Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

Recursive Algebraic Modelling of Gene Signalling, Communication and Switching

S. Pereverzyev Jr., R.S. Anderssen

RICAM-Report 2008-24

Recursive Algebraic Modelling of Gene Signalling, Communication and Switching

Sergiy Pereverzyev Jr.[1] and Robert S. Anderssen[2]

[1] Industrial Mathematics Institute, Johannes Kepler University Linz,
Altenbergerstrasse 69, A-4040 Linz, Austria,
pereverzyev@indmath.uni-linz.ac.at

[2] CSIRO Mathematical and Information Sciences, PO Box 664, Canberra,
ACT 2601, Australia,
Bob.Anderssen@csiro.au

Abstract

In the modelling of genetic signalling, communication and switching (GSCS), there is a need to identify the various mechanistic models, which nature has discovered, in terms of simple positional information rules (Wolpert (1969a)). The discovery of such simple rules, however, is a highly non-trivial process; in part, because of the complexity of the plethora of organs and organisms that such protocols are able to construct; in part, because the same rules will not apply universally to all related genetic processes (e.g. the genetic dynamics within epidermal cells and trichomes). In this paper, based on the model proposed by Young (1983) for pea leaf development, a framework is formulated for the explorative combinatorial algebraic mechanistic modelling of the GSCS control of some specific plant development process. Its indicative utility is illustrated with a notional mechanistic model for the genetic control of the positioning of trichomes on Arabidopsis leaves, by using different rules that take account of the different genetic activities of the GL1, TRY and CPC genes. The overall goal of the computational modelling is an illustration of how, for an algebraic model of some biological process, to utilize, in an iterative manner, the available biological knowledge in the formulation of model that captures the essence of the biology being investigated. In particular, it illustrates how, when all factors are taken into account, the resulting model can be quite elementary.

1. Introduction

In problem-solving and modelling, once the matter under investigation has been specified, the first step is the conceptualization of the mathematical framework within which a model will be formulated, simulated, analysed, and validated. Conceptualization is challenging for various reasons. On the one hand, it does not only demand familiarity with the use of a wide variety of mathematical concepts and techniques, but it also requires a sufficiently comprehensive understanding of the problem context to the point where that understanding helps to guide the choice of the framework and the formulation of a representative model within it. On the other hand, because of the comprehensive scope of mathematical knowledge and the huge range of opportunities that this knowledge represents for model formulation, the process of

conceptualization is akin to solving an inverse problem as the various possibilities are identified, assessed and compared.

This inverse problem aspect is clearly apparent when formulating a model for simulating some genetic signalling, communication and switching (GSCS) process. In the specific matter to be investigated, such as the genetic control of the positioning of trichomes on Arabidopsis leaves, there is always a multiplicity of choices for how the modelling might be performed, complicated by the fact that information about the process to be modelled is limited with the involved mechanisms often unknown or poorly understood (Schellmann et al. (2002)). In addition, the estimation of the parameters in the chosen model, in order to reproduce a given biological pattern, can, on occasions, be acutely sensitive to small changes in the pattern.

For the simulation of pattern formation in plants, such as phyllotactic positioning of leaves on the branches of plants (Kuhlemeier (2007)), various macroscopic and cellular models have been proposed and implemented. At the macroscopic level, ordinary and partial differential equations have been successfully used to simulate observed pattern formation process (Meinhardt (1982), Jonsson et al. (2006)). At the cellular level, rewriting, implemented using L-systems methodologies, is a popular tool (Prusinkiewicz and Lindenmayer (1990) and Smith et al. (2006)). To date, the goal of such modelling has been to illustrate how the evolving patterns of some particular biological processes can be simulated, without necessarily understanding the underlying mechanism.

Though, as explained by Kuhlemeier (2007; p. 148), the sophistication and utility of current mechanistic models is limited because “the lack of experimental data remains a bottleneck”, this should not limit the formulation of different modelling alternatives when they are based on widely accepted assumptions about the genetic and biological processes involved.

Depending on the matter under investigation, the inverse problem nature of genetics is a two-way street. On the one hand, one has the type of consideration outlined above. On the other hand, genetics itself has the aspects of an inverse problem in the sense that small changes in the expressed genes can cause corresponding large changes in the phenotype. Mutant studies of trichomes on Arabidopsis leaves yield excellent examples. If both the TRY and CPC genes are switched off then the regular position of the trichomes is totally disrupted with the trichomes forming irregularly located clumps (Larkin et al. (2003) and Hulskamp (2004; Figure 2)), whereas if the GL1 gene is switched off then either no or very few trichomes initiate (Larkin et al. (2003)).

