


Available online at [www.sciencedirect.com](http://www.sciencedirect.com)SCIENCE  DIRECT®

Agricultural Systems xxx (2004) xxx–xxx

---



---

 AGRICULTURAL  
SYSTEMS
 

---



---

[www.elsevier.com/locate/agsy](http://www.elsevier.com/locate/agsy)

## Merging genomic control networks and soil-plant-atmosphere-continuum models <sup>☆</sup>

S.M. Welch <sup>a,\*</sup>, J.L. Roe <sup>b</sup>, S. Das <sup>c</sup>, Z. Dong <sup>d</sup>  
R. He <sup>e</sup>, M.B. Kirkham <sup>f</sup>

<sup>a</sup> Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA<sup>b</sup> Division of Biology, Kansas State University, Manhattan, KS 66506, USA<sup>c</sup> Department of Electrical and Computer Engineering, Kansas State University, Manhattan, KS 66506, USA<sup>d</sup> Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA<sup>e</sup> Department of Computing and Information Sciences, Kansas State University, Manhattan, KS 66506, USA<sup>f</sup> Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA

Received 11 November 2003; received in revised form 2 June 2004; accepted 23 July 2004

---

### Abstract

Advances in genomic science make it desirable to include genomic controls in soil-plant-atmosphere-continuum (SPAC) models by methods proposed in this paper. Molecular genetic concepts suggest that a differential equation similar to ones used in neural networks can be used to model single-gene elements of larger systems. Natural modifications to the equation incorporate temperature dependency. Multi-gene components based on this element function as Boolean logic gates, linear arithmetic units, delays, differentiators, integrators, oscillators, coincidence detectors, and bi-stable devices. Related genetic circuitry from real organisms is shown. Genomic integration with SPAC models entails whole-plant modeling with realistic morphology. Plants are networks of parts, iterated in time and space under genetic control, that induce and modulate conservative SPAC mass/energy flows. Network developmental

---

*Abbreviation: ABA, abscisic acid; OOP, object-oriented programming; SVAT, soil vegetative atmosphere transfer; SPAC, soil plant atmosphere continuum; LD, long day; SD, short day; DNA, deoxyribonucleic acid; mRNA, messenger ribonucleic acid*

<sup>☆</sup> Contribution number 03-81-J from the Kansas Agric. Exp. Stn., Kansas State Univ.

\* Corresponding author. Tel.: +1 785 532 7236; fax: +1 785 532 6094.

E-mail address: [welchsm@ksu.edu](mailto:welchsm@ksu.edu) (S.M. Welch).

rules can be stated as Lindenmayer grammars whose symbols represent plant parts programmed as software objects. A structure is presented for simulators based on these concepts. The discussion argues that prior object-oriented plant modeling approaches (i) do not reflect how plants actually develop morphologically and (ii) may represent processes in tactically unwise ways at a time when genomics is advancing knowledge of process interactions. Finally, genomics and expanding computing power redefine concepts of model “simplicity” and “complexity” to favor increased realism.

© 2004 Elsevier Ltd. All rights reserved.

*Keywords:* L-systems; Development; Simulation; Physiology; Environmental physics; Genetics; Object-oriented programming

---

## 1. Introduction

### 1.1. Motivation

In 1994, the American Society of Agronomy held a symposium on the status of crop modeling. In published comments Passioura (1996, p. 690) criticized mechanistic models as having “failed to meet their aspirations” because they are based on “untestable guesses about the processes that control growth”. Monteith (1996, p. 696) said that, if Aristotle’s advice against seeking undue precision were followed, all crop modeling would be put “on hold until we could describe: (i) the [*governance of stage-specific assimilate allocation*] and (ii) [*root*] uptake ... in relation to ... growth, anatomy, and activity”. Others were less harsh. Sinclair and Seligman (1996, p. 701) said models “should be viewed ... as heuristic tools” and Boote et al. (1996) gave examples of model use but also suggested “cautions and limitations” (p. 704). All of these views are more tentative than perceptions in hydrology, pollutant transport, and meteorology, wherein model use is “routine, if not indispensable” (Baker, 1996, p. 689).

It is noteworthy that the first two authors, with expertise ranging from physiology to physics, both allude to process control issues. Plants are not passive responders to the environment but operate under controls at least as elaborate as those in advanced, cybernetically managed machines (Csete and Doyle, 2002). This is in contrast to current models, which often behave less plastically than real plants (Welch et al., 2003). As a result, some have suggested that improved process controls should be a modeling priority (Hammer, 1998). The core of plant process control lies in the genome. Combining analog and digital features, gene switching between active and inactive states modulates metabolism, initiates biosynthesis of (and response to) hormones, integrates sensory data, controls morphological differentiation, and regulates other functions too numerous to list.

Initial efforts to integrate genomics and crop simulation have targeted plant breeding in order to model more realistically the effects and potential gains from selection (Messina, 2003). White and Hoogenboom (1996) regressed genetic coefficients in a dry bean (*Phaseolus vulgaris*) model on the presence of dominant alleles

at seven loci affecting phenology, growth habit, and seed size. Chapman et al. (2003) simulated sorghum breeding program alternatives by merging QU-GENE (Podlich and Cooper, 1998) with the APSIM crop model (McCown et al., 1996). Additional activities are described by Weiss (2003).

There is, however, another reason for integrating genomics that creates an opportunity for a broader synthesis. As general circulation models have improved, increasingly realistic sub-grid cell soil-vegetation-atmosphere-transfer (SVAT) models compute plant-controlled air/surface transfers of radiation, moisture, sensible heat, and momentum (Budyko, 1974; Dickenson, 1984; Sellers et al., 1986, 1996a,b; Arora, 2003). SVAT models: (i) focus on biosphere–atmosphere exchanges rather than economic yields, but (ii) have major physiological and morphological commonalities with crop models, while (iii) being more strongly grounded in physical principles.

Climate change modelers realize that decade- to century-length simulations will need to incorporate biogeographical shifts in plant distribution (Arora, 2003). Less recognized are possible functional changes due to plant evolution. For example, Ward et al. (2000) showed in a five-generation experiment with *Arabidopsis thaliana* (L.) Heynh that  $2 \times \text{CO}_2$  appeared to select for a shorter lifespan and unchanged or reduced biomass. This result could not have been predicted from earlier, single-generation studies on  $\text{C}_3$  plants. Combining genomics and SVAT models would enable both improved representation of process control mechanisms as well as incorporation of selection effects. In addition, the richer use of physics already present in SVAT models could improve realism across all plant modeling applications including crop breeding.

## 1.2. Rationale

In this paper, we propose a framework for merging genomic control into plant process models. However, rather than simply attempting to graft genomics onto current crop simulation or SVAT models, both of which have become quite specialized, we instead begin near their common ancestors. In a seminal paper, Philip (1966) coined the phrase “soil-plant-atmosphere-continuum” (SPAC) and cataloged the “minimum information [*for*] a determinate nonstationary model of energy and water transfer in the SPAC” (p. 247). Almost all listed items are found in modern SVAT models, and not a few of the fundamentals (including the basic decomposition into soil, plant, and environment) are also present in crop simulation models. Philip, himself, saw SPAC models as serving both agronomic and ecological ends (1966, p. 248). Therefore, in the spirit of a “back to basics” approach, we shall refer to “SPAC models” from here on.

Geometry is a frequent topic in Philip’s paper (1966, pp. 247, 250, 252–253, 255, 256–257, and 262). Indeed, he views geometric influence as so pervasive that “its neglect appears to have led a substantial body of work related to the SPAC into unreality” (p. 247). Beyond just flows of water and nutrients, morphology also influences light interception, the biomass and location of resource sinks/sources, etc. While geometric realism has certainly improved since 1966 (e.g., by layering schemes), most models remain crude morphological approximations. This complicates

incorporation of genomics since gene expression is tissue specific as well as environmentally responsive so inappropriate spatial positioning can lead to inaccurate control simulations.

Three stated reasons for limiting morphological detail are: (i) more complex models have “vastly greater data requirements” unlikely to be offset by information gains (Jones, 1978, p. 622), (ii) more intricate models would be difficult to apply in multiple situations (ibid.), and (iii) geometric realism “would add greatly to the difficulty of the formal mathematical problem of the SPAC” (Philip, 1966, p. 250). All three reasons are based on outdated assumptions. The first is that morphological complexity automatically equates to data complexity. It is now known that simple, repetitively applied rules can generate detailed morphology (Mandelbrot, 1983; Prusinkiewicz and Hanan, 1989). In plants, morphology is the result of a developmental sequence of genetic state transitions – thus, the “rules” operate at the genomic level. [Although genomic, these “rules” can be triggered physiologically – e.g., phytochrome cycling alters patterns of gene expression to yield different seedling phenotypes (Casal and Sanchez, 1998, p. 322)].

The second assumption is that morphology must be a hard-wired model feature, leaving a least-common-denominator structure (e.g., roots, stems, and leaves) as the only available approach. In reality, nothing prevents a suitably constructed set of rules from generating intricate, environmentally dependent structures. Rather than manually building detailed morphology, researchers can focus on control mechanisms (genomic, physiological, and physical-constraint based) and let the computer produce the resulting geometric complexity. This paradigm closely parallels reality in which a phenotype results from a genetic control program executing in a given environment. Lindenmayer systems (Lindenmayer, 1968; Prusinkiewicz and Hanan, 1989; Mech and Prusinkiewicz, 1996) provide a well developed, concise encoding of genesis rules for plant structure.

The third premise is that only human beings can write equations and organize their numerical solution. Under this assumption, realism is defeated by the numbers of leaves, stems, and roots for which equations are needed. However, mathematicians and engineers routinely use software packages (e.g., Macsyma<sup>TM</sup>, Mathematica<sup>TM</sup>, and Maple<sup>TM</sup>) that derive equations symbolically. Commanded to “differentiate Eq. (2) and substitute the result in (7)”, these programs produce a formula as a result and will compute its numeric solution if required. Thus, the genetic “rule” that triggers differentiation of a new leaf could also produce new equations to represent it in a SPAC model that enlarges automatically as development progresses. Whether done symbolically or otherwise, we call this process of model alteration *equation management*.

