres in the tissue $T_2$. The cells centres belonging to manually defined target pattern are marked by red triangles. Green triangles correspond to the cell centres in interpolated WUS pattern. The interpolated pattern contains more cells than the target pattern.

(a) WUS target pattern of $T_1$.

(b) Interpolated WUS pattern of $T_1$.

Figure 5.22: (a) A cross section of WUS target pattern of $T_1$ defined manually. (b) A cross section of WUS pattern calculated for an interpolated set of parameters for $T_1$. The cells in WUS pattern are marked by red color. One might notice that the interpolated pattern is consisted of cells neighbouring the cells of the target pattern. See Figure 5.23 for a representation of cell centres of target pattern and interpolated pattern.

control the WUS expression domain. To further explore this hypothesis, we adimensionalized one of the models proposed by Jönsson *et al.* where there are two diffusive signals originating from L1 layer and cells situated on the stem. This can be a simple way of taking into account our hypothesis, since the gradients of the signals inform the internal layers of meristem about the geometry of external layers. To have a better intuition about the adimensionalized model we carried out a mathematical study of the repressor model on a row of cells. We then estimated the parameters of the adimensionalized model on a series of three dimensional digitized meristems corresponding to snapshots of a growing and dividing meristem. Strikingly we obtained rather accurate WUS expression domains by supposing that the parameters of the model slowly evolve over time. By interpolating the optimal parameter values we proposed candidate WUS target pattern different from the initial target pattern. The initial target pattern was manually defined for the last meristem. Then using the lineages, the WUS pattern was defined for the first three meristem. In other words we supposed that WUS is expressed in a mother cell in tissue $i-1$, if it is expressed in one of its daughter cells in tissue $i$. The candidate target pattern that we proposed correspond to a situation in which the WUS domain is controlled by gradients of signals in the meristem, independently of the lineages of WUS expressing cells. This implies that a daughter cell of a WUS expressing mother cell may not express WUS gene. This fact was

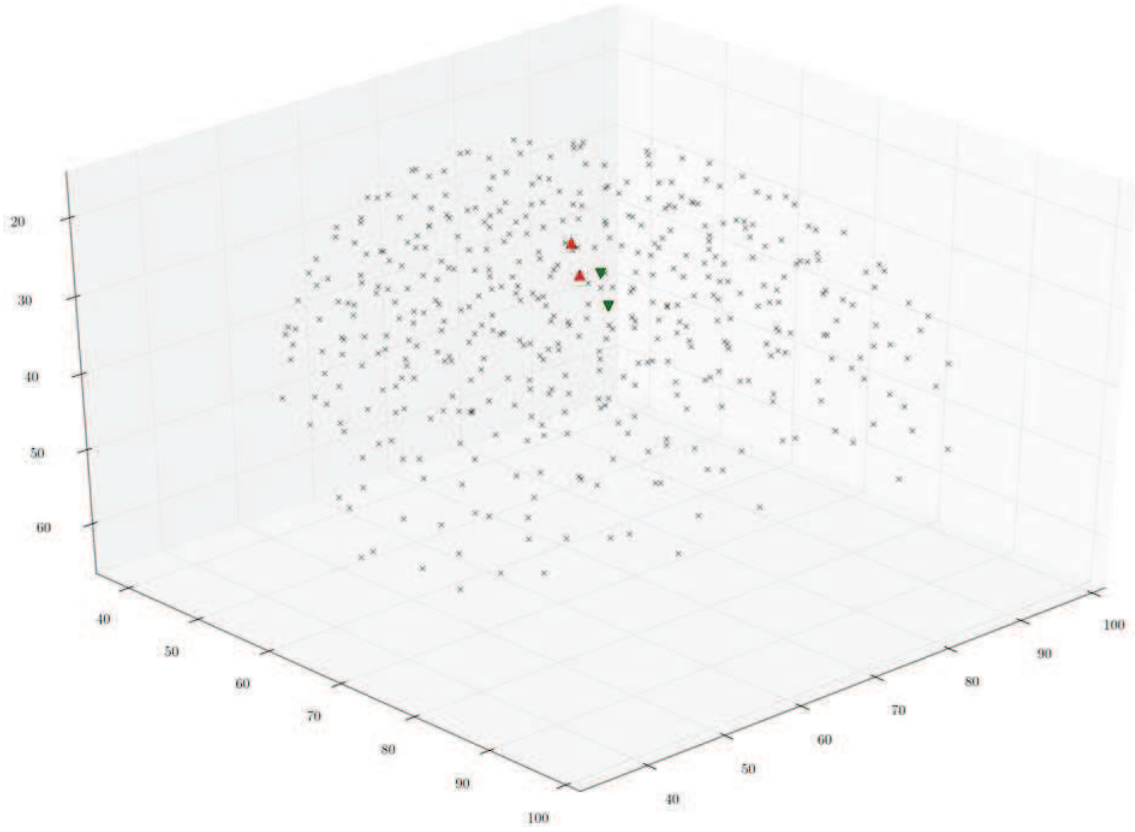



Figure 5.23: The WUS pattern created by interpolated parameter values for tissue $T_2$. The cell centres of tissue $T_2$ are presented by small $\times$ signs. The cells centres belonging to manually defined target pattern are marked by red triangles. Green triangles correspond to the cell centres in interpolated WUS pattern.

actually observed in our study, since the obtained candidate patterns were different from those defined using cell lineages.

The main conclusion to be drawn from this study is that the hypothesis is partially valid regarding our manually defined target pattern. To completely validate this hypothesis we need to characterize WUS target pattern using images obtained from confocal images where WUS is marked by a reporter gene as GFP. At the time being these informations are not available. Hopefully they will be available in near future allowing us to test our hypothesis.

# Conclusion

In this dissertation, in Chapter 1, we presented the state of art about phyllotaxis at different scales, i.e. macroscopic scale, meristematic scale and scale of genes and plants hormones. In Chapter 2, after a brief overview of gene regulation processes from biological point of view, we presented the formalisms for modelling gene regulatory networks. We presented cellular automata as an efficient modelling tool for gene regulatory network.