It is such mutant comparisons that are the source of the available information (data) about the genetic activity of some biological module. The biological modularization concept of Hartwell et al. (1999) will be closely followed in the logic developed below. The inverse problems situation is clear: phenotype depends discontinuously on gene changes with the model of the forward problem from phenotype to genetics unknown or poorly understood. Accordingly, the modelling strategy proposed and analysed here is the categorization of the genes into equivalence classes that relate to the different known biological processes occurring and then the algebraic modelling of the interaction within and between the genes in these equivalence classes. For

example, in the initiation and positioning of trichomes on Arabidopsis leaves, the equivalence classes could be the genes controlling the switching of the fate of an epidermal cell to become a trichome, the genes which subsequently control the single cellular endoreduplication of the trichome and the genes that inhibit adjacent epidermal cells switching to a trichome fate.

The question addressed in this report is "How does one model the genetics of the coordination and exploitation of positional information by a group of cells performing a specific task?". The specific activity modelled is the initiation and positioning of trichomes on the adaxial layer of Arabidopsis leaves. The foundations, on which the proposed recursive algebraic methodology is based, are:

(i) The work of Young (1983) on the modelling of the development of the pea leaf. An examination of the methodology that he proposed yields a framework in which the recursive algebraic modelling of biological processes can be formulated, simulated, analysed, and interpreted.

(ii) The positional information concept proposed by Wolpert (1969a, 1969b). It yields a framework for the analysis of biological processes which fits naturally with the module concept of Hartwell et al. (1999), the neighbourhood nature of biological processes and gene regulatory networks.

(iii) The concept of Hartwell et al. (1999), Sellers (1989) and Milner (1964) that, as articulated by Sellers (1989; p.295), "any mechanism can be decomposed into a sum of irreducible parts".

The overall goal of the deliberations below is to illustrate how, in algebraic biological modelling, the available biological data, when taken fully into account, can lead to model simplification.

The paper is organized in the following manner. Relevant background is introduced in Section 2 that is related to the positional information concept of Wolpert (1969a, 1969b), the modularization of biology and biological modelling (Hartwell et al. (1999)) and leaf morphogenesis (Donnelly et al. (1999) and Fleming et al. (2005, 2006)). In Section 3, an analysis of the utilized by Young (1983) recursive algebraic methodology to model pea leaf morphogenesis is used to identify the general structure of the framework in which algebraic models of biological processes can be formulated. The essential steps in this framework are detailed in Section 4. A categorization of what are believed to be the key genes in the initiation and positioning of trichomes on Arabidopsis leaves is given in Section 5. This categorization becomes the basis in Section 6 for the formulation of a recursive algebraic model to notionally simulate the initiation and positioning of trichomes on leaves where representative changes in the parameters generate known mutants.

2. Background - Positional Information, Biological Modularization

The foundations on which the proposed recursive algebraic modelling is based are positional information, biological modularization and leaf morphogenesis. A brief summary of the necessary background for each is given below.

2.1. Positional Information

As explained by Wolpert (1969b), "The idea that the behaviour of a cell is determined by its position in the embryo" dates from the late 19th century experiments of Hans Driesch with sea urchin embryos. Driesch showed that such embryos developed normally even when parts were removed. Similar results have been shown to hold for the growth of plant organs out from the meristem (Steeves and Sussex (1989)).

From a plant perspective, it is necessary to establish what corresponds to the "embryos" for the various organs, as distinct from that of the seed. As explained below, it will be assumed that the tissue layer adjacent to the meristem boundary is where the embryonic positioning of plant organs occurs (i.e. the positional identity of plant cells is determined in this tissue layer). As Poethig (1997; p. 1078) has remarked "One function of the shoot meristem may be to make tissue that can make leaves rather than to make leaves".

For the identification of the simple protocols that evolution has discovered, developed and exploited in a wide variety of different ways, the positional information concept of Wolpert represents the natural framework in which to work. In essence, the concept asserts that as cells form and evolve, they acquire various positional identities in a coordinate system and then interpret this information according to their genetic constitution and developmental history. The appropriateness of this concept is reflected in the fact that it is not only biologically relevant, but also engenders simplification by localizing the mechanism of pattern formation.

In addition, it will be assumed that both the accumulation of positional information and its subsequent exploitation is a local neighbourhood activity of cells. This is possible because there is a copy of the genome in every cell with a nucleus. As explained by Pennell et al. (1995), "In plants, cells differentiate according to their position with relation to neighbouring cells.". Consequently, if it is thereby assumed that the fate of a cell is determined by its neighbours, a natural corollary is that the positional information rules, in order to be simple, should be algebraic. It also represents justification for turning to the modularization and mechanistic conceptualizations of Hartwell et al. (1999) and Sellers (1989).