The sections that follow expand these ideas. Tutorial and review information describe basic gene function and present a simple differential equation model. We have extended this model to incorporate thermal effects and demonstrated the ability of small sets of genes to implement elementary signal processing operations. This new material is presented next. Whole plants can then be modeled as networks of parts whose creation and functioning are under genomic control. Lindenmayer systems can describe the former while object-oriented programming can implement the

latter. Both formalisms are reviewed. We have created a way to meld the two technologies that we call L/OOP. A simplistic model demonstrates the functions comprising any L/OOP simulation. It should be understood that scaling the example up to realistic physiology is direct, albeit lengthy. The final section: (i) relates L/OOP principles to current schemes of crop model organization and (ii) discusses alternative views of model simplicity and complexity.

## 2. A single gene switching model

### 2.1. System identification

Genes are sections of a DNA molecule that encode the construction of proteins and may be active or inactive at any point in time. If active, the coded instructions are *transcribed* into RNA strands that communicate information used to synthesize the protein product. Lewin (2000) and Buchanan et al. (2000) give comprehensive introductions to the multi-step process. Drawing on Fry and Peterson (2002), we provide just enough detail to motivate the mathematical model that follows.

In plants and other eukaryotes DNA wraps around complexes of modified histone molecules to form *nucleosomes*. Further winding creates higher-level spirals collectively called *chromatin*. Often DNA in this configuration cannot be accessed by transcription mechanisms, so the coiled genes are inactive (Fig. 1(a)). Activation is a series of biochemical events that vary from gene to gene within an overall pattern. Various molecules attach to the *promoter region*, an area of DNA upstream of the segment coding for the protein. Such attaching molecules are currently the focus of intensive research (Horn and Peterson, 2002; Kouzarides, 2002; Peterson, 2002; Reyes et al., 2002). First chromatin is *remodeled* so that the DNA relaxes into a more exposed shape. Attachment sites are created for molecules arriving later (Fig. 1(b)). Last to attach are the components of the transcription mechanism, including the enzyme RNA polymerase (Fig. 1(c)). With all elements in place, transcription begins (Fig. 1(d)), producing a strand of *messenger RNA* (mRNA) that ultimately guides protein synthesis. An *expressed* gene is one being transcribed and protein production from mRNA templates is *translation*. Proteins and mRNA usually degrade in time, so fixed concentrations require ongoing transcription. Genes that are always active are said to be *constitutively* expressed.

The molecular events preceding transcription provide a means for gene regulation. If any of the molecules involved is (or becomes) unavailable, transcription cannot begin (or continue) and the gene is inactive. Alternatively, *repressor* molecules may bind to the promoter and block transcription. Transcription, once started, can be accelerated or slowed by other chemical species. Collectively, molecules that modulate transcription rates are called *transcription factors*. Of course, transcription factors are also gene products whose function may be regulated. Thus, genes control each other's activity in ways best represented as a *graphical model* or network (Martinez-Zapater et al., 1994; Koornneef et al., 1998; Simpson et al., 1999; Blazquez, 2000).

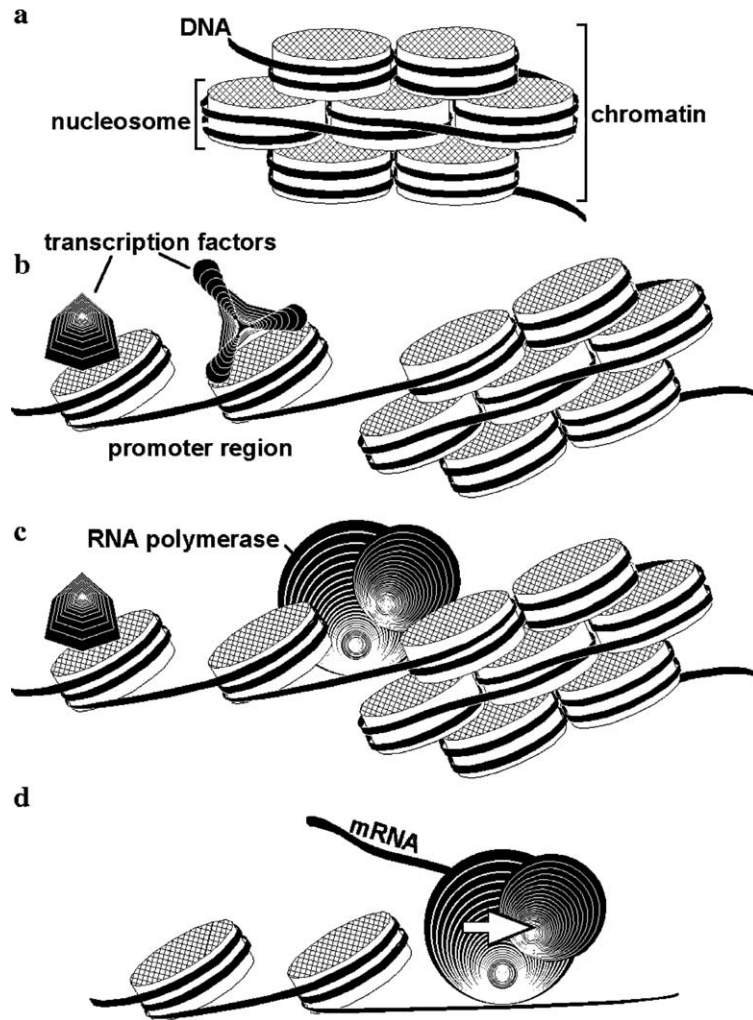


Fig. 1. Transcription steps. (a) Inactive genes with their DNA coiled around histone complexes form nucleosomes. (b) Early transcription factors remodel chromatin to make DNA accessible. (c) Subsequent attachment of the RNA polymerase transcription mechanism. (d) Transcription begins, producing mRNA.

## 2.2. A basic model

Graphical models are qualitative and cannot estimate outcomes of experiments other than the ones they are based on. In contrast, the intelligent systems community has many ways to annotate networks mathematically that permit quantitative prediction. Bioinformaticists are exploring these *connectionist* schemes to balance tractability, utility, and realism in dynamic gene expression modeling. Examples include: (i) Boolean (ON/OFF) networks (Frank, 1998; Liang et al., 1998; Mendoza and

Alvarez-Buylla, 1998, 2000; Szallasi and Liang, 1998; Samsonova and Serov, 1999; Akutsu et al., 2000; Ideker et al., 2000; Maki et al., 2001), (ii) Petri (concurrent information flow) nets (Goss and Peccoud, 1999; Matsuno et al., 2000), (iii) S-systems (continuous time models motivated by chemical kinetics) (Liang et al., 1998; Akutsu et al., 1999; Tominaga et al., 1999; Akutsu et al., 2000; Maki et al., 2001), (iv) differential equation models (Wolf and Eeckman, 1998; Chen et al., 1999), (v) neural network models (Reinitz and Sharp, 1995; D’Haeseleer et al., 1999; Weaver and Workman, 1999; Marnellos et al., 2000), and (vi) Bayesian networks (Friedman et al., 2000; Barash and Friedman, 2001; Hartemink et al., 2001).

In one of the simplest differential equation models, state variables represent dynamic, normalized (*ergo* dimensionless) mRNA and protein levels (Baldi and Hatfield, 2002, p. 151). Such values are obtained from Northern or Western blot analysis, rt-PCR readings, and microarrays. All biochemical levels are assumed to represent active forms. Let  $P$  be such a level and assume that  $R$  and  $\lambda P$  (both nonnegative) are, respectively, the production and degradation rates of  $P$  per unit time. Let  $\mathbf{g}$  be a dimensionless factor relating other gene products to production of  $P$ . Then,

$$\frac{dP}{dt} = R\mathbf{g} - \lambda P. \quad (1)$$

Because genes exhibit ON/OFF and intermediate behavior,  $\mathbf{g}$  is assumed to be a monotone, zero-to-one *transfer function* whose domain is an index of net transcriptional effect. A simple measure is the linear form  $\beta_0 P_0 + \beta_1 P_1 + \dots + \beta_n P_n$  where the  $P_i$ s are the levels of gene products or external inputs affecting  $P$  expression directly and the  $\beta_i$ s are effect strengths ( $>0$  for promotion and  $<0$  for repression). The  $\beta_0$  value relates constitutive influences, if present, and  $P_0$  (identically 1) represents their (possibly multi-gene) source. The single gene model can be written compactly as

$$\frac{dP}{dt} = R\mathbf{g}\left(\sum_{i=0}^n \beta_i P_i\right) - \lambda P. \quad (2)$$

This “rough network model” (Chen et al., 1999) has some shortcomings, such as a limited provision for protein–protein interactions. Suppose proteins  $A$  and  $B$  react as  $A + B \rightarrow AB$ , and  $AB$  regulates gene  $C$ . If (but only if)  $A$ ,  $B$ , and  $C$  all exhibit graded responses, the dynamics are controlled by the nonlinear Law of Mass Action (Waage and Guldberg, 1864), which the model can only approximate. Lesser issues are that: (i) degradation rates are independent of other gene states and (ii) the model averages over intra-cellular spatial compartmentalization that could slow interaction rates. However, examples below successfully adapt Eq. (2) despite these formal difficulties.

Eq. (2) also has major advantages. First, it is similar to the most common continuous time neural network (see Pearlmutter, 1995, p. 1213, Eqs. (3) and (4)). Thus, many parameter estimation methods used with neural networks also apply here (Pearlmutter, 1995). For example, Welch et al. (2003) used a common neural networking transfer function,  $\mathbf{g}_{\text{NN}}(x) = 1/(1 + \exp(-x))$ , in a static model of flowering time control in *A. thaliana*. Second, for some  $\mathbf{g}$ s, Eq. (2) has analytic solutions, at least over certain ranges. Closed form solutions display the entire behavior of a

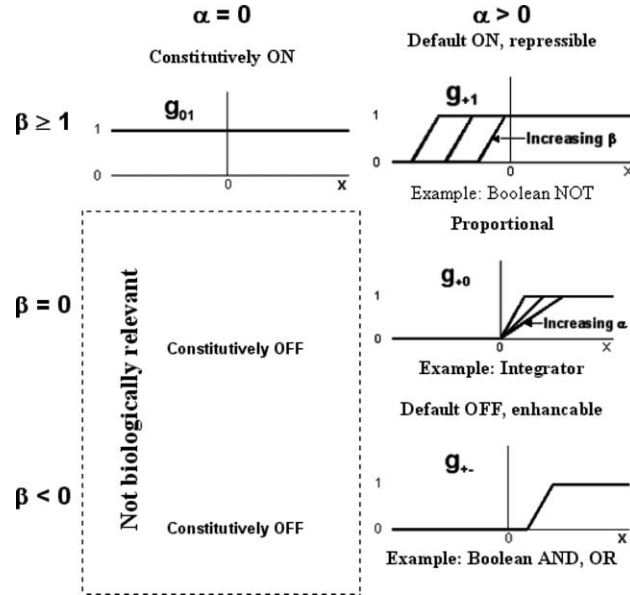


Fig. 2. Useful forms of  $g_{x\beta}$ . Graphs of each form are shown along with parameter effects, and operations that use them. The default status of  $g_{+1}$  and  $g_{+-}$  refers to their value given a zero input.

process. This visibility can reveal phenomena that might escape notice in a finite number of simulation runs. Two analyses below use a  $g$  with three linear segments

$$g_{x\beta}(x) = \max(\min(\alpha x + \beta, 1), 0), \quad (3)$$

where  $\beta$  is  $\beta_0$  in Eq. (2),  $x = \beta_1 P_1 + \dots + \beta_n P_n$ , and  $\alpha (\geq 0)$ , is the slope of the middle (rising) segment of  $g$ . Fig. 2 names some  $g_{x\beta}$ s according to their use.