One of the main objectives of this dissertation was first to show how the three scales are related to each other, which can lead us to better understand phyllotaxis. To achieve this goal, in Chapter 3, we began by studying measured divergence angles for both wild-type and *ahp6* mutant. We observed perturbations in the sequences of divergence angles. We carried out a statistical analysis that confirmed the hypothesis that permutations of successive organs along the stem can explain perturbations. Based on this hypothesis we designed a combinatorial model and algorithms. The application of the algorithms allowed us to explain nearly all of the perturbations observed by permutations. The results confirmed the hypothesis of permutations and shows that permutations occur frequently in *Arabidopsis thaliana*. The algorithms can be combined with other tools in order to test the ability of phyllotaxis models in reproducing realistic sequences.

The frequency of permutations observed in measured sequences of divergence angles in Chapter 3, suggested that permutations should not be neglected in phyllotaxis studies. In particular the models should be able to regenerate such patterns. We used the algorithms of Chapter 3 to assess the ability of a model to reproduce realistic sequences of divergence angles. The model was meant to study the role of noise in generating permutation patterns. The noise affected the size of central zone as well as the threshold of organ initiations. We found out that the noise induced another type of perturbations that we interpreted as inversions of generative spiral. We proposed a method that used the algorithms of Chapter 3 to study the inversions. The analysis of the simulations of the model using this method, highlighted the importance of the central zone size robustness as the noise added to the size of central zone induced spiral inversions that we considered as unrealistic.

The regulation of central zone is the subject of Chapter 5. In this chapter we moved to the

scale of gene regulatory network to study the mechanism controlling the size of meristem. In particular the negative feedback network between WUS, CLV1, and CLV3 is treated as the interactions between these genes control the size of central zone. We simulated a three dimensional implementation of a model proposed by Jönsson *et al.* [Jönsson *et al.*, 2005]. In this model two signals originating from L1 layer and stem represses the expression of WUS. Our simulations of this model led us to hypothesize that the geometrical shape of meristem can have a role in the WUS expression domain. To further investigate this idea we proposed a general method to estimate the parameters of a model based on differential equations for a given target pattern. We used this method to optimise parameters of the model on a three dimensional digitized meristem corresponding to a meristem tracked over a period of time. The optimal parameter values reproduced a target patterns that were obtained using cell lineages. By interpolating optimal parameter values we proposed plausible, alternative candidate patterns. The main conclusion that can be drawn from our study is that the geometrical shape of meristem can define the expression zone of WUS. Since the candidate patterns differ from the target patterns, we believe that our tools can be used to track WUS expressing cells over time.

The general goal that motivated this thesis, was to improve our understanding of phyllotaxis. One major observation that we made is the fact that phyllotaxis is not the perfectly regular phenomenon that was presented in most studies on this subject. The few studies that were concerned with perturbations did so by considering mostly noise and randomness in phyllotaxis. We showed that the perturbations can be highly structured, and in our case, they were well described by permutation patterns. Importantly, these patterns, were very frequent in wild-type plants. This raises the question of whether these patterns are a general feature of phyllotaxis in other species as well as other mutants. To answer this question, more biological data need to investigated. Recent phenotyping techniques will allow for such data to be produced massively in the coming years. In this context, the tools developed in Chapter 3 as well as extensions as presented in Chapter 4, will participate in the global effort of the scientific community to decipher the link between genotype and phenotype.

One of the major elements playing a role in phyllotaxis, is the central zone, which is defined in terms of specific gene expression patterns. The role of genes in defining the geometrical properties in biology has been already studied [Coen *et al.*, 2003]. In Chapter 5, our study of the central zone formation led us to a seemingly opposite point of view, where the geometry of a biological tissue determines the underlying gene patterns. Some further thoughts show that these points of view are complementary rather than opposite. For instance in Chapter 5, it is not only the geometry of the SAM that controls the expression zone of WUS gene, but also some pre-defined gene patterns implicitly modelled in external layers of the SAM. Therefore, a feedback between the geometry of biological tissues and gene patterns, appears to be at the

heart of morphogenetic processes. The methods that we introduced in Chapter 5, can be useful to understand this feedback. Actually, these tools allow us to distinguish the patterns defined by cell lineages from those that are under control of geometry and propose candidate patterns which, we hope, will be confronted with experimental data in the future.

# Appendix A

# Software implementation

## A.1 Gene regulatory network simulator: a prototype

We developed a prototype for a simulator of gene regulatory network. The simulator gets as input a 4D digitized meristem as explained in Chapter 5 and a model described in a file of Python functions, see Figure A.1. The simulator allows to run the model and to simulate it on the 4D meristem. The simulations can be saved or loaded and can be tracked in time. Colour palates can be used to visualize gene concentrations in cells. By pointing the mouse pointer to a cell, different information about the cell such as its volume, its center can be obtained. It has an integrated two dimensional viewer that can be used to explore the 3D meristem and view the generated gene patterns. This is a prototype of a gene regulatory networks simulator. In the future, other features as a 3D viewer will be added. The code is about one thousand lines and is written using Python programming language and the graphical user interface is implemented using PyQT.