2.2. Biological Modularization

"Cellular functions … carried out by 'modules'"
Hartwell et al. (1999).

The Hartwell et al. (1999) approach, which is utilized in the sequel, has an affinity with the ideas of Sellers (1989), but takes a quite different position to the algorithmic "executable biology" ideas of Fisher and Henzinger (2007).

From a mathematical modelling perspective, without some rationale for how to proceed, the simulation of genetic dynamics is a “dog’s breakfast”. However, the understanding of the genetics of biological processes has reached the point where the nature of how biology works acts as the guide. The modularization that biology itself exploits, in terms of conserved genes in stable genetic processes linked to other processes with less conserved genes lays the foundation as to how the modelling should be organized and performed.

This is astutely explained in Hartwell et al. (1999). Clearly, this modularization of biology represents independent validation of the concept of positional information as it supports the idea that biological processes operate in modules.

However, for the modelling, it is not only a matter of identifying the module the mechanism of which one is aiming to understand and quantify, but also the identification of the stages occurring in the process being modelled. As outlined below in the discussion of leaf morphogenesis, the change from one stage to the next corresponds to a change in the genetic dynamics and thereby to the need to identify the mechanism via which it has occurred.

2.3. Leaf Morphogenesis

Here, it is assumed that, biologically, the growth of a leaf involves, as a minimum, the following stages (Donnelly et al. (1999), Fleming et al. (2005, 2006), Tsukaya (2002))

1. The initial growth out of the meristem until the petiole (stem) forms. At this stage, the only significant expansion that occurs is that occurring at the meristem boundary between the tissue being formed, that will subsequently become the cells of the leaf, and the cells of the meristem. It is this initial growth that determines the basic shape and structure within a leaf. (It appears to play a homologous role similar to that of embryo formation in other organisms.)

2. The formation of the petiole. It is this stage that is the switch between leaf formation and leaf growth.

3. The leaf expansion. Partial differential equation modelling (e.g. Meinhardt (1982)) is more appropriate at this stage. Many, if not all, of the trichomes will have already formed.

Here, the focus is this first stage. It is assumed that the initial growth out from the meristem occurs in the following manner

(i) The basic shape and internal structure of a leaf are determined at the boundary between the cells that have fully formed as leaf cells and the cells within the meristem controlling the formation of the tissue layer that will next become the next layer of leaf cells (Poethig (1997)).

(ii) The positional information of the current tissue layer growing out from the meristem is determined by the signalling and communication occurring

between the cells that have fully formed and the cells within the meristem controlling the formation of the current tissue layer.

3. The Algebraic Modelling of Pea Leaf Morphogenesis

An example of a simple recursive algebraic protocol that successfully models an actual pattern formation activity, pea leaf morphogenesis, has been given by Young (1983). His model simulates, as a discrete dynamical system, the progressive growth of a pea leaf as an assemblage of leaflets and tendrils attached to a network of rachis axes. Here, the biological module is the morphogenesis of the pea leaf, with the various primordia that form corresponding to sub-modules.

The progressive growth of the pea leaf is modelled under the following assumptions:

A1. Progressive Dynamical System Growth

The size of the last primordium to form determines the next stage in the development. In this way, the sizes of the various primordia define the positional information associated with the subsequent development. Clearly, the size of a primordium (used below to determine the dynamics of growth) can be reinterpreted from various perspectives such as the concentration of some signalling hormone or enzyme.

A2. Initialization

The progressive growth starts with an initial primordium of size m_0 which

(i) grows in size according to the equation

$$m_1 = f_1(m_0),$$

then

(ii) subdivides in to three meristem-primordia (consisting of two lateral primordia of size m_2 and a central primordium of size m_3) to produce the basic rachis structure of Figures 2 and 3 in Young (1983) with

$$m_2 = f_2(m_1), \qquad m_3 = f_3(m_1).$$

For the model proposed by Young (1983),

$$f_1(m_0) = g.m_0, \qquad g = \text{constant},$$

$$f_2(m_1) = k.m_1 + c, \qquad k, c = \text{constants},$$

$$f_3(m_1) = m_1 - 2f_2(m_1).$$

In order to take account of the dynamics of the underling GSCS, some appropriate model for the switching must be defined. For Young (1983), the choice was whether or not various thresholds were exceeded. In particular, Young showed that, with only two thresholds T_1 and T_2, which depends on the concentrations of various gene products (specifically af and tl), the progressive growth model ccould simulate the growth of known mutants as well as the wild type. In fact, the importance, from an algebraic modelling perspective, of Young's model is its ability not only to simulate, through changes in the values of some key parameters, the growth of the wild type pea leaf but also of known mutants.