### 2.3. Thermal effects

Energy balances, the resulting temperature distributions, and their biological effects are key parts of SPAC models. However, aside from shock and vernalization studies, most molecular geneticists have viewed temperature only as a variable to be optimized for experimental throughput (exceptions are Blazquez et al., 2003; Halliday et al., 2003). We have extended Eq. (2) to include temperature. Assume that product-related processes are rate-limiting so that temperature only affects  $R$  and  $\lambda$  (Welch et al., 2003). Many mathematical expressions can be used to relate  $R$  and  $\lambda$  to temperature. Johnson and Thornley (1985) categorize several by process type: diverse chemical reactions, diffusion, viscosity, translocation, and lipid phase changes. Some expressions exhibit an optimum temperature, while others that do not may still be useful within restricted temperature ranges.

Mathematical derivations based on Eq. (2) often involve products and quotients of  $R$  and  $\lambda$ . For example, if  $P$  is fully ON ( $g = 1$ ), its maximum expression level is  $R/\lambda$  and it traverses fixed subintervals of  $[0, R/\lambda]$  in times proportional to  $1/\lambda$ . Therefore,

functions whose forms remain the same under multiplication and division simplify analysis. We shall call this property *form invariance*. Three examples from chemistry and biology are Arrhenius factors (Johnson and Thornley, 1985, p. 122),  $Q_{10}$  values (ibid.), and beta functions (Yin et al., 1995):

$$\begin{aligned} A_T(\kappa, \alpha) &= \kappa e^{(-\alpha/T)}, \\ Q_T(\kappa_r, Q_{10}, T_r) &= \kappa_r Q_{10}^{(T-T_r)/10}, \\ B_T(\kappa, \alpha, \beta, T_L, T_U) &= \kappa(T - T_L)^\alpha (T_U - T)^\beta, \end{aligned} \quad (4)$$

where  $T$  is temperature in °K for  $A_T$  and °C elsewhere.  $T_L$ ,  $T_r$ ,  $T_U$  are, respectively, lower, reference, and upper threshold temperatures assumed to be constant within submodels. All other symbols may differ by gene. Quotients and products of  $A_T$ ,  $Q_T$ , and  $B_T$  have different parameter values but the same form. That is,

$$\begin{aligned} A_T(\kappa_1, \alpha_1)/A_T(\kappa_2, \alpha_2) &= A_T\left(\frac{\kappa_1}{\kappa_2}, \alpha_1 - \alpha_2\right), \\ Q_T(\kappa_{r,1}, Q_{10,1})/Q_T(\kappa_{r,2}, Q_{10,2}) &= Q_T\left(\frac{\kappa_{r,1}}{\kappa_{r,2}}, \frac{Q_{10,1}}{Q_{10,2}}\right), \\ B_T(\kappa_1, \alpha_1, \beta_1)/B_T(\kappa_2, \alpha_2, \beta_2) &= B_T\left(\frac{\kappa_1}{\kappa_2}, \alpha_1 - \alpha_2, \beta_1 - \beta_2\right), \end{aligned} \quad (5)$$

where the submodel-wide constants  $T_r$ ,  $T_L$ , and  $T_U$  are omitted to simplify notation.

$A_T$  and  $Q_T$  increase strictly with temperature.  $B_T$  can take many useful shapes including constant, linear, quadratic, and both symmetric and skewed unimodal curves. Other temperature functions (e.g., for transition state theory) have the form  $T^{n/2}A_T(\kappa, \alpha)$  where  $n$  is a non-zero integer (Johnson and Thornley, 1985). Although algebraically lacking form invariance, their curves are nearly identical to  $A_T$  at biological temperatures. Thus, we assume hereafter that  $R$  and  $\lambda$  are form-invariant temperature functions.

#### 2.4. General comments

The importance of bridging between the plant genome, phenome, and environment cannot be overstated (Cooper et al., 2002; Arora, 2003). Studies of cellular processes are informing these issues and have additional relevance to urgent topics ranging from health to security. In response, some researchers are pursuing the goal of complete *virtual cell* simulations (Tomita et al., 1999). This bottom-up approach uses more realistic biochemical kinetic equations than those presented above and may be expected to pay significant dividends in time. However, (i) high-throughput instruments to monitor eukaryote biochemistry are still in the development stage (Ezzell, 2002; Fiehn, 2002; Phelps et al., 2002) and (ii) the computing technology needed to make such simulation routine and applicable at the organism level is years away (Butler, 1999). Thus, it is reasonable to exploit an approximate but considerably more tractable model, while remaining ready to extend it as necessary. Our

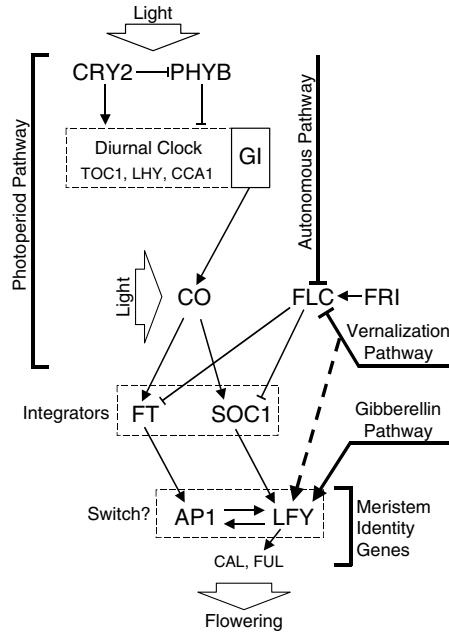


Fig. 3. *A. thaliana* flowering control (citations in the text). Heavy lines bracket or represent key gene sets including pathways that sense photoperiod, delay for winter (vernalization), respond to hormones (gibberellin), and generate default flowering signals (autonomous). Integrator genes merge autonomous and photoperiod path outputs. Meristem identity genes define the vegetative and reproductive states. The dashed boxes show putative component functions.

analyses show that genes behaving according to Eq. (2) can perform many useful tasks when linked into small networks: Boolean logic gates, linear arithmetic units, delays, differentiators, integrators, oscillators, coincidence detectors, and bi-stable devices. Three component types, putatively identified by dashed boxes in Fig. 3, are discussed next using results from various species. Table 1 lists gene abbreviations and selected phenotypic traits.

### 3. Genetic circuit components

#### 3.1. Oscillators

Organisms have endogenous rhythms driven by a diurnal clock based on negative molecular genetic feedback loop(s). If a set of genes necessary for sustained oscillations has product levels that remain largely on the rising segment of  $\mathbf{g}_{\alpha\beta}$ , then

$$\frac{d}{dt}\mathbf{P} = \mathbf{AP} + \mathbf{b}, \tag{6}$$

Table 1  
Genetic abbreviations used in the text

Abbreviation	Full name and function
<i>ABI3</i>	<i>ABSCISIC ACID INSENSITIVE 3</i> Transcription factor: regulates genes in ABA signal transduction
<i>API</i>	<i>APETALA 1</i> MADS-box transcription factor: promotes the transition to flowering
<i>CAL</i>	<i>CAULIFLOWER</i> MADS-box transcription factor: promotes flowering
<i>CCA1</i>	<i>CIRCADIAN CLOCK ASSOCIATED 1</i> Myb DNA binding protein: element of circadian clock oscillator
<i>CO</i>	<i>CONSTANS</i> Zinc finger transcription factor: promotes flowering
<i>CRY2</i>	<i>CRYPTOCHROME 2</i> Blue light receptor: positively regulates <i>CO</i> in flowering promotion
<i>FLC</i>	<i>FLOWERING LOCUS C</i> MADS-box transcription factor: inhibits flowering through repressing <i>SOCI</i>
<i>FRI</i>	<i>FRIGIDA</i> FRIGIDA promotes expression of the floral repressor <i>FLC</i>
<i>FRQ</i>	<i>FREQUENCY (Neurospora)</i> Diurnal oscillator component controlling clock period
<i>FT</i>	<i>FLOWERING LOCUS T</i> Kinase inhibitor: promotes flowering
<i>FUL</i>	<i>FRUITFUL</i> MADS-box transcription factor: promotes fruit and pod development
<i>GI</i>	<i>GIGANTEA</i> Nuclear protein: promotes flowering by involving phytochrome signaling
<i>Hd1</i>	<i>Heading date 1 (rice, homologous to CO)</i> Zinc finger transcription factor: promotes flowering in rice
<i>Hd3a</i>	<i>Heading date 3a (rice, homologous to FT)</i> Kinase inhibitor: promotes flowering in rice
<i>LFY</i>	<i>LEAFY</i> Transcription factor: commits the meristem to floral development
<i>LHY</i>	<i>LATE ELONGATED HYPOCOTYL</i> Myb DNA binding protein: element of circadian clock oscillator
<i>PHYB</i>	<i>PHYTOCHROME B</i> Red/far-red photoreceptor: clock entrainment and shade avoidance; inhibits flowering
<i>SOCI</i>	<i>SUPPRESSOR OF OVEREXPRESSION OF CO 1</i> MADS-box transcription factor: promotes flowering by activating <i>LFY</i>
<i>TOCI</i>	<i>TIMING OF CAB EXPRESSION1</i> Autoregulatory psuedo response regulator: element of diurnal clock oscillator
<i>WC1</i>	<i>White-Collar-1 (Component of Neurospora White Collar Ccomplex)</i> Blue light photoreceptor transcription factor
<i>WC2</i>	<i>White-Collar-2 (Component of Neurospora White Collar Ccomplex)</i> Temperature compensation transcription factor

where  $\mathbf{P}$  is a vector of product levels,  $\mathbf{b}$  is a vector containing any constitutive quantities, and  $\mathbf{A}$  is a circuit connectivity and parameter matrix. Eq. (5) is readily solved analytically (O'Neil, 1991, Section 14.5), which, as the next example shows, can be a source of useful biological insight.