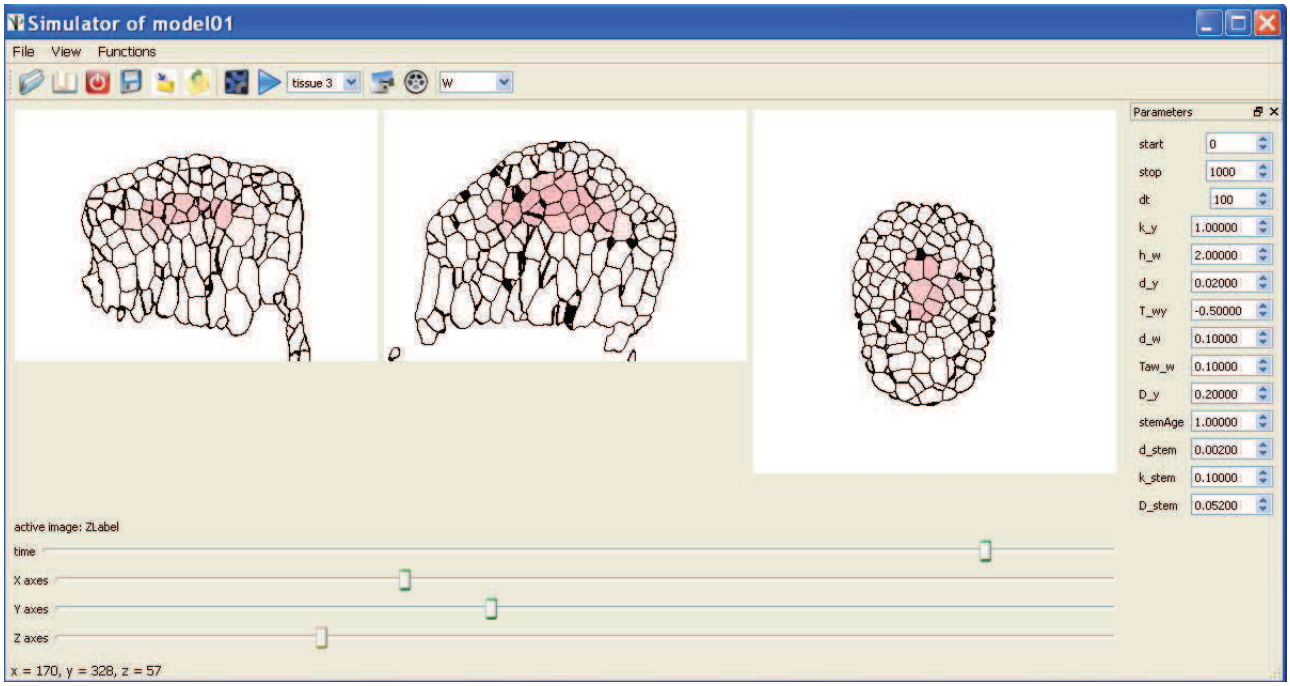



Figure A.1: Snapshot of the gene regulatory network simulator. The simulator has an integrated two-dimensional viewer that can be used to explore a 4D meristem. A model described in a python file can be loaded. On the right pane the parameters of the model can be dynamically viewed and modified. The simulator simulates the model. The simulation can be saved or loaded.

# Appendix B

# Software implementation

We implemented a prototype of two dimensional simulator of gene regulatory networks using the cellular automata. An artificial two dimensional tissue can be created and discretized using hexagonal cells, see Figure B.1. The size of tissue, size of cells and time step can be modified. The model to simulate should be described in a Python file. Colour palates can be used to represent gene concentration in cells.

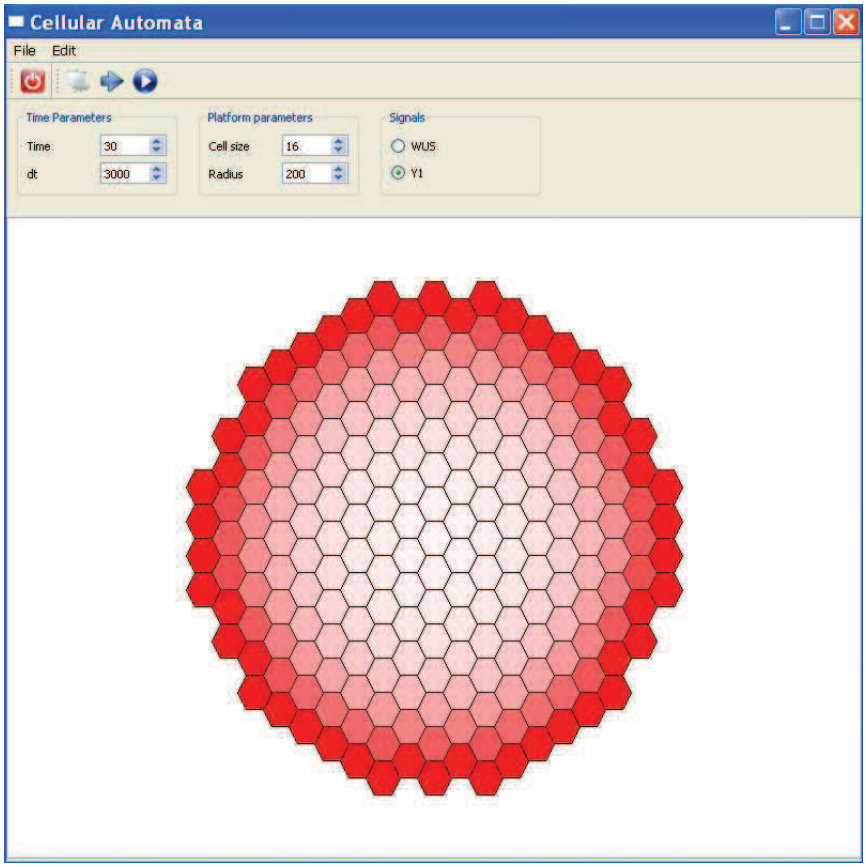



Figure B.1: Snapshot of the simulator of gene regulatory networks using cellular automata.

# Appendix C

# Definition of (hidden) Variable-order Markov chains

In the following, we first introduce high-order Markov chains. Then, variable-order Markov chains and hidden Markov models based on first-order and variable-order Markov chains. In the case of an $r$th-order Markov chain $\{S_t; t = 0, 1, \ldots\}$, the conditional distribution of $S_t$ given $S_0, \ldots, S_{t-1}$ depends only on the values of $S_{t-r}, \ldots, S_{t-1}$ but not further on $S_0, \ldots, S_{t-r-1}$,

$$P\left(S_t = s_t | S_{t-1} = s_{t-1}, \ldots, S_0 = s_0\right) = P\left(S_t = s_t | S_{t-1} = s_{t-1}, \ldots, S_{t-r} = s_{t-r}\right)$$

In our context, the random variable represents theoretical divergence angles and can take the five possible values $\alpha$, $2\alpha$, $-\alpha$, $3\alpha$ and $-2\alpha$. These five possible values correspond to the Markov chain states. A $J$-state $r$th-order Markov chain has $J^r\left(J-1\right)$ independent transition probabilities. Therefore, the number of free parameters of a Markov chain increases exponentially with the order. Let the transition probabilities of a second-order Markov chain be given by

$$p_{hij} = P\left(S_t = j | S_{t-1} = i, S_{t-2} = h\right) \quad \text{with } \sum_j p_{hij} = 1.$$