Remark 1. In many ways, a test for the appropriateness of an algebraic biological model should be its ability to simulate, with representative changes in key parameters, the morphology of both the wild type and known related mutants.

Remark 2. Modelling the switching in this way allows the growth of one branch of the peal leaf to be uncoupled from that of the other branches. Exemplification is given below in Rule 2. Though, there, the thresholds are assumed to be constant, they could be defined to depend on the past and/or current status of the growth.

A3. Growth

The dynamics of the growth is modelled using the rules:

Rule 1. The size of the next primordia to form is determined, in a dynamical manner, by given algebraic formulas defined only in terms of the sizes of the primordia from which they have just formed.

Rule 2. The disjoint dynamic stages in the subsequent development of primordia just formed are determined by the following size threshold constraints:

(a) $m_2 > T_1$, (or $m_3 > T_1$): the initiation repeats with m_2 (or m_3) as the new values for m_0.

(b) $T_2 < m_2$, $m_3 < T_1$: leaflet forms, and the process stops.

(c) $m_2 < T_2$ or $m_3 < T_2$: tendril forms, and the process stops.

An examination of the structure of this model of Young (1983) shows clearly the types of components required in order to simulate some modular biological process such as the positioning of trichomes on Arabidopsis leaves; namely,

Geometry: A protocol which defines the progressive stages of development (cf. **A1** above).

Initiation: A protocol which defines how the whole process starts.

Rules: Protocols which define how the GSCS operates.

4. A Framework for the Mechanistic Modelling of Plant Development

The above analysis of the methodology proposed by Young (1983) illustrates how to formulate a framework for the mechanistic modelling of some developmental process. Once the nature of the developmental process to be investigated has been articulated, the proposed algebraic framework involves the following components

(a) A Conceptualization about How the Development will be Modelled. The conceptualization that underlies, implicitly, Young's (1983) modelling is that the changing "sizes" (positive or negative) of various cellular components (hormones, transcription factors, etc) model the "signalling and communication" between cells with "switching" from one cell fate to another occurring when threshold conditions are satisfied. Because of its simplicity and appropriateness biologically, this conceptualization has a universal appeal and is the foundation on which the model formulation is performed below.

(b) The Geometry of the Development. The development of all biology organs follow a well-defined geometry the essential components of which must be identified and modelled. For the geometry of planar cellular structures in plants, the Voronoi tessellation is popular because of its visual appeal, optimality properties and ease of construction. In addition, from a geometrical "size" perspective, because the starting point for the tessellation's construction is the cell centres of the feature being visualized, the resulting structure is not unrepresentative of an actual planar structure. Biologically, its possible short-coming is the geometric assumption it imposes about the position and shape of cell boundaries.

(c) The Initiation. For the initiation protocol, one must turn to the known genetics for insight. At this point, the known specifics of the genetic dynamics of the situation under investigation must be brought into the picture. Even though, currently, there is a limited understanding about what the underlying mechanism might be, modelling can only progress once some representative hypothesis is made about what is actually occurring.

(d) The Development. From a biological modelling perspective, the identification of rules, that model the development, forces the need, in terms of the available genetic information, to categorize the genes in terms of their known properties.

In essence, this framework has a similar structure to that involved with the modelling of some evolutionary processes such as diffusion with step (a) taking on the role of the equation, step (b) the domain on which the evolution takes place, step (c) the initial conditions and step (d) the (numerical) solution process. The big difference associated with modelling biological development is that the domain changes shape.

5. A Categorization of Genes involved with Trichome Initiation and Positioning

The goal of the deliberations below is to illustrate how the above framework can be utilized to mechanistically model the initiation and positioning of the trichomes on Arabidopsis leaves. The key step, explained in some detail below, is the assessment of the available biological data to determine how they yield insight about the initiation and growth. In addition, this process gives insight about the limitations in the currently available biological data and, thereby, the planning of future mutant experiments.

Consequently, the subsequent discussion should be seen as a description of how such a framework needs to be applied in the identification of an algebraic mechanistic model that describes the genetic dynamics of the GSCS process being examined.

Assumption (a). On the basis of the above discussion, it will be assumed that the positional information for the epidermal cells in a (Arabidopsis) leaf is established in the boundary layer of tissue between the meristem (stem) cells and the epidermal cells that have become leaf cells. Consequently, as a cell forms at the meristem boundary to enter the tissue layer, the fate of the subsequent leaf cell, that it will become, is determined jointly by its neighbours above in the tissue layer (next to become leaf cells) and below in the meristem.