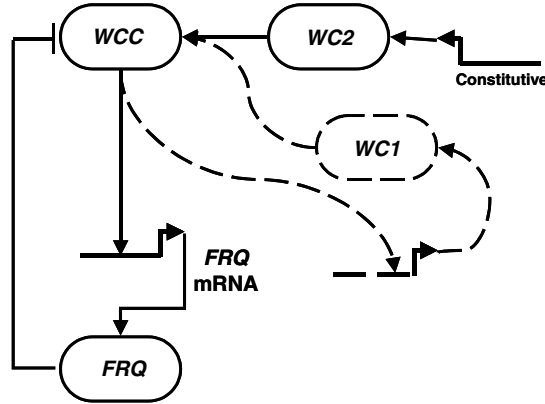


Fig. 4. *Neurospora* diurnal clock (simplified from Loros and Dunlap, 2001). Ovals are proteins or protein complexes. The  $\text{—}\blacktriangleright$  glyph is the accepted symbol for transcription (cf: Fig. 1D). Links between genes denote stimulation ( $\rightarrow$ ) or repression ( $\perp$ ). The dashed elements are excluded from Eq. (7) (see text).

In *A. thaliana* (Alabadi et al., 2001), a three-gene set has a critical role in sustaining oscillations, although the exact interactions are not yet fully established. In contrast, over 40 years of work with the clock in *Neurospora* has revealed many of its molecular details (Loros and Dunlap, 2001). The principle feedback loop (Fig. 4) also has three elements: (i) the White Collar Complex (WCC) heterodimer, (ii) mRNA for the *FREQUENCY* (FRQ) gene, and (iii) the *FRQ* protein product. WCC promotes FRQ transcription and the FRQ protein inhibits WCC. The *Neurospora* clock therefore embodies exactly the type of interactions described earlier as problematic for models based on Eq. (2).

A commentator suggested modeling FRQ mRNA and protein levels and comparing results to the temperature dependency data in Liu et al. (1998). Our model of this system includes the *WHITE-COLLAR-2* (WC2) (Linden and Mancino, 1997) component of WCC (Fig. 4) but, for simplicity, omits the second protein, *WHITE-COLLAR-1* (WC1) (Ballario et al., 1996). WC2 is: (i) more abundant than either FRQ or WC1, (ii) constitutively expressed, and (iii) involved in temperature compensation (unlike WC1). More significantly, (iv) FRQ and WC1 do not react without WC2, and (v) blockage of WCC formation by FRQ appears to be the limiting factor that establishes oscillatory feedback (Loros and Dunlap, 2001 and citations therein).

We used the  $\mathbf{g}_{+0}$  transfer function, making the conservative assumption that there are no significant, as-yet-undiscovered constitutive effects. The resulting equations are

$$\frac{d}{dt} \mathbf{P} = \begin{bmatrix} -\lambda_1 & 0 & R_1\beta_{13}\alpha & R_1\beta_{14}\alpha \\ R_2\beta_{21}\alpha & -\lambda_2 & 0 & 0 \\ 0 & R_3\beta_{32}\alpha & -\lambda_3 & 0 \\ 0 & 0 & R_4 & -\lambda_4 \end{bmatrix} \mathbf{P}, \quad (7)$$

where  $\mathbf{P}$  is a column vector whose elements are, in order from top to bottom, [WCC, FRQ mRNA, FRQ, WC2].  $\mathbf{P}$ , in combination with each row of the matrix, defines one equation of a four-equation system according to the rule for matrix multiplication. For example, the third row specifies the equation for  $P_3$  (i.e., FRQ) that reads

$$\frac{d}{dt} \text{FRQ} = (0)(\text{WCC}) + (R_3\beta_{32}\alpha)(\text{FRQ mRNA}) + (-\lambda_3)(\text{FRQ}) + (0)(\text{WC2}). \quad (8)$$

The last matrix row does not include  $\alpha$  because  $\mathbf{g}_{+0} = 1$  for WC2. The model also omits the WCC response to light (Loros and Dunlap, 2001) because the Liu et al. (1998) data were taken in continuous darkness. Collected at 21 and 28 °C, the authors say that the data are representative of a wider temperature range. Therefore, we set  $\lambda_i = A_T(\kappa_{i1}, \alpha_{i1})$  and  $R_i = A_T(\kappa_{i2}, \alpha_{i2})$  to reduce the number of parameters. Observation constrains the solution: (i) to be positive, (ii) oscillate stably, (iii) have a temperature independent period, (iv) show temperature responses in FRQ amplitude and average level, but (v) not in FRQ mRNA. Space prohibits printing the full solution (available on request), but the mRNA and FRQ protein results are

$$\begin{bmatrix} \text{FRQ mRNA} \\ \text{FRQ} \end{bmatrix} = \begin{bmatrix} c_{21} \sin\left(\frac{2\pi}{p}t\right) + c_{22} \cos\left(\frac{2\pi}{p}t\right) + c_{23} \\ c_{31}R_3 \sin\left(\frac{2\pi}{p}t\right) + c_{32}R_3 \cos\left(\frac{2\pi}{p}t\right) + c_{33}\left(\frac{R_4}{\lambda_4}\right) \end{bmatrix}, \quad (9)$$

where  $p$  is the period. The  $c_{ij}$  are combinations of the constants  $\alpha_{ij}$ ,  $\kappa_{ij}$  (from  $R_i$  and  $\lambda_i$ ) and  $\beta_{ij}$  embedded in Eq. (6). The results show that (i) temperature does not appear in (and therefore does not influence) the FRQ mRNA solution, (ii)  $R_3$  controls the FRQ oscillation amplitude, and (iii) the FRQ diurnal average depends on the WC2 temperature response (via the  $R_4/\lambda_4$  term). The last point is significant. The derivation of Eq. (9) embodied constraints (iii–v) above, but the model was never “told” that WC2 is the “source” of the temperature effects. This fact emerged from the mathematics in agreement with the prior (unexplained) observation that WC2 is involved in temperature compensation.

Eq. (9) appears to have 13 degrees of freedom – 6 for  $c_{ij}$ s, 1 for  $p$ , and 2 each for the Arrhenius parameters in  $R_3$ ,  $R_4$  and  $\lambda_4$ . The actual number is 9 because the constraints eliminate 5 of the 13 but 1 must be added because raw mRNA and protein data are normalized against different standards. A nonlinear least squares fit of the constrained model (Fig. 5) tracks the data well, showing that protein–protein interactions are not automatically beyond the reach of Eq. (2).

### 3.2. Coincidence detection

In plants, the *CONSTANS* (*CO*) gene (Suarez-Lopez et al., 2001) receives both circadian clock and light inputs (Fig. 3). Light may regulate *CO* post-transcriptionally by either enhancing translational efficiency or slowing degradation (Suarez-Lopez et al., 2001). *CO* is strongly implicated as the *external coincidence* detector mediating between exposure to light and progress toward flowering. The external

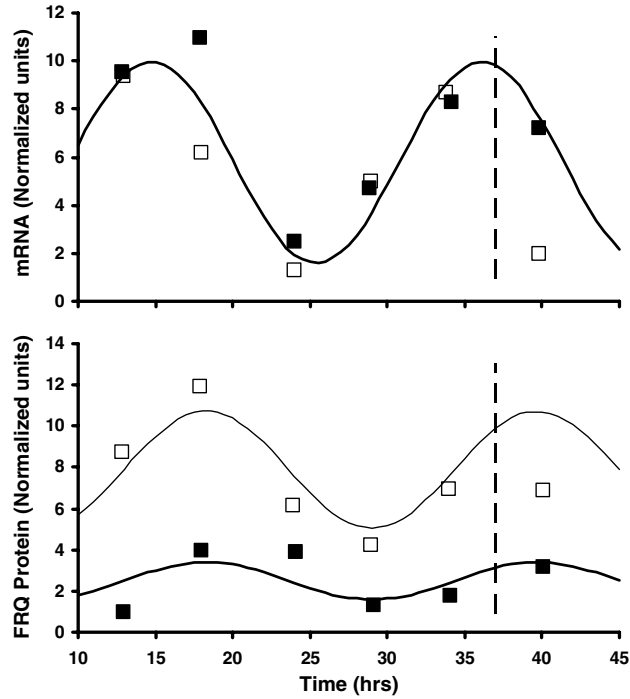


Fig. 5. *Neurospora* clock model. The markers (Liu et al., 1998) show *FRQ* mRNA (top) and *FRQ* protein (bottom) at 21 °C (solid) and 28 °C (open). The lines are a fit of Eq. (9) to all data left of the vertical dashes. Data at 40 hrs (of darkness) were dropped as the 28 °C points showed drift. mRNA was modeled as temperature independent. The thin (thick) protein line is for 28 °C (21 °C).

coincidence model (Bunning, 1936; Pittendrigh, 1972; Samach and Coupland, 2000; Davis, 2002; Roden et al., 2002; Yanovsky and Kay, 2002) states that development rates only respond to light during certain clock phases.

Kojima et al. (2002) give long (LD, 15 h) and short day (SD, 9 h) time series for *HEADING DATE 1* (*Hdl*), the rice (*Oryza sativa*) homolog of *CO*. *Hdl* (i) peaks higher and later under LD, (ii) has both a smaller amplitude and lower average under SD, and (iii) has broader troughs and sharper peaks under SD than LD (Fig. 6(top)). These features suggest (i) a greater daytime net production rate and (ii) a sinusoidal clock signal distorted by a transfer function that is concave up over some relevant input range. A model based on Eq. (2) that incorporates these traits is

$$\frac{d}{dt}(Hd1) = \left. \begin{matrix} R_D \\ R_L \end{matrix} \right\} \mathbf{g}_{NN}(C(t)) - (Hd1) \left\{ \begin{matrix} \lambda_D \\ \lambda_L \end{matrix} \right. \quad (10)$$

where the  $R$ s and  $\lambda$ s are constants and L and D denote light and dark. The clock input is  $C(t) = A \sin(2\pi t/p + \theta) + \mu$ , where  $A$  is amplitude,  $p$  is period,  $\theta$  is phase angle, and  $\mu$  combines the clock average and the  $\mathbf{g}_{NN}$  bias value (see Welch et al., 2003).