These transition probabilities can be arranged as a $J^2 \times J$ matrix where the row $(p_{hi0}, \ldots, p_{hiJ-1})$ corresponds to the transition distribution attached to the [state $h$, state $i$] memory. If for a given state $i$ and for all pairs of states $\left(h, h^{'}\right)$ with $h \neq h^{'}$, $p_{hij} = p_{h^{'}ij}$, i.e. once $S_{t-1}$ is known, $S_{t-2}$ conveys no further information about , the $J$ memories of length 2 [state $h$, state $i$] with $h = 0, \ldots, J-1$ can be grouped together and replaced by the single [state $i$] memory of length 1 with associated transition distribution $(p_{i0}, \ldots, p_{iJ-1})$. This illustrates the principle for building a variable-order Markov chain. In a variable-order Markov chain, the order (or memory length) is variable and depends on the "context" within the sequence.

The memories of a Markov chain can be arranged as a memory tree such that each vertex (i.e. element of a tree graph) is either a terminal vertex or has exactly $J$ "offspring" vertices. In practice, the memories corresponding to unobserved contexts are not included in the memory

tree (this is the case for the memory $3\alpha2\alpha$ that was not observed). The memories associated with the $J$ vertices (memories of length $r+1$) deriving from a given vertex (memory of length $r$) are obtained by prefixing the parent memory with each possible state. For instance, *in Fig. x*, the four second-order memories $\alpha2\alpha$, $2\alpha2\alpha$, $-\alpha2\alpha$ and $-2\alpha2\alpha$ derive from the first-order memory $2\alpha$. A transition distribution is associated with each terminal vertex of this memory tree.

Because of the noisy character of the measurements, we built a noisy or hidden Markov model (Ephraim and Merhav, 2002) based on a variable-order Markov chain. In this hidden Markov model, the non-observable variable-order Markov chain represents the succession of theoretical divergence angles along the axes while the von Mises observation distributions represent measurement uncertainty. A hidden variable-order Markov chain can be viewed as a two-level stochastic process, i.e. a pair of stochastic processes $\{S_t, X_t\}$ where the "output" process $\{X_t\}$ is related to the "state" process $\{S_t\}$ , which is a finite-state variable-order Markov chain, by the observation distributions. The probability density function of the von Mises observation distribution (in degrees) for state $j$ is given by

$$g_j\left(x_t;\mu_j,\kappa\right)=\frac{1}{360I_0\left(\kappa\right)}\exp\left\{\kappa\cos\left(x_t-\mu_j\right)\frac{\pi}{180}\right\}.$$

where $I_0\left(\kappa\right)$ is the modified zero order Bessel function of the first kind, $\mu_j$ the location parameter (analogous to the mean) and $\kappa$ the concentration parameter (precision or inverse variance) (Mardia and Jupp, 2000). The von Mises distribution, also known as the circular Gaussian distribution, is a univariate Gaussian-like periodic distribution for a variable $x\in[0,360^\circ)$. The distribution has period $360^\circ$ so that $g_j\left(x_t+360;\mu_j,\kappa\right)=g_j\left(x_t;\mu_j,\kappa\right)$. The definition of the observation distributions expresses the assumption that the output process at time $t$ depends only on the non-observable Markov chain at time $t$.

# Appendix D

# Selection of the Memories of a Variable-order Markov chain

The order of a Markov chain can be estimated using the Bayesian information criterion (BIC). For each possible order $r$, the following quantity is computed

$$\mathrm{BIC}\,(r) = 2\log L_r - J^r\,(J-1)\log n.$$

where $L_r$ is the likelihood of the observed sequences for the $r$th-order estimated Markov chain, $J^r\,(J-1)$ is the number of independent transition probabilities of a $J$-state $r$th-order Markov chain and $n$ is the cumulated length of the observed sequences. The principle of this penalized likelihood criterion consists in making a trade-off between an adequate fitting of the model to the data (given by the first term in (A1)) and a reasonable number of parameters to be estimated (control by the second term, the penalty term). In practice, it is infeasible to compute a BIC value for each possible variable-order Markov chain of maximum order $r \leq R$ since the number of hypothetical memory trees is very large. An initial maximal memory tree is thus built combining criteria relative to the maximum order and to the minimum count of memory occurrences in the observed sequences. This memory tree is then pruned, using a two-pass algorithm which is an adaptation of the context-tree maximizing algorithm [Csiszár *et al.*, 2006]: a first dynamic programming pass, starting from the terminal vertices and progressing towards the root vertex for computing the maximum BIC value attached to each sub-tree rooted in a given vertex, is followed by a second tracking pass starting from the root vertex and progressing towards the terminal vertices for building the memory tree.

# Bibliography

[Abramowitz *et al.*, 1970] Abramowitz, M., Stegun, I. A., *Handbook of mathematical functions with formulas, graphs and mathematical tables*, Dover Publications, New York (1970).

[Adler *et al.*, 1997] Adler I., Barabé D., Jean R.V., *A History of the Study of Phyllotaxis*, Annals of Botany, 80 (3):231-244 (1997).

[Alber *et al.*, 2002] Alber M., Kiskowski M., Glazier J., Jiang Y., *On Cellular Automaton Approaches to Modeling Biological Cells*, in IMA 134: Mathematical systems theory in biology, communication, and finance, Springer-Verleg, New York. , pp. 12, (2002).

[Allard *et al.*, 1951] Allard, H.A., *The ratios of clockwise and counterclockwise spirality observed in the phyllotaxy of some wild plants*, The Journal of the Southern Appalachian Botanical Club 16:1 (1951).

[Alon, 2007] Alon U., *An introduction to systems biology design principles of biological circuits*, Chapman & Hall/CRC Mathematical and Computational Biology Series (2007).