Assumption (b). Here, the geometry is defined as an evolving pattern of equally sized hexagonal cells. Pseudo-justification for hexagonal cells can be found in Larkin et al. (1996), Table 1, where it is recorded that the average number of cell boundaries for epidermal cells is 6. The utility of this assumption is the ease with which such an evolving pattern can be generated, and the multiple scale interpretation that can be given to such a pattern in which each hexagon can represent a cell or a neighbourhood of cells. Consequently, the leaf grows out from the meristem as a succession of new layers of tissue cells that subsequently become, above the tissue layer, either epidermal leaf cells or trichomes.

Assumption (c). For the initiation protocol, one must turn to the known genetics for insight. From Larkin et al. (2003), it is known that the principal genes involved with trichome initiation and development are

- GL1. "…highest levels of *gl1* … in early stages of developing trichomes…", "…increase in GL1 expression appears to proceed detectable expansion of trichome precursor cells", and "…the commitment to trichome fate involves a positive feedback loop regulating GL1 expression"

- GL2. "Loss-of-function *gl2* mutants produce small trichomes with reduced branching and aberrant expansion, …"

- GL3. "Loss-of-function *gl3* mutants produce a reduced number of trichomes that are smaller and have fewer branches than wild type."

- TTG. Null alleles of *ttg* have a "hairless phenotype"

- TRY. “In *try* mutants, trichomes arise adjacent to other trichomes at a much higher frequency than in wildtype …”, “*try* mutants which also constitutively overexpress GL1 … have more trichomes than plants that overexpress…” both TRY and GL1

- CPC. “Constitutive expression of CPC in leaves eliminates trichome production, and *cpc* mutants have an increased number of trichomes …”

- TRY and CPC. “Both *try* and *cpc* are expressed most strongly in developing trichomes and not in other epidermal cells, even though their function is to inhibit neighbouring cells from developing as trichomes.”

The next step is, on the basis as to what is currently known, from mutant, GFP and other experiments, about the developmental properties of these genes, is to hypothesize about their categorization in terms of their roles. In the modelling of genetic dynamics, this is a fundamental step as the categorization is the basis for the identification at the gene level of biological modules in the sense outlined in Hartwell et al. (1999).

For the categorization of the above genes, it is necessary to take account of the fact that trichome development involves not only the initiation of some event like a change of cell fate but also the subsequent control of that change in fate. In the case of trichome initiation, the subsequent control involves two aspects: the endoreduplication within the trichome and the control of the positioning of trichomes in the neighbourhood of one already formed. It is this framework that will be used here to perform the categorization of the above genes.

Assumption (d). The deliberations of Assumption (c) suggest, as a notional starting point for the explorative mechanistic modelling, the following conclusions on the basis of which the algebraic rules will be formulated:

C1. When the concentration of *gl1* in a cell just being formed in the tissue layer is sufficiently high, then that cell’s fate changes to be a trichome.

C2. The level of *cpc* controls the threshold that the concentration of *gl1* must exceed before a newly forming tissue cell will change its fate to become a trichome.

C3. The gene products *gl2* and *gl3* control the subsequent endoreduplication development of a trichome after its initiation.

C4. In the tissue cells next to form below a newly formed trichome cell, *ttg* and *try* reset the concentration of *gl1* to a lower level. In this way, they perform a polar inhibition. The cells about the newly initiated trichome already have their fate to be epidermal cells. However, during subsequent expansion of the leaf after the petiole forms, there is a possibility that new trichomes will initiate. This situation is not pursed here.

Both *gl1* and *gl3* must be in high concentration as the fate of an epidermal cell changes to be a trichome to ensure the normal development of the trichome.

6. An Algebraic Model for Trichome Initiation and Positioning

In this section, we propose a framework for the explorative recursive algebraic mechanistic modelling of GSCS. Specifically, we consider GSCS in the plant development process related to the positioning of trichomes on Arabidopsis leaves. The framework is motivated by Young's model for the pea leaf development discussed in Section 3.

First of all, we propose to represent the leaf as an array of hexagonal tiles (Figure 1). An individual tile can be viewed as a model of either a single cell or a collection of cells. Each tile has positional information quantitatively determined by a hexagonal number. The value of this number will be responsible for the switching that performs the trichome initiation. As in Young's model, it is assumed that this switching is controlled by a threshold condition, i.e. when the hexagonal number exceeds some threshold, a trichome will initiate in the corresponding tile. From the discussions in the previous section, it can be surmised that the hexagonal number is related to the concentration of the expression of the gene GL1. Therefore, we denote the value of the threshold by T_{gl1}.

In order to have an effective reference to the hexagonal numbers, we introduce the concept of a horizontal layer of tiles which is the collection of tiles with centres lying along the same horizontal line. Then, the hexagonal number is denoted by $P(i, j)$, where i is the position of the layer counting from the top, and j is the position of the tile within that layer counting from the left.