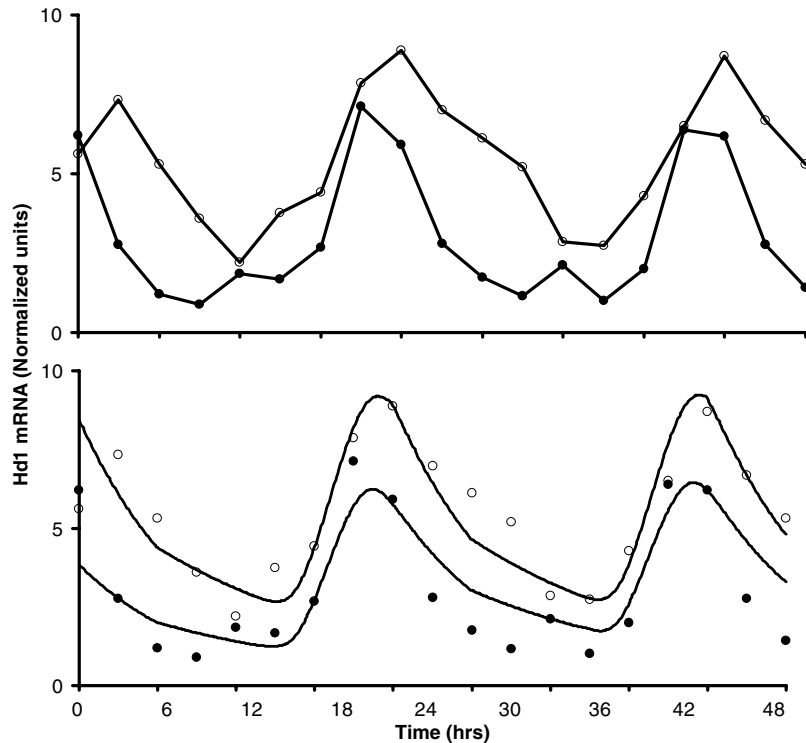


Fig. 6. External coincidence detection. (Top) Expression data from Kojima et al. (2002) on *Hdl* under LD and SD (open and closed markers, respectively). (Bottom) The results of fitting Eq. (9) via nonlinear least sum of absolute differences to balance the effects of large and small values.

Fig. 6 (bottom) shows a fit to the pooled data by the method of Wraith and Or (1998).

The higher LD average of *Hdl* led us to examine the relation of its mean expression to development rate. Multi-cycle averages of *Hdl* values were obtained from Eq. (10) at various photoperiods. The averages were scaled to the [0, 1] interval, subtracted from 1, and plotted (solid line, Fig. 7). The subtraction reversed larger (i.e., LD) and smaller values to visually mimic a short day plant response. Kojima et al. (2002) obtained their data from the Nipponbare variety at 25 °C. Developmental rates were estimated for this system using the model of Yin et al. (1997a,b), and re-scaled to aid comparison (dashed line, Fig. 7). The *Hdl* values are constant at low photoperiods unlike the Yin et al. (1997a,b) model whose calibration reflects the aggregated impact of short days on all processes. However, both models show rate declines above ca. 10 h where, presumably, *Hdl* begins to dominate. This is the first time mRNA levels have been linked to a direct calculation of critical short day length, a key calibration parameter in some crop models (Tsuji et al., 1994; Irmak et al., 2000; Welch et al., 2002).

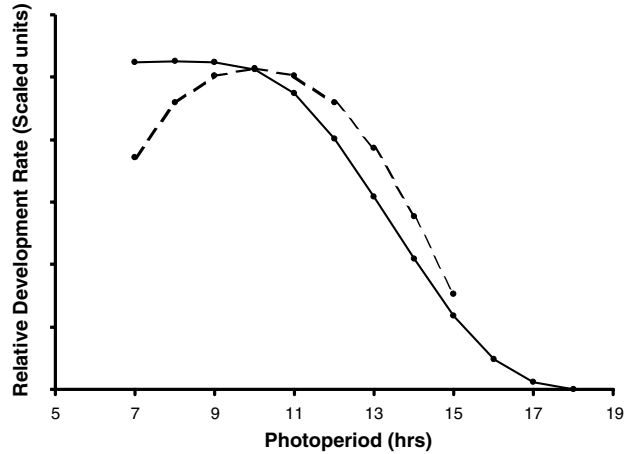


Fig. 7. A critical short day length effect. (Solid line) Time-averaged predicted values of *Hdl* at indicated photoperiods, transformed as described in the text. (Dashed line) Developmental rates predicted by the Yin et al. (1997a,b) model, and scaled. Only lateral position comparisons are valid as vertical scaling is arbitrary.

### 3.3. Bi-stable devices

Bi-stable devices, sometimes called *switches*, are single-bit, zero/one memories that record the (non)occurrence of key events/conditions like the arrival of new growth stages. The word “switch” connotes a rapid change of state (Thornley and Johnson, 2000, Chapter 6) but the transition may be quite slow, so long as it is faster than the processes being regulated. Herein, “bi-stable device” refers to a class of circuits, among which switches are the fastest operators. Bi-stable devices use positive feed-

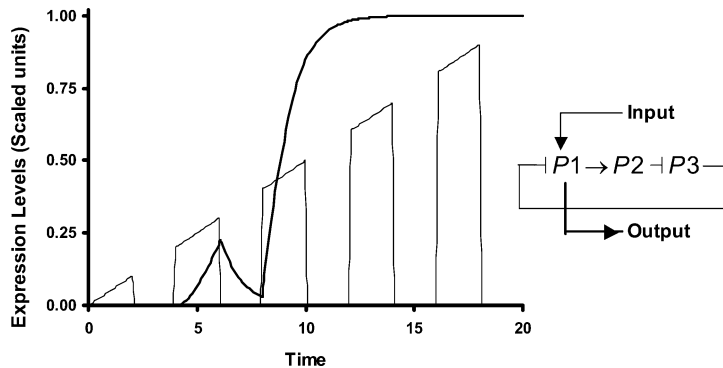


Fig. 8. A bi-stable switch. Small inputs to the switches of Davidson et al. (2002) (modeled with Eq. (2)) produce small output shifts that reverse when the stimulus abates. However, a persistent change of state results when the input exceeds a threshold even slightly.

back to drive outputs toward the extremes of  $g_{\alpha\beta}$ . The device remains in the zero(one)-state given small inputs, but converts to (and stays in) the one(zero)-state when inputs move it even slightly past a triggering threshold (Fig. 8).

Davidson et al. (2002) present a highly detailed circuit controlling endomesoderm differentiation in sea urchin (*Strongylocentrotus purpuratus*) embryos. Several three-gene loops, connected as  $P1 \rightarrow P2 \dashv P3 \dashv P1$ , switch on major network subsystems (“ $\rightarrow$ ” and “ $\dashv$ ” denote promotion and inhibition, respectively). It seems clear that these switches implement positive feedback as “double-negative” loops. The input (Fig. 8) can be a temporally increasing stimulus or, as in the sea urchin, a chemical gradient. In the latter case, separate cells diverge developmentally according to their local concentrations.

Bi-stable devices can also be mutually stimulatory ( $P1 \rightleftharpoons P2$ ) or repressive ( $P1 \dashv \dashv P2$ ) two-gene loops. A likely example is the *LFY*  $\rightleftharpoons$  *API* loop of meristem identity genes (Fig. 3). *LFY* levels increase with time to a threshold that initiates flowering (Blazquez et al., 1997). Parallel signals also up regulate *API* and the *LFY*  $\rightleftharpoons$  *API* loop “amplifies” (Mouradov et al., 2002, Fig. 4) the meristem identity signal, leading to a permanent transition and floral differentiation.

#### 4. Synthesis of whole-plant models

Plants are networks of parts. Over time, parts differentiate into interconnected new parts that grow and later senesce. These activities interweave many genomic and physiological processes that distribute within the plant structure and operate under physical constraints. Selecting ways to represent plant structure is an early, essentially artistic step in modeling. Two existing structural concepts, plant parts as *programming objects* and *Lindenmayer systems*, can aid in merging genomics with SPAC modeling.

##### 4.1. Object-oriented programming (OOP)

OOP is the current dominant programming language paradigm and is described in a crop modeling context by Sequeira et al. (1997). In OOP, objects (called *instances*) can be created or destroyed. When representing plant parts, instances have associated routines (*methods*) useful for localized simulation. The *locality* concept is central to our approach. Tissue parcel responses can only be based on: (i) material, energy, and information collocated either by synthesis or transport or (ii) currently active internal programming. Thus, all biological activity is ultimately local. Additional aspects of OOP are convenient, but not essential. For example, methods for some processes (e.g., general carbon processing, which lacks compartmentalization) may apply to all plant part classes, while others (e.g., stomatal control) may only be needed by a few. In OOP, methods only have to be defined once for each *class* of plant part. Some other OOP features are less help to modeling, as described by Acock and Reddy (1997).

#### 4.2. Lindenmayer systems

Lindenmayer (1968) studied developmental patterns of filamentous algae. He wrote sets of rules (called *grammars*) like

$$\vec{A} \rightarrow \overleftarrow{A} \overrightarrow{B}, \quad \overleftarrow{A} \rightarrow \overleftarrow{B} \overrightarrow{A}, \quad \overrightarrow{B} \rightarrow \overrightarrow{A} \overleftarrow{B} \rightarrow \overleftarrow{A}, \quad (11)$$

where  $A$  and  $B$  denoted cell sizes and the arrows indicated polarity. Filament growth was simulated by starting with a single symbol and repeatedly rewriting it using the rules to guide substitution. For example, starting with  $\vec{A}$ , three successive rewrites give the lengthening filaments  $\overleftarrow{A} \overrightarrow{B}$ ,  $\overleftarrow{B} \overrightarrow{A} \overrightarrow{A}$ , and  $\overleftarrow{A} \overleftarrow{A} \overleftarrow{B} \overrightarrow{A} \overrightarrow{B}$ . Symbol rewriting is enormously powerful and Lindenmayer grammars (now called *L-systems*) can, theoretically, model any dynamic process. L-system symbols need not be cell types but can be assigned any useful meaning. In 3D plant imaging L-systems, for example, symbols are equated to drawing commands (Smith, 1984; Prusinkiewicz and Hanan, 1989). L-system grammars have evolved into full programming languages, albeit ones that are specialized due to their continued reliance on symbol rewriting (Borovikov, 1995). The most elaborate models (<http://www.cpai.uq.edu.au/>) are justifiably called *virtual plants* (Room et al., 1996). For reviews of recent progress and many L-system applications, see Prusinkiewicz (1998, 1999) and Birch et al. (2003).