[Alvarez-Buylla *et al.*, 2007] Alvarez-Buylla E., Benítez M., Dàvila E., Chaos À. Espinosa-soto C., Padilla-Longoria P., *Gene regulatory network models for plant development*, Current opinion in plant biology, 10:83-91 (2007).

[Balzter *et al.*, 1998] Balzter H., Braun P.W., Kohler W., *Cellular automata models for vegetation dynamic*, Ecological modelling, 107:113-125 (1998).

[Barbier *et al.*, 2006] Barbier de Reuille P., Bohn-Courseau, I. Ljung K., Morin H., Carraro N., Godin C., Traas J. *Computer simulations reveal properties of the cell-cell signaling network at the shoot apex in arabidopsis*, Proceedings of the National Academy of Sciences of United States of America, 103(5):1627-1632, (2006).

[Barbier, 2005] Barbier de Reuille P. *Toward a dynamic model of the shoot apical meristem of Arabidopsis thaliana*, PhD thesis, Universite Montpellier II, December 2005

[Beal, 1873] Beal, W.J., *Phyllotaxis of cones*, The American naturalist, 7(8) (1873).

[Ben-Dor *et al.*, 1999] Ben-Dor A., Shamir R., Yakhini Z., *Clustering gene expression patterns*, Journal of computational biology, 6(3-4): 281-297 (1999).

[Besnard *et al.*, 2011] Besnard, F., Refahi, Y., Morin, V., Chambrier, P., Legrand, J., Brunoud, G., Farcot, E., Marteaux, B., Cellier, C., Das, P., Bishopp, A., Helariutta, Y., Godin, C., Traas, J., Guédon, Y. and Vernoux, T. *Cytokinin inhibitory fields control phyllotaxy by converting spatial to temporal information*, Submitted (2011).

[Bower *et al.*, 2001] Bower J., Bolouri H., *Computational modelling of genetic and biochemical networks*, Massachusette Institute of technology (2001).

[Bravais *et al.*, 1837] Bravais L, Bravais A. *Essai sur la disposition des feuilles cuivisériès*, Annales des Sciences Naturelles Botanique 7: 42-110; 193-221; 291-348; 8: 11-42 (1837).

[Chapman *et al.*, 1987] Chapman, J.M., Perry R., *Diffusion model of phyllotaxis.* Annals of Botany, 60: 377-387. (1987).

[Clark *et al.*, 1993] Clark S.E., Running M.P., Meyerowitz E.M. *CLAVATA1, a regulator of meristem and flower development in Arabidopsis*, Development, 119:397-418 (1993).

[Coen *et al.*, 2003] Coen E., Rolland-Lagan A., Matthews M., Bangham J.A., Prusinkiewicz P., *The genetics of geometry*, Proceedings of the National Academy of Sciences of United States of America, 101(14):4729-4735 (2003).

[Comtet, 1972] Comtet, L., *Sur les coefficients de l'inverse de la série formelle* $\sum n!t^n$, Compt. Rend. Acad. Sci. Paris A 275:569-572 (1972).

[Comtet, 1974] Comtet L., *Advanced Combinatorics*, D.Reidel, Dordrecht, Holland (1974).

[Cooke, 2006] Cooke T.J., *Do Fibonacci numbers reveal the involvement of geometrical imperatives or biological interactions in phyllotaxis?*, Botanical Journal of the Linnean Society, 150:3-24 (2006).

[Couder, 1998] Couder Y., *Initial transitions, order and disorder in phyllotactic patterns: the ontogeny of* Helianthus annuus. *A case study*, Acta Societatis Botanicorum Poloniae, 67:129-150 (1998).

[Csiszár *et al.*, 2006] Csiszár, I. and Talata, Z., *Context tree estimation for not necessarily finite memory processes, via BIC and MDL.* IEEE Transactions on Information Theory, 52:1007-1016 (2006).

[de Jong, 2002] de Jong H., *Modeling and simulation of genetic regulatory systems: A literature review*, Journal of computational biology, 9:67-103 (2002).

[de Jong *et al.*, 2004] de Jong H., Geiselmann J., Batt G., Hernandez C., Page M., *Qualitative simulation of the initiation of sporulation in Bacillus subtilis*, Bull. Math. Biol., 66(2):261-300 (2004).

[Douady *et al.*, 1996(a)] Douady S, Couder Y., *Phyllotaxis as a self organizing process Part I: The spiral modes resulting from time-periodic iterations*, Journal of Theoretical Biology 178:255-274 (1996) .

[Douady *et al.*, 1996(b)] Douady S, Couder Y., *Phyllotaxis as a self organizing process Part II: The whorled modes and their relation to the spiral ones*, Journal of Theoretical Biology 178:275-294 (1996) .

[Douady *et al.*, 1996(c)] Douady S, Couder Y., *Phyllotaxis as a self organizing process part III: The simulation of Transient regimes of ontogeny*, Journal of Theoretical Biology 178:295-312 (1996) .

[Edwards, 2000] Edwards R., *Analysis of continuous-time switching networks*, Physica D, 146:165-199 (2000).

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

[Ermentrout, 1993] Ermentrout G.B., Edelstein-Keshet L. *Cellular automata approaches to biological modeling*, J. theor. Biol., 160:97-133 (1993).

[Farcot, 2006] Farcot E., *Geometric properties of piecewise affine biological network models*, Journal of Mathematical Biology, 52(3):373-418 (2006).

[Fernandez *et al.*, 2010] Fernandez R., Das P, Mirabet V., Moscardi E., Traas J., Verdeil J., Malandain G., Godin C. *Imaging plant growth in 4D: robust tissue reconstruction and lineaging at cell resolution*, Nature Methods, 7:547-553 (2010).

[Sbalzarini, 2009] F.Sbalzarini I., *Spatiotemporal Modelling and Simulation in Biology*, Course notes, (2009).

[Geier, *et al.*, 2008] Geier F., Lohmann J., Gerstung M., Maier A., Timmer J., Fleck C., *A Quantitative and Dynamic Model for Plant Stem Cell Regulation*, PlosOne, 3, 2008.

[Gibson *et al.*, 2001] Gibson M., Mjolsness E., *Modelling the activity of single genes*, Computational modelling of genetic and biochemical networks, Massachusette Institute of technology, 3-48 (2001).