On the basis of the discussions in Section 2.1, a tile, being formed in the tissue layer, reads the positional information of its neighbours above, and, depending on the situation, responds in an appropriate manner. The neighbours, of a newly forming tile (i, j), are illustrated in Figure 2. In order to have an effective reference for them, let $n_l = n_l(i, j)$ and $n_r = n_r(i, j)$ denote the positions of the tiles, on layer $(i-1)$, to the left and right of this tile, and $n_c = n_c(i, j)$ denote the position, on layer $(i-2)$, of the tile immediately above the *(i,j)*-tile (see Figure 2 for illustrations).

After a tile has formed with its positional value $P(i, j)$, the genetics within that cell responds to the value of $P(i, j)$. As already indicated, a tile switches to be a trichome if the value of $P(i, j)$ exceeds an appropriate threshold, such as T_{gl1}. The actual value of $P(i, j)$ is determined in the following manner. The hexagonal number in the top tile has some initial value P_0 that is below the threshold. Thus, in the top tile no trichomes initiates. Also, it is known (Schellmann, et al. (2002)) that trichomes do not initiate on the leaf boundary. It is therefore assumed that the hexagonal numbers in the lateral tiles (i.e. tiles located on the boundary) also take the value P_0. The values of $P(i, j)$ for the inner tiles will, in general, increase in some systematic manner.

On the basis of the discussion in Section 5, in particular Assumption (d) C4, the hexagonal numbers of the tiles in the tissue layer, just below a tile with a trichome, will be reset to the initial value P_0.

For the $P(i, j)$ values in the tiles not adjacent to a newly initiated trichome, it is assumed that the hexagonal number in the tile is equal to the weighted sum of hexagonal numbers in its neighbours above, and, on the basis of that value either switches to become a trichome or keeps that value.

There is a duality in the interpretation of how a given tile (cell) assumes its hexagonal number value $P(i, j)$. Either the tile above the tissue layer imposes the value on the tiles forming below in the tissue layer, or a tile forming below in the tissue layer assumes its value on the basis of how it interprets the situation in the tiles just formed above. It is the latter interpretation which will be assumed in the discussion below.

The above considerations can be formally summarised in the following recursive rules for determining values of $P(i, j)$:

1. $P(1,1) = P(2,1) = P(2,2) = P_0$.
2. If the tile (i, j) is located on the periphery of the leaf, then $P(i, j) = P_0$.
3. If the $P(i, j)$ value in *at least* one of the neighbours above the tile (i, j) exceeds some threshold T_{gl1} (i.e. if $P(i-1, n_l) \geq T_{gl1}$, or $P(i-1, n_r) \geq T_{gl1}$, or $P(i-2, n_c) \geq T_{gl1}$; this also means that trichome growth has commenced in at least one of the neighbours above), then the hexagonal number in this tile is reset to P_0 (i.e. $P(i, j) = P_0$). Otherwise, $P(i, j)$ is determined by the following algebraic formula:

$$P(i, j) = w_l P(i-1, n_l) + w_c P(i-2, n_c) + w_r P(i-1, n_r),$$

where w_l, w_c, w_r are some constants satisfying $w_l + w_c + w_r \geq 1$.

In the situations considered below, it is assumed that $w_l = w_c = w_r = 1$ and $P_0 = 1$.

As shown in Figure 3, the algorithm starts with three hexagonal tiles (Figure 3(a)) and then progressively generates successive layers with the hexagonal numbers determined according to the above rules. Figures 3(b)-(h) show this progression at different stages. The trichomes will initiate in the tiles whose hexagonal number exceeds the threshold.

Remark 2. Following the discussions in the previous section, the value of the threshold T_{gl1} is probably determined by *cpc*.

Remark 3. The algebraic properties of hexagonal numbers are explored in O'Keefe et al. (2008).

7. Results and Conclusions

In Figure 4, it is shown how the regular pattern of trichome initiation, using the above rules, is sensitive to the value of the threshold T_{gl1}. Figure 4(a) shows the hexagonal leaf pattern with trichomes everywhere. Figures 4(b)-(f) show the changing pattern of the trichome initiation when the threshold T_{gl1} varies. As the value of T_{gl1} increases, the number of trichomes decreases (non-monotonically), the positions of the trichomes move down from the leaf tip and the spacing between the various trichomes increases.

The proposed framework has the potential to provide models that simulate not only the development of the wild type but also of known and unknown mutants. This is achieved through changes in the threshold values that determine the switching from one cell fate to another. This is fully consistent with the known biology of development.

If one assumes, as here, that the positional information and the fate of the cells on a leaf are determined in the tissue layer between the leaf cells and the meristem (stem) cells, then the inhibition, at the embryo leaf growth stage, needs only be polar in the direction of the newly forming cells at the meristem boundary.