Some L-systems models exist that are relevant to SPAC processes – e.g., water relations in maize (*Zea mays*) roots (Doussan et al., 1998a,b), shortwave radiation (Mech and Prusinkiewicz, 1996), and meristem temperature (Fournier and Andrieu, 1998). However, two factors may have hindered L-systems' broader application. First, L-system models view interaction with the environment as a process of information exchange (Prusinkiewicz, 1998, p. 133) that may involve symbol rewriting (Fournier and Andrieu, 1998, Fig. 1). In contrast, physicists view system behavior as determined by extended sets of equations that (i) include both environmental and plant variables and (ii) function as constraints that collectively permit only one result. A practical question becomes how these equation sets are to be built and maintained. Second, despite the power of symbol rewriting, its embodiment in specialized programming languages limits its visibility beyond the L-systems community. An analogous situation once existed for OOP. Conceived in 1962 and implemented via specialized languages like SIMULA and SMALLTALK, OOP concepts did not predominate until after they were added to a widely accepted language, C, in 1980.

Our approach (named L/OOP) addresses these issues by reinterpreting L-system symbols to represent plant part objects, *sensu* OOP. In L/OOP, symbol rewriting is replaced by the normal OOP operations of object creation and destruction as guided by L-system rules. Equation system management is centralized as described below.

#### 4.3. Melding the methodologies

Fig. 9 is a schematic, instantaneous snapshot of a L/OOP run. The linked plant part network interacts with functional components that would be elements of any

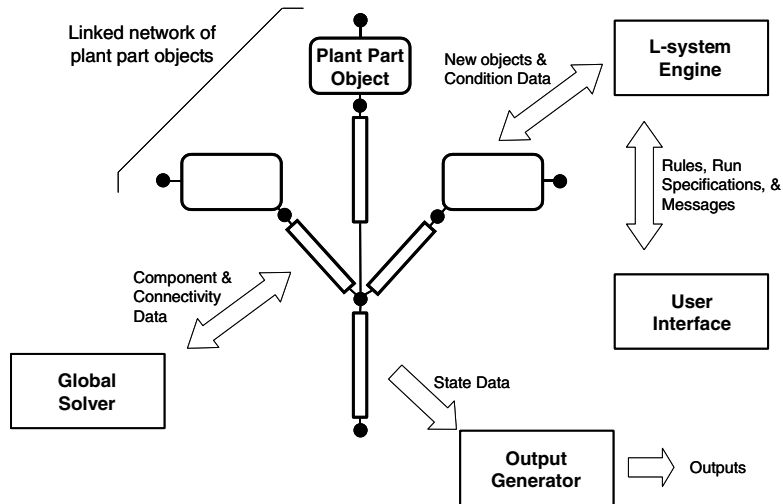


Fig. 9. The L/OOP concept. The rectangular components provide fixed services (see text), exchanging data as shown. The plant is a morphologically realistic network of objects (stems and leaves depicted) and elaborates under L-system control while executing a continuous time simulation.

L/OOP implementation. The Global Solver computes values that depend on whole plant morphology (e.g., anything affecting the energy budget), a task requiring *equation management*. It also performs numerical integration. The Output Generator summarizes results as desired. Our L/OOP example (below) uses the public domain POVray ([www.povray.org](http://www.povray.org)) ray tracing software and the shareware Animagic GIF animator ([www.download.com](http://www.download.com)) to create plant growth motion sequences (see [www.oznet.ksu.edu/agronomy/people/welch/Modeling%20Intra-plant%20Water%20Flows\\_files/frame.htm](http://www.oznet.ksu.edu/agronomy/people/welch/Modeling%20Intra-plant%20Water%20Flows_files/frame.htm)). The User Interface function is clear from the figure.

Plant part objects perform three tasks. First, they store part-specific state data (volume, area, concentration, position, temperature, age, mass, *etc.*). Second, they produce all data used by the Output Generator. Third, and most important, they compute all values determined by local information alone (e.g., hydraulic resistance from capillary size). Some local values are time derivatives like the right hand sides of gene Eq. (2). The corresponding, integrated gene values are controls/signals for physiological processes and/or developmental switch states.

Developmental switches connect to conditional elements in L/OOP rules like

$$\text{Cond} : \langle \text{Apex} \rangle \rightarrow [ \langle \text{Petiole} \rangle [ \langle \text{Lflt} \rangle ] [ \langle \text{Lflt} \rangle ] [ \langle \text{Lflt} \rangle ] \langle \text{Apex} \rangle ]. \quad (12)$$

The  $\langle \rangle$  brackets allow symbols have longer names. Rule (12) says that when an  $\langle \text{Apex} \rangle$  instance detects that condition Cond becomes true, it will differentiate into a trifoliate leaf with a new  $\langle \text{Apex} \rangle$  for later development. To do so, the differentiating  $\langle \text{Apex} \rangle$  notifies the L-system Engine, which reconfigures the plant by creating new plant part objects (Fig. 10) and connecting them as specified by the square brackets

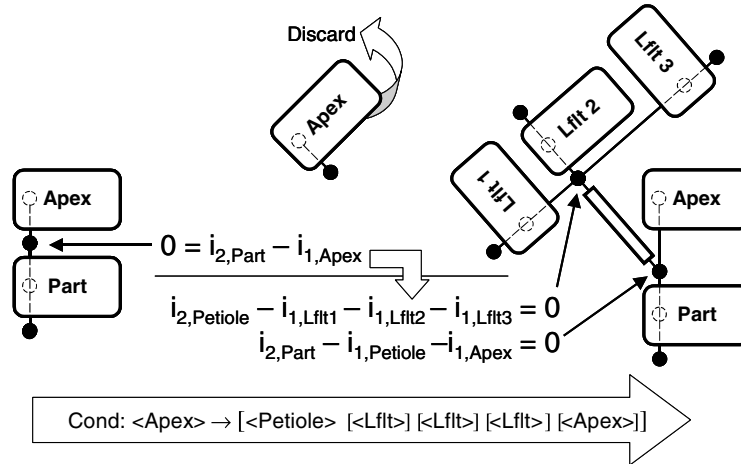


Fig. 10. Applying an L-system rule. When signaled, the L-system Engine creates and links new plant parts. Equation management responds to new network structure (Kirchhoff's Current Law shown) and to internal plant part circuitry (suggested by dashes). Although having little physical impact, apices contain key genomic features.

(Prusinkiewicz and Hanan, 1989). By analogy to DNA, a L/OOP grammar is stored as text that is available within all plant parts.

#### 4.4. Equation management

When development changes plant structure, the system of governing equations also changes, a process that must be managed. Management is automatic for equations like (2) that are local to individual plant parts – new equations are created along with new parts. But variables like water and radiation fluxes depend on whole plant morphology. One of us (He) studied related software engineering issues in an elementary L/OOP testbed similar to the Doussan et al. (1998a,b) root model. Their model was a pure electronic resistance network (Slichter, 1899; Gradmann, 1928; van den Honert, 1948; Kirkham, 2002). While some view electronic networks as simplistic (Hillel, 1971, p. 210; Hillel, 1998, p. 573), the analogy remains in common use (e.g., Martre et al., 2001). Our testbed is a canopy model that extends the Doussan et al. (1998a,b) model by: (i) adding intra-plant water storage and (ii) letting growth respond to water stress. Appendix A gives a unified, efficient mathematical theory for arbitrary networks; to maintain a common vocabulary, electronic terminology is used below.

The testbed builds on the single-part, plant model bounded by the dashed line in Fig. 11. The small circles are nodes. A circuit segment (a–d in Fig. 11) connects two nodes and contains one current source (e.g., for transpiration), resistor, or storage compartment. Each segment has a current ( $I_{seg}$ ) and voltage drop ( $V_{seg}$ ), where seg is a–d. Segments have orientations that may be arbitrary, although the arrows in

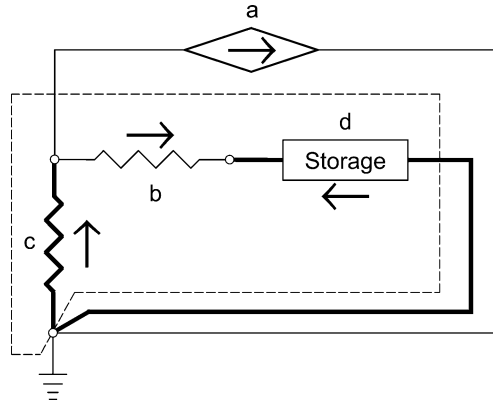


Fig. 11. Plant part example. See the main text and Appendix A for details.

Fig. 11 point from higher to lower voltages. The equation management prototype models storage with a content-dependent voltage source controlled by a Höfler–Thoday curve (Jones, 1992, p. 78). This ties total tissue water potential to relative water content ( $\theta$ ), which indexes stress. Roderick (2001) argued for defining growth volumetrically rather than on a dry matter basis since the former relates more closely to pressure–volume work and, thus, to thermodynamics. Therefore, to link flows to growth, we defined the units of storage content ( $Q$ ), current, resistance ( $R_{seg}$ ), and voltage to be  $m^3$ ,  $m^3 s^{-1}$ ,  $MPa s m^{-3}$ , and  $J m^{-3}$ , respectively. Thus, if  $S$  is the size of a plant part in  $m^3$ ,  $\theta = Q/S$ .