[Gierer *et al.*, 1972] Gierer A., Meinhardt H., *A theory of biological pattern formation*, Kybernetik, 12:30-39 (1972).

[Glass, 1975] Glass L., *Combinatorial and topological methods in nonlinear chemical kinetics*, J. Chem. Phys. 63:1325-1335 (1975).

[Glass *et al.*, 1973] Glass L., Kauffman S.A., *The logical analysis of continuous, nonlinear biochemical control networks*, J. Theor. Biol. 39, 103-129 (1973).

[Gouzé *et al.*, 2003] J.L. Gouzé, T. Sari, *A class of piecewise linear differential equations arising in biological models*, Dynamical systems, 17:299–316 (2003).

[Gordon *et al.*, 2009] Gordon S., Chickarmane V., Ohno C., Meyerowitz E.M., *Multiple feedback loops through cytokinin signaling control stem cell number within the Arabidopsis shoot meristem*, Proceedings of the National Academy of Sciences of United States of America, 106:16529-16534 (2009).

[Gray *et al.*, 1983] Gray P., Scott S.K., *Autocatalytic reactions in the isothermal, continuous stirred tank reactor: Isolas and other forms of multistability*, Chemical Engineering Science, 38:29-43 (1983).

[Gray *et al.*, 1984] Gray P., Scott S.K., *Autocatalytic reactions in the isothermal, continuous stirred tank reactor: Oscillations and instabilities in the system* $A + 2B \rightarrow 3B$; $B \rightarrow C$, Chemical Engineering Science, 39:1087-1097 (1984).

[Gray *et al.*, 1985] Gray P., Scott and S.K., *Sustained oscillations and other exotic patterns of behavior in isothermal reactions*, Journal of Physical Chemistry, 89:22-32 (1985).

[Green, 1992] Green P.B., *Pattern formation in shoots: a likely role for minimal energy configurations of the tunica*, International Journal of Plant Sciences, 153(3):S59-S75 (1992).

[Green, 1996] Green P.B., *Expression of form and pattern in plants ? a role for biophysical fields*, Seminars in Cell & Developmental Biology, 7:903-911 (1996).

[Heisler *et al.*, 2006] Heisler M., Jonsson H., *Modeling auxin transport and plant development*, Journal of Plant Growth Regulation, 25:302-312 (2006).

[Hirsch *et al.*, 2004] Hirsch M., Smale S., Devaney R., *Differential equations, dynamical systems & an introduction to chaos*, Second Edition, Elsevier (2004).

[Hobe *et al.*, 2003] Hobe M., Müller R., Grünewald M., Brand U. Simon R. *Loss of CLE40, a protein functionally equivalent to the stem cell restricting signal CLV3, enhances root waving in Arabidopsis*, Dev. Genes. Evol., 213:371-381 (2003).

[Hofmeister, 1868] Hofmeister W., *Allgemeine Morphologie der Gewähse*, Handbuch der Physiologischen Botanik, Leipzig, Engelmann 1:405-644 (1868).

[Hohm *et al.*, 2009] Hohm T., Zitzler E., *A multiobjective evolutionary algorithm for numerical parameter space characterization for reaction diffusion systems*, Pattern Recognition in Bioinformatics, Lecture Notes in Computer Science, 2009:162-174, (2009).

[Hohm *et al.*, 2010] Hohm T., Zitzler E., Simon R., *A Dynamic Model for Stem Cell Homeostasis and Patterning in Arabidopsis Meristems*, Plos One, 5, (2010).

[Jacob *et al.*, 1961] Jacob F., Monod J., *Genetic regulatory mechanisms in the synthesis of proteins*, Journal of Molecular Biology 3:318 (1961).

[Jean *et al.*, 1994] Jean R.V., *Phyllotaxis: A Systemic Study in Plant Morphogenesis*, Cambridge University Press, Cambridge, UK (1994).

[Jeune *et al.*, 2004] Jeune B., Barabé J., *Statistical Recognition of Random and Regular Phyllotactic Patterns*, Annals of Botany, 94:913-917 (2004).

[Jeune *et al.*, 2006] Jeune B., Barabé J., *A stochastic approach to phyllotactic pattern analysis*, Journal of Theoretical Biology 238:52-59 (2006).

[Jennifer *et al.*, 2007] Jennifer P.C., Kieber J.J., *Cytokinin signaling: two-components and more*, Trends in plant science, 13(2):85:92 (2007).

[Jönsson *et al.*, 2003] Jönsson H., Shapiro B.E., Mjolsness E., Meyerowitz E.M., *Signalling in multicellular models of plant development*, On Growth, Form and Computers. Academic Press, London, 21:156-161, (2003).

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

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

[Kauffman, 1993] Kauffman S.A., *The origin of order: self organization and selection in evolution*, Oxford University Press, New York (1993).

[Kier *et al.*, 1996] Kier L.B., Cheng C.K, Testa B., Carrupt P.A., *A cellular automata model of enzyme kinetics*, Journal of molecular graphics, 14:227:231, (1996).

[King, 2006] King A., *Generating Indecomposable permutations*, Discrete Mathematics, 306:508-518 (2006).

[Koch *et al.*, 1994] Koch, A.J., Meinhardt, H. *Biological pattern formation: from basic mechanisms*, Reviews of Modern Physic, 66(4):1481-1510 (1994).

[Krupinski *et al.*, 2010] Krupinski P., Jönsson H. *Modeling Auxin-regulated Development*, Cold Spring Harbor Laboratory Press, (2010).

[Kuhlemeier, 2007] Kuhlemeier C., *Phyllotaxis*, Trends in Plant Science, 12:143-150 (2007).

[Lindsay *et al.*, 2006] Lindsay DL, K. Vipen V, Bonham-Smith PC *Cytokinin-induced changes in CLAVATA1 and WUSCHEL expression temporally coincide with altered floral development in Arabidopsis* Plant Science, 170:1111-1117 (2006).

[Livio, 2002] Livio M., *The Golden Ratio: The Story of Phi, the World's Most Astonishing Number*, Broadway books, (2002).