The overall goal of the above computational modelling is an illustration of how, for an algebraic model of some biological process, to utilize, in an iterative manner, the available biological knowledge in the formulation of a model that captures the essence of the biology being investigated. In particular, it illustrates how, when all factors are taken into account, the resulting model, though quite elementary, is able to generate a plethora of possibilities as occurs in the genetic manipulation of plants.

Consequently, the analysis and interpretation of such a model is not necessary straightforwardly simple. For example, there is a need, on the basis of published results about trichome initiation, positioning, and development, to give more precise biological interpretation of the threshold(s) and to modify the model. In addition, the relationship between the number of generated trichomes and the value(s) of the threshold(s) requires further investigation, as does the dependence of this relationship on the total number of cells.

Acknowledgements. Both authors wish to acknowledge the financial support received from RICAM to undertake this research during the 2007-2008 Special Semester on Quantitative Biology Analysed by Mathematical Methods. The second author wishes to acknowledge the helpful advice received from various CSIRO colleagues including Christine O'Keefe, Peter Waterhouse and John Watson.

References

P. M. Donnelly, D. Bonetta, H. Tsukaya, et al. (1999) Cell cycling and cell enlargement in developing leaves of Arabidopsis, Develop. Biol. 215 (1999), 407-419.

J. Fisher and T. A. Henzinger (2007) Executable cell biology, Nature Biotech. 25 (2007), 1239-1250.

A. J. Fleming (2006) Leaf initiation: The integration of growth and cell division, Plant Molecular Biology 60 (2006), 905-914.

Fleming (2005) The control of leaf development, New Phytologist 166 (2005), 9-20.

L. H. Hartwell, J. J. Hopfield, S. Leibler and A. W. Murray (1999) From molecular to modular cell biology, Nature 402 (1999), C47-C52.

M. Hulskamp (2004) Plant trichomes: A model for cell differentiation, Nature Reviews 5 (2004), 471-480.

H. Jonsson, M. G. Heisler, B. E. Shapiro BE, et al. (2006) An auxin-driven polarized transport model for phyllotaxis, Proc. Nat. Acad. Sci. USA 103 (2006), 1633-1638.

C. Kuhlemeier (2007) Phyllotaxis, Trends in Plant Science 12 (2007), 143-150.

J. C. Larkin, M. L. Brown and J. Schiefelbein (2003) How do cells know what they want to be when they grow up? Annu. Rev. Plant Biol. 54 (2003), 403-430.

J. C. Larkin, N. Young, M. Prigge and M. D. Marks (1996) The control of trichome spacing and number in Arabidopsis, Development 122 (1996), 997-1005.

H. Meinhardt (1982) Models of biological pattern formation, Academic Press, 1982.

P. C. Milner (1964) The possible mechanisms of complex reactions involving consecutive steps, J. Electrochem. Soc. III (1964), 228-232.

C. M. O'Keefe, S. Pereverzyev Jr. and R. S. Anderssen (2008) The algebra of hexagonal numbers, preprint.

R. I. Pennell, Q. C. B. Cronk, L. S. Forsberg, et al. () Cell-context signalling,

R. S. Poethig (1997) Leaf morphogenesis in flowering plants, The Plant Cell 9 (1997), 1077-1087.

P. Prusinkiewicz and A. Lindenmayer (1990). *The Algorithmic Beauty of Plants*. Springer-Verlag, 1990.

S. Schellmann, A. Schnittger, V. Kirik, et al. (2002) TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in Arabidopsis, EMBO J. 21 (2002), 5036-5046.

P. H. Sellers (1989) Combinatorial aspects of enzyme kinetics, in Applications of Combinatorics and Graph Theory to the Biological and Social Sciences, Editor – F. Roberts, Springer-Verlag, 1989.

R. S. Smith, S. Guyomarc'h, T. Mandel, et al. (2006) A plausible model of phyllotaxis, PNAS 103 (2006), 1301-1306.

T. A. Steeves and I. M. Sussex (1989) Patterns in Plant Development, Cambridge University Press, Cambridge, 1989.

H. Tsukaya (2002) Leaf development, in The Arabidopsis Book, Amer. Soc. Plant Biologists, 2002.

L. Wolpert (1969a) Positional information and spatial pattern of cellular differentiation, J. Theor. Bio. 25 (1969a), 1-74.

L. Wolpert (1969b) One hundred years of positional information, Trends in Genetics 12 (1969b), 259-364.

J. P. W. Young (1983) Pea leaf morphogenesis: A simple model, Annals Bot. 52 (1983), 311-316.

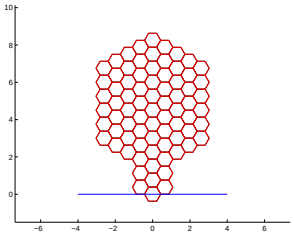

Figure 1. Representation of the leaf as an array of hexagonal tiles.