To avoid complexity in a software engineering testbed, growth was modeled by a logistic equation (Prusinkiewicz et al., 1994) that we adjusted as appropriate by a multiplicative water stress factor. Although simple, our strategy is general and can be elaborated as desired. In particular, let  $G$  be the growth rate resulting from arbitrarily involved genomic and physiological equations controlling plant size. Then,

$$\begin{aligned} S' &= G(S, \theta), & I_a(t) &= I_c - I_b, \\ Q' &= I_b, & V_d(\theta) &= I_c R_c(S) + I_b R_b(S, \theta) \end{aligned} \quad (13)$$

where  $t$  is time, arguments indicate specified (but not necessarily all) functional dependencies and/or time series data, and  $'$  denotes time derivatives. The two right-hand equations combine Ohm's and Kirchhoff's Laws and directly reflect network topology. The two left-hand equations must be integrated to advance the model through time.  $G$ ,  $R_b$ ,  $R_c$ , and  $V_d$  are calculated by the plant part object's methods. For example,  $R_c$  can be calculated using Poiseuille's Law for capillaries whose dimensions relate to  $S$ .  $R_b$  might additionally incorporate yield threshold turgor pressure and volumetric extensibility effects indexed by  $\theta$  (Hsiao, 2000). Applying data for  $I_a$  (from a vapor pressure difference submodel in the testbed) yields the explicit model

$$\begin{aligned} S' &= G, \\ Q' &= \frac{(V_d - R_c I_a)}{(R_b + R_c)}, \end{aligned} \quad (14)$$

where all right-hand variables are functions of  $S$ ,  $Q$ , and  $t$ . Convenient initial conditions are  $S(0)$  and  $\theta(0)$ , which determine all other needed values.

Because the network in Fig. 11 is so simple, Eq. (14) was derived manually by eliminating  $I_b$ . However, after the L-System Engine adds new plant parts (each with its own growth rate, flows, etc.) the derivation has to be repeated, mandating automation. Fortunately, easily updated matrices can represent networks of arbitrary structure in flow equations. The simulation steps become: (i) a developmental rule is triggered within a plant part, which requests the L-system Engine to modify the network (Fig. 10); (ii) the Global Solver selects segments comprising a *spanning tree* (bolded in Fig. 11) that connects all nodes without looping; (iii) uses it to compute the required matrices; and (iv) the Global Solver integrates all (local plus matrix flow) differential equations until some plant part detects the next developmental event. Steps (i)–(iv) repeat for the duration of the run. Appendix A gives the mathematical details of (ii)–(iv), including the common case where resistances depend on currents. The L/OOP testbed utilizes fourth-order Runge–Kutta integration (Press et al., 1992) of a less efficient but easier to program version of Eq. (A.6).

## 5. Discussion and conclusions

It is easy to think of physiology as intermediate between genomics and physics, controlled by the former and constrained by the latter. Failure to maintain cognizance across this entire spectrum in the face of expanding knowledge has many adverse consequences: limited understanding, error, missed opportunities, poor prioritization of effort, inability to exploit results, duplication, etc. Quantitative modeling (Kitano, 2002) is a natural response to these hazards that supplies a common vernacular spanning the total range of inquiry. Mathematics is the everyday language of environmental physicists; it is not unfamiliar to many physiologists; and the emerging importance of bioinformatics proves its relevance to genomic scientists.

### 5.1. Object organization

It is not trivial to integrate all the disparate phenomena above into a single model. Stripped to its essentials, our suggestion is that one should model plant models after plants. Genes operate within tissues, which, for convenience, can be equated to plant parts like meristems, leaves, petioles, root segments, etc. The parts are also the sites of all metabolic activity. It is the parts that grow and differentiate into other parts.

The parts are connected into a network that defines the pathways and influences the rates of intra-plant transport. The sizes and locations of the parts delimit the boundary of the microenvironment. These facts all argue that plant parts should be the basic system objects.

This is not a novel idea (Acock and Reddy, 1997; Sequeira et al., 1997; Lemmon and Chuk, 1997; Gauthier et al., 1999), but genomics leads to rethinking the application of object concepts. The cited papers share two features: (i) plant parts were organized hierarchically from large units (mainstems or root systems) to smaller components (fruits and leaves), and (ii) physiological processes like photosynthesis were modeled as separate objects. Gauthier et al. (1999) note that morphology is established botanical convention but process details are still being worked out. They argue that concrete knowledge is best kept separate from mathematical abstractions still in flux.

However, top-down morphological hierarchies do not reflect how plants actually develop as much as they do human perceptions of the end result. In reality, meristems differentiate into parts that grow larger – a mechanism directly congruent with L-system rules. Inclusion of higher-level morphological objects merely complicates class libraries unnecessarily. Sequeira et al. (1997) distinguish “kind of” from “part of” relationships. They use the latter to sequence the flow of control when responding to commands like “Update”, which high-level objects (i) first apply to themselves and (ii) then to the parts that comprise them. However, this order presumes that high-level objects have processes separate from those of their component parts, an idea difficult to reconcile with the locality concept. Furthermore, a leaner class library actually reduces the amount of code needed to accomplish the second task.

Beyond flow control, high-level morphological classes are also used to aggregate output data (Gauthier et al., 1999). However, rather than distort model verisimilitude, we urge realism. In the real world, quantities of interest are obtained by sampling and data analysis. The obvious parallel is to simulate these activities within the Output Generator by having it collate the desired information.

Representing processes as separate objects can also be questioned. Sequeira et al. (1997) correctly state that processes can either be plant part methods or objects in their own right. They favor the latter as a convenience that allows researchers in different laboratories to model separate processes. This exploits a core concept of OOP, object communication across clearly defined interfaces. However, genomic insights are likely to alter “process boundary” definitions dramatically. For example, the autonomous pathway in *A. thaliana* (Fig. 3) contains carbon nutrition genes (Blazquez, 2000). Reared in total darkness and sprayed with sugar solution, *A. thaliana* will flower despite being a long day plant (Roldan et al., 1999). Also, the *CO* protein complexes with the *ABSCISIC ACID INSENSITIVE 3 (ABI3)* product (Rock, 2000), inviting speculation about links between stress affects on flowering time and photoperiod perception. Both examples blur the lines between phenology and other plant processes. Until matters clarify, it seems shortsighted to raise walls between laboratories in the form of borders built in code.

## 5.2. Issues of detail

As the 19th century concluded, Turner (1893) theorized on the philosophical impact of the American frontier at its closing. A century later, the advent of genomics brings biological reductionism close to its own conclusion. However, the vistas of high-speed computation continue to expand, seemingly without limit. These twin frontiers must impact the philosophy of modeling and, in particular, how modelers think about “detail”. The L/OOP testbed used a three-rule grammar and generated 320 plant parts totaling 1280 circuit segments, roughly equivalent to V-23 stage soybean [*Glycine max* (L.) Merr.] in terms of individual leaflets, internodes, and petioles (Welch, unpub. data). Is this model “detailed” because so many plant parts were simulated or “simple” because so few rules were used?

In statistics, many assumptions (normality, homoskedasticity, etc.) became fixed in practice because of underlying theory derived in the pre-computer era. Given only manual calculation, relaxing these assumptions was impractical (Efron and Tibshirani, 1991). However, calculation is now “free”. The result has been an explosion of computationally intensive techniques (resampling methods, etc.) that do not require old simplifications. By analogy, does the amount of morphological detail matter when it is a freely generated byproduct of controls that are the real research focus?

There are several obvious questions. Does this mean one should always model at the lowest level possible? The phrase “lowest level” suggests a perceptual bias. If one thinks of processes, their controls, and applicable physical constraints, it is hard to see which of them is “low”. Modeling is always a tradeoff between the time invested, the accuracy achieved, and the objectives served. However, the balance point is now shifting rapidly toward increased realism in the light of new knowledge and computing capability.

Will realistic models be as difficult to understand as the plants they represent? Suppose some oracle let an investigator read, dissect, run, trace, and otherwise test the source code of a complete and correct, molecular-level plant simulation. Is there any doubt that the researcher’s insight would advance much more rapidly than it could by experimentation on real plants alone? Even though he or she starts with far less knowledge of the hypothetical model than the authors of real models have? Furthermore, the most realistic models (even when qualitative) never exceed current knowledge, but rather define its limits. Thus, model realism can never delay understanding.

Will it be necessary to simulate every gene? Candidly, that is exactly what the phrase “virtual plant” means to some genomic scientists (SIBS, 2000). However, close attention to important regulatory genes is the efficient path to greater realism. For example, the genes in Fig. 3 initiate the entire reproductive cascade. As genetic networks increasingly clarify, the decision processes in physiological models will probably migrate to genomic representations. Even so, much plant physiology will likely remain in code sections quite recognizable from existing crop models.

During the question period of the symposium mentioned in the introduction, there was much discussion on the relative merits of “complex” vs. “simple” mod-

els, generally to the disparagement of the former. In reality, complexity is largely a matter of viewpoint and the techniques available to cope with detail. Plants consist of parts iterated in space and time. Genetic networks contain large sections that are turned *ON* or *OFF* by single, high-level regulatory switches. Genetic interactions and physical laws can both be stated in concise mathematics. Appropriately applied, these features can serve as organizing principles around which to build tractable models. Late in the symposium, an audience member asked, “Is there a so-far-undiscovered, *elegant* theory of the plant?” [The question earned a standing ovation for the very surprised individual who raised it.] If such a theory does exist, it will undoubtedly reveal that simplicity and complexity are merely two parts of a single whole.

### Acknowledgments

The concepts in this paper were originally presented at a symposium in honor of Dr. Gaylon Campbell at the 2001 annual meeting of the Agronomy Society of America. This research was partially supported by grants to Kansas State University from the National Science Foundation (32115), United States Department of Agriculture (2003–35304–13217), and by Hatch Project 507 of the Kansas Agricultural Experiment Station.

**Appendix A.** This appendix summarizes the calculations required for arbitrary, multi-plant-part circuits. The networks here contain only resistors and current or voltage sources. Except for  $S$  and  $Q$ , the notation is that of [Swamy and Thulasiraman \(1981, Section 11.2\)](#) who derive Eq. (A.6) for fixed resistances. We have extended this to resistances that are arbitrary functions of flux (Eq. (A.7)).

A *spanning tree* is any set of segments that provides a unique path between any two nodes. The spanning tree (bolded) in [Fig. 11](#) includes all voltage sources (i.e.,  $V_d$ ), no current sources, and omits  $b$ , the storage input segment. Each segment not in a spanning tree determines a unique loop by linking the two segment ends with a (unique) path through the tree. Loops have orientations, whose assignment may be arbitrary. The loops in [Fig. 11](#) can be described in a matrix,  $\mathbf{B}_{fl}$ , as follows:

$$\begin{array}{c} a \quad b \quad c \quad d \\ a \quad \left[ \begin{array}{cccc} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & -1 \end{array} \right] \\ b \end{array} \quad (\text{A.1})$$

There is a column for each network segment and rows for segments not in the spanning tree. The leftmost columns are in the same order as the rows. A 0 signifies that a particular loop does not contain a given segment. A 1 indicates that the loop orientation matches that of a segment it traverses; otherwise the entry is  $-1$ .