[Lohmann *et al.*, 2002] Lohmann J.U., Weigel D., *Building beauty: The genetic control of floral patterning*, Developmental Cell, 2:135-142 (2002).

[Lu *et al.*, 1996] Lu P., Porat R., Nadeau J.A., O'NeillS.D., *Identification of a Meristem L1 Layer-Specific Gene in Arabidopsis That Is Expressed during Embryonic Pattern Formation and Defines a New Class of Homeobox Genes*, The Plant Cell, 8(12):2155-2168 (1996).

[Luthi *et al.*, 1998] Luthi P.O., Chopard B., Preiss A., Ramsden J.J., *A cellular automaton model for neurogenesis in Drosophila*, Physica D, 118:151-160, (1998).

[Mardia *et al.*, 2000] Mardia, K. V. and Jupp, P. E. *Directional Statistics*, John Wiley & Sons: Chichester, 2000.

[Meinhardt, 1974 ] Meinhardt H, *The formation of morphogenetic gradient and fileds*, Ber. Deutch. Bot. Ges. 87: 101-108 (1974).

[Mitchison *et al.*, 1977] Mitchison, G. H. *Phyllotaxis and Fibonacci series*, Science 196: 270:275 (1977).

[Mirabet *et al.*, 2011] Mirabet V., Besnard F., Vernoux T., Boudaoud A., *Noise and robustness in phyllotaxis*, submitted.

[Murray, 1989] Murray J.D., *Mathematical biology*, Springer-Verlag, Berlin (1989).

[Murray, 2003] Murray J.D., *Mathematical biology II: spatial models and biomedical applications*, Springer-Verlag, Berlin (2003).

[Owen *et al.*, 1999] Owen M, Sherratt J., Myers S., *How far can a juxtacrine signal travel?*, Proceedings of Royal Society, 266:579-585 (1999).

[Parcy *et al.*, 1998] Parcy F., Nilsson Ove, Maximilian A.B., Lee I., Weigel D, *A genetic framework for floral patterning*, Nature, 395:561-566 (1998).

[Parcy *et al.*, 2005] Parcy F., *Flowering: a time for integration*, International Journal of Development Biology 49:585-593 (2005).

[Parida, 2008] Parida L., *Pattern Discovery in Bioinformatics: Theory & Algorithms*, Chapman & Hall/CRC Mathematical Computational Biology series, (2008).

[Peaucelle *et al.*, 2007] Peaucelle A., Morin H., Traas J., Laufs P., *Plants expressing a miR164-resistant CUC2 gene reveal the importance of post-meristematic maintenance of phyllotaxy in Arabidopsis*, Development, 134:1045-1050 (2007).

[Pawlowski *et al.*, 2004] Pawlowski P.H., Zielenkiewicz P., *Biochemical kinetics in changing volumes*, Acta Biochimica Polonica, 51:231-243 (2004).

[Pilkington, 1929] Pilkington, M., *The regeneration of the stem apex.*, New Phytologist, 28: 37-53 (1929).

[Plahte, *et al.* 1998] Plahte E., Mestl T., Ohmolt S.W., *A methodological basis for description and analysis of systems with complex switch-like interactions*, Journal of Mathematical Biology, 36:321-348 (1998).

[Prasad *et al.*, 2011] Prasad K., Grigg S.P., Barkoulas M., Yadav R.K., Sanchez-Perez G.F, Pinon1 v., Blilou I., Hofhuis H., Dhonukshe P., Galinha1 C., Mähönen A.P., Muller W.H., Raman S., Verkleij A.J., Snel B., Reddy G.V., Tsiantis M., Scheres B., *Arabidopsis plethora transcription factors control phyllotaxis*, Current Biology 21(13):1123-1128 (2011).

[Prigogine *et al.*, 1968] Prigogine I., Lefever R., *Symmetry breaking instabilities in dissaptive systems.*, J. Chem. Phys., 48:1695-1700 (1968).

[Ptashne *et al.*, 2002] Ptashne M., Gann A., *Genes and signals*, Cold spring Harbor, New York, (2002).

[Ragni *et al.*, 2008] Ragni L., Belles-Boix, E, Günl1 M., Pautot V., *Interaction of KNAT6 and KNAT2 with BREVIPEDICELLUS and PENNYWISE in Arabidopsis Inflorescences*, The Plant Cell, 20:888-900 (2008).

[Reddy *et al.*, 2004] Reddy, G.V., Heisler, M.G., Ehrhardt, D.W., Meyerowitz, E.M., *Real-time lineage analysis reveals oriented cell divisions associated with morphogenesis at the shoot apex of Arabidopsis thaliana*, Development 131:4225-4237 (2004).

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

[Refahi *et al.*, 2010] Refahi, Y., Guédon, Y., Besnard, F., Farcot, E., Godin, C., Vernoux, T., *Analyzing perturbations in phyllotaxis of Arabidopsis thaliana.* In: 6th International Workshop on Functional-Structural Plant Models, Davis, California, 170-172 (2010).

[Refahi *et al.*, 2011] Refahi, Y., Guédon, Y., Besnard, F., Farcot, E., Godin, C., Vernoux, T., *A Combinatorial Model of Phyllotaxis Perturbations in Arabidopsis thaliana*, Lecture Notes in Computer Science (Proceedings of Combinatorial Pattern Matching 2011), 6661/2011:323-335 (2011).

[Reinhardt *et al.*, 2000] Reinhardt, D., Mandela T., Kuhlemeier, C. *Auxin Regulates the Initiation and Radial Position of Plant Lateral Organs*, Plant Cell, 12:507-518 (2000).

[Reinhardt *et al.*, 2001] Reinhardt, D., and Kuhlemeier, C. *Phyllotaxis in higher plants* In Meristematic Tissues in Plant Growth and Development. M.T. McManus and B.E. Veit, eds (Boca Raton, FL: CRC Press), 172-212, (2001).

[Reinhardt *et al.*, 2003(a)] Reinhardt D., Pesce E.R., Stieger P., Mandel T., Baltensperger K., Bennett M., Traas J., Frim J., Kuhlemeier C., *Regulation of phyllotaxis by polar auxin transport*, Nature, 426:255-260 (2003).