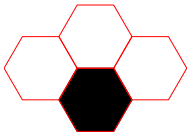

Figure 2. A tile (i, j) (black one) and its left $(i-1, n_l)$, right $(i-1, n_r)$, and central $(i-1, n_c)$ neighbours.

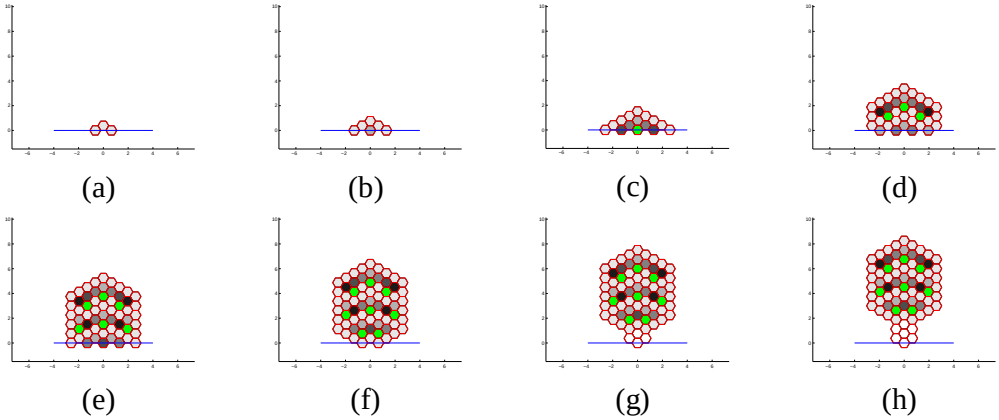

Figure 3. An example of leaf development at different stages. Hexagonal numbers are determined according to the introduced rules. The color of the tile represents the corresponding hexagonal number with white corresponding to zero and black to some maximal value. The tiles where the hexagonal number exceeds the prescribed threshold are marked green. In this example $T_{gl1} = 10$.

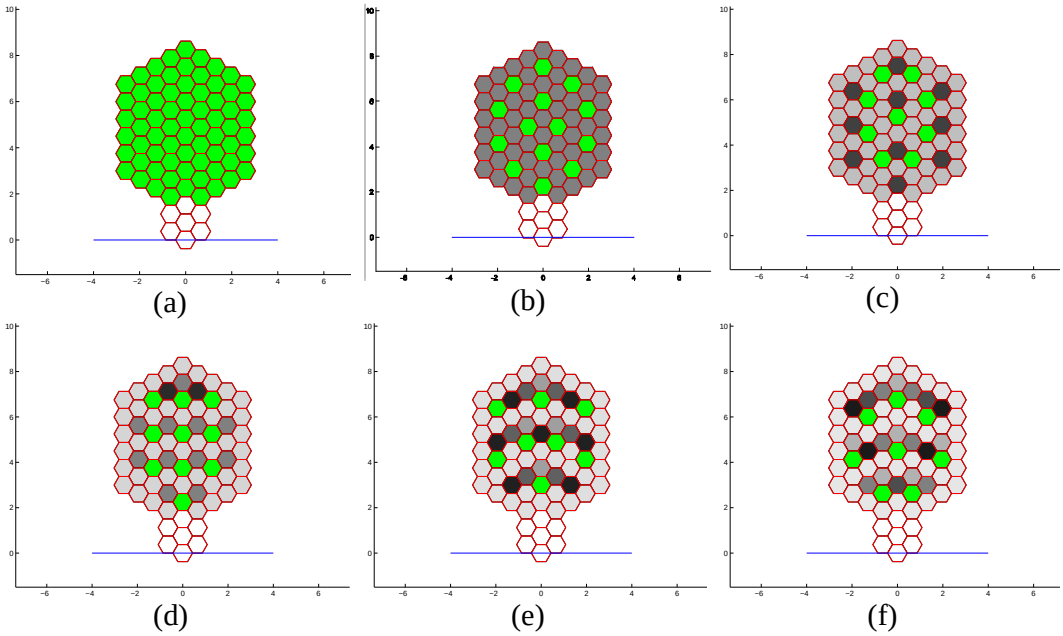

Figure 4. Final distribution of hexagonal numbers for different values of the threshold T_{gl1}. (a) $T_{gl1} = 1$; (b) $T_{gl1} \in [2,3]$; (c) $T_{gl1} \in [4,5]$; (d) $T_{gl1} \in [6,7]$; (e) $T_{gl1} \in [8,9]$; (f) $T_{gl1} \in [10,13]$.