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

[Reinhardt, 2005] Reinhardt D., *Regulation of phyllotaxis*, International Journal of Developmental Biology, 49: 539-546 (2005).

[Richards, 1951] Richards F.J., *Phyllotaxis: its quantitative expression and relation to growth in the apex*, Philosophical Transactions of the Royal Society of London, B, 235:509-564 (1951).

[Sablowski, 2007] Sablowski R., *The dynamic plant stem cell niches*, Current opinion in plant biology, 10:639-644, (2007).

[Sarkar, 2000] Sarkar P., *A brief history of cellular automata*, ACM Computing Surveys, 32 (2000).

[Schoof *et al.*, 2000] Schoof H., Lenhard M., Haecker A., Mayer K.F.X., Jürgens G., Laux T., *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).

[Schoute, 1913] Schoute, J.C., *Beiträge zur Blattstellungslehre.* rce. Trav. Bot. Néerl. 10: 153-339 (1913).

[ Schroeder, 1953] Schroeder C.A., *Spirality in Citrus* 114, 3: 350-352 (1953).

[Sharma *et al.*, 2003] Sharma V., Carles C., Fletcher J., *Maintenance of stem cell populations in plants*, Proceedings of the National Academy of Sciences of United States of America 30:11823-11829 (2003).

[Shipman *et al*, 2005] Shipman P.D., Newell A.C., *Polygonal planforms and phyllotaxis on plants*, Journal of Theoretical Biology 236:154-197 (2005).

[Smith *et al.*, 2006] Smith R.S., Kuhlemier C., Prusinkiewicz P., *Inhibition fields for phyllotactic pattern formation: a simulation study*, Canadian Journal of Botany, 84:1635-1649, (2006).

[Smith *et al.*, 2005] Smith R.S., Guyomarc'h S., Mandel T., Reinhadrt D., Kuhlemier C., Prusinkiewicz P., *A plausible model of phyllotaxis*, Proceedings of the National Academy of Sciences of United States of America, 103(5):1301-1306 (2005).

[Smolen *et al.*, 1999] Smolen P., Baxter D., Byrne J., *Modeling transcriptional control in gene networks - methods, recent results, and future directions* , Bulletin of Mathematical Biology, 62:2:247-292, (1999).

[Smyth *et al.*, 1990] Smyth D.R., Bowman J.L., Meyerowitz E.M. *Early flower development in Arabidopsis*, Plant Cell, 2:755-767 (1990).

[Snow *et al.*, 1931] Snow M, Snow R., *Experiments on phyllotaxis. I. The effect of isolating a primordium.* Philosophical Transactions of the Royal Society London 221B:1-43 (1931).

[Snow *et al.*, 1952] Snow, M., Snow R., *Minimum areas and leaf determination*, Procedeings of Royal Society, B139:545-566 (1952).

[Stoma *et al.*, 2008] Stoma S., Lucas M., Chopard J., Schaedel M., Traas J., Godin C., *Flux-based transport enhancement as a plausible unifying mechanism for auxin transport in meristem development*, PLOS Computational Biology, (2008).

[Soille *et al.*, 1999] Soille, P. *Morphological Image Analysis: Principles and Applications*, Springer-Verlag, Berlin, 89-125 (1999).

[Vernoux *et al.*, 2010] Vernoux T., Besnard F., Traas J., *Auxin at shoot apical meristem*, Cold Spring Harbor Perspectives in Biology, published online (2010).

[Thomas *et al.*, 1973] Thomas R., *Boolean formalization of genetic control circuits*, Journal of Theoretical Biology, 42:563-585 (1973).

[Thomas *et al.*, 1990] Thomas R., D'Ari R., *Biological feedback*, CRC Press, Boca Raton, FL. (1990).

[Traas *et al.*, 2010] Traas J., Monéger F., *Systems Biology of organ initiation at shoot apex*, Plant Physiology, 152:420-427 (2010).

[Turing, 1952] Turing A.M., *The chemical basis of morphogenesis* Phil. Trans. R. Soc. Lon. B 237:37-72 (1952).

[Veen *et al.*, 1977] Veen A. H. & Lindenmayer, A. *Diffusion mechanism for phyllotaxis.* Plant Physiology, 60: 127-139(1977).

[Wardlaw, 1968] Wardlaw C. W. *Essays on Forms in Plants* Manchester, Manchester University Press (1968).

[Weigel *et al.*, 2002] Weigel D., Jürgens G., *Stem cells that make stems*, Nature, 415:751-754 (2002).

[Weimar, 1996] Weimar J.R., *cellular automata for reaction-diffusion systems*, Parallel Computing, 23:1699-1715 (1997).

[Weimar, 1998] Weimar J.R., *Simulation with cellular automata*, Logos-Verlag, Berlin, (1998).

[Weimar, 2002(a)] Weimar J.R., *Three-dimensional Cellular Automata for Reaction-Diffusion Systems*, Fundamenta Informaticae , 52:275-282 (2002).

[Weimar, 2002(b)] Weimar J.R., *Cellular automata as a simulation tool*, Course notes, (2002).

[Wolfram, 2002] Wolfram, S., *A New Kind of Science*, Wolfram Media, Inc., Champaign, IL, (2002).

[Wurthner *et al.*, 1999] Wurthner J.U., Mukhopadhyay A.K., Peimann C.-J., *A cellular automaton model of cellular signal transduction*, Computers in Biology and Medicine, 30:1-21 (1999).

[Young, 1978] Young D.A., *On the diffusion theory of phyllotxis*, Journal of Theoretical Biology, 71:421-432 (1978).

[Zeron, 2007] Zerón M.S., *Complex Dynamic Behavior in Simple Gene Regulatory Networks*, AIP Conference Proceedings, 885(1):161-184, (2007).

[INRA webpage] http://www.inra.fr/internet/Directions/DIC/ACTUALITES/SIA2001/arabidopsisintro.html

[openalea webpage] http://openalea.gforge.inria.fr/doc/vplants/phyllotaxis_analysis/doc/_build/html/contents.html

[www.math.smith.edu/phyllo] http://www.math.smith.edu/phyllo