Let the subscripts 1, 2, and 3 refer to collections of current sources, resistors, and voltage sources, respectively. Then a loop matrix can be always be partitioned as

$$\mathbf{B}_{ft} = \begin{bmatrix} \mathbf{U} & \mathbf{B}_{12} & \mathbf{B}_{13} \\ \mathbf{0} & \mathbf{B}_{22} & \mathbf{B}_{23} \end{bmatrix}, \quad (\text{A.2})$$

where  $\mathbf{U}$  is an identity matrix. For Fig. 11 the partition is

$$\mathbf{B}_{ft} = \begin{bmatrix} [1] & [0 & 1] & [0] \\ [0] & [1 & 1] & [-1] \end{bmatrix}. \quad (\text{A.3})$$

Next, voltage drops across resistors must be specified. Swamy and Thulasiraman (1981) use Ohm's Law in the form  $\mathbf{V}_2 = \mathbf{Z}_2 \mathbf{I}_2$ , where  $\mathbf{Z}_2$  is a diagonal matrix of resistance values. In Fig. 11

$$\mathbf{Z}_2 = \begin{bmatrix} R_b & 0 \\ 0 & R_c \end{bmatrix}, \quad (\text{A.4})$$

In plants, Ohm's Law is often violated. Jones (1978) models resistance as hyperbolic with transpiration but his data actually show a variety of curvilinear forms plus hysteresis that becomes extreme under high soil moisture deficits. Let voltage be a general function of current and some parameters  $\mathbf{z}$

$$\mathbf{V}_2 = \mathbb{Z}_2(\mathbf{I}_2, \mathbf{z}). \quad (\text{A.5})$$

Finally, define  $\mathbf{I}_l$  to be a column vector of flows through resistors not in the spanning tree. In Fig. 11,  $\mathbf{I}_l = [I_b]$ . For Ohm's Law resistors, a linear equation (Swamy and Thulasiraman, 1981, Eq. (11.18)) relates  $\mathbf{I}_l$  to the current and voltage sources

$$(\mathbf{B}_{22} \mathbf{Z}_2 \mathbf{B}_{22}^T) \mathbf{I}_l = -\mathbf{B}_{23} \mathbf{V}_3 - \mathbf{B}_{22} \mathbf{Z}_2 \mathbf{B}_{12}^T \mathbf{I}_1. \quad (\text{A.6})$$

The equation for general resistors is

$$\mathbf{B}_{22} \mathbb{Z}_2(\mathbf{B}_{12}^T \mathbf{I}_1 + \mathbf{B}_{22}^T \mathbf{I}_l, \mathbf{z}) = -\mathbf{B}_{23} \mathbf{V}_3. \quad (\text{A.7})$$

Inserting the vectors and matrices for Fig. 11 in (A.6) yields

$$(R_b + R_c) I_b = V_d - R_c I_a \quad (\text{A.8})$$

from which the second right-hand side in (14) is immediately recovered. For complex plant part networks,  $\mathbf{I}_l$  will contain the storage input currents needed for  $\dot{\mathbf{Q}}$  as these were explicitly excluded from the spanning tree. The consolidated growth equation,  $\mathbf{S}' = \mathbf{G}$ , completes the model.

The computational efficiency of this procedure is high. First, fast methods for finding spanning trees (e.g., Suraweera, 1989) allow rapid, automatic construction of  $\mathbf{B}_{ft}$ . Second, all matrices in (A.6) and (A.7) are sparse, dramatically reducing computer memory and processing requirements (Press et al., 1992). Third, it is easy to design networks where  $\mathbf{I}_l$  consists entirely of storage input currents. Then explicit (implicit) integration can be applied to (A.6) ((A.7)) with  $\mathbf{I}_l$  replaced by  $\mathbf{Q}'$ . This yields a system with only one equation per plant part, a considerable savings over naïve methods that generate an equation for each circuit node.

## References

- Acock, B., Reddy, V.R., 1997. Designing an object-oriented structure for crop models. *Ecol. Model.* 94, 33–44.
- Akutsu, T., Miyano, S., Kuhara, S., 1999. Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. *Proceedings of the Pacific Symposium on Biocomputing*, vol. 4. World Publishing Co., Singapore, pp. 17–28.
- Akutsu, T., Miyano, S., Kuhara, S., 2000. Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics* 16, 727–734.
- Alabadi, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Mas, P., Kay, S.A., 2001. Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. *Science* 293, 880–883.
- Arora, V.K., 2003. Simulating energy and carbon fluxes over winter wheat using coupled land surface and terrestrial ecosystem models. *Agric. Forest Meteorol.* 118, 21–47.
- Baker, J.M., 1996. Use and abuse of crop simulation models. *Agron. J.* 88, 689.
- Baldi, P., Hatfield, G.W., 2002. *DNA Microarrays and Gene Expression*. Cambridge University Press, Cambridge, UK.
- Ballario, P., Vittorioso, P., Mangrelli, A., Tallora, C., Cabibbo, A., Mancino, G., 1996. White-collar-1, a central regulator of blue light response in *Neurospora crassa*, is a zinc-finger protein. *EMBO J.* 15, 1650–1657.
- Barash, Y., Friedman, N., 2001. Context specific Bayesian clustering for gene expression data. In: *Proceedings of the 5th Annual International Conference on Computational Molecular Biology RECOMB-2001. ACM-SIGACT*, New York, pp. 2–11.
- Birch, C.J., Andrieu, B., Fournier, C., Vos, J., Room, P., 2003. Modelling kinetics of plant canopy architecture concepts and applications. *Euro. J. Agron.* 19, 519–533.
- Blazquez, M.A., 2000. Flower development pathways. *J. Cell Sci.* 113, 3547–3548.
- Blazquez, M.A., Ahn, J.H., Weigel, D., 2003. A thermosensory pathway controlling flowering time in *Arabidopsis thaliana*. *Nature Genet.* 33, 168–171.
- Blazquez, M.A., Soowal, L.N., Lee, I., Weigel, D., 1997. LEAFY expression and flowering initiation in *Arabidopsis*. *Development* 124, 3835–3844.
- Boote, K.J., Jones, J.W., Pickering, N.B., 1996. Potential uses and limitations of crop models. *Agron. J.* 88, 704–716.
- Borovikov, I.A., 1995. L-systems with inheritance: An object-oriented extension of L-systems. *ACM SIGPLAN Notices* 30, 43–60.
- Buchanan, B., Gruissem, W., Jones, R.L., 2000. *Biochemistry and Molecular Biology of Plants*. American Society of Plant Physiologists, Maryland, USA, p. 1397.
- Budyko, M.I., 1974. *Climate and Life*. Academic Press, p. 508.
- Bunning, E., 1936. Die endonome Tagesrhythmische als Grundlage der photoperiodischen Reaktion. *Ber. Deut. Bot. Ges.* 54, 590–607 (In German).
- Butler, D., 1999. Computing 2010: from black holes to biology. *Nature* 402, C67–C70.
- Casal, J.J., Sanchez, R.A., 1998. Phytochromes and seed germination. *Seed Sci. Res.* 8, 317–329.
- Chapman, S., Cooper, M., Podlich, D., Hammer, G., 2003. Evaluating plant breeding strategies by simulating gene action and dryland environment effects. *Agron. J.* 95, 99–113.
- Chen, T., He, H.L., Church, G.M., 1999. Modeling gene expressions with differential equations. *Proceedings of the Pacific Symposium on Biocomputing*, vol. 4. World Publishing Co., Singapore, pp. 17–28.
- Cooper, M., Chapman, S.C., Podlich, D.W., Hammer, G.L., 2002. The GP problem: quantifying gene-to-phenotype relationships. *In Silico Biol.* 2, 151–164.
- Csete, M.E., Doyle, J.C., 2002. Reverse engineering or biological complexity. *Science* 295, 1664–1669.
- D'Haeseleer, P., Wen, X., Fuhrman, S., Somogyi, R., 1999. Linear modeling of mRNA expression levels during CNS development and injury. *Proceedings of the Pacific Symposium on Biocomputing*, vol. 4. World Publishing Co., Singapore, pp. 41–52.
- Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Caestani, C., Yuh, C., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G.,

- Pan, H., Schilstra, M.J., Clarke, P.C., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., . A Genomic regulatory network for development. *Science* 295, 1669–1678.
- Davis, S.J., 2002. Photoperiodism: The coincidental perception of the season. *Curr. Biol.* 12, R841–R843.
- Dickenson, R.E., 1984. Modeling evapotranspiration for three-dimensional global climate models. In: Hanson, E., Takahashi, T. (Eds.), *Climate Processes and Climate Sensitivity*. Am. Geophys. Union, pp. 58–72.
- Doussan, C., Pages, L., Vercambre, G., 1998a. Modelling of the hydraulic architecture of root systems: An integrated approach to water absorption Model description. *Ann. Bot.* 81, 213–223.
- Doussan, C., Vercambre, G., Pages, L., 1998b. Modelling of the hydraulic architecture of root systems: An integrated approach to water absorption Distribution of axial and radial conductances in maize. *Ann. Bot.* 81, 225–232.
- Efron, B., Tibshirani, R., 1991. Statistical data analysis in the computer age. *Science* 253, 390–395.
- Ezzell, C., 2002. Proteins rule. *Scient. Am.* 286, 40–47.
- Fiehn, O., 2002. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* 48, 155–171.
- Fournier, C., Andrieu, B., 1998. A 3D architectural and process-based model of maize development. *Ann. Bot.* 81, 233–250.
- Frank, S.A., 1998. Population and quantitative genetics of regulatory networks. *J. Theo. Biol.* 197, 281–294.
- Friedman, N., Linial, M., Nachman, I., Pe'er, D., 2000. Using Bayesian networks to analyze expression data. *J. Comput. Biol.* 7, 601–620.
- Fry, C.J., Peterson, C.L., 2002. Unlocking the gates of gene expression. *Science* 295, 1847–1848.
- Gauthier, L., Gary, C., Zekki, H., 1999. GPSF: a generic and object-oriented framework for crop simulation. *Ecol. Model.* 116, 253–268.
- Goss, P.J.E., Peccoud, J., 1999. Analysis of the stabilizing effect of ROM on the genetic network controlling Cole1 plasmid replication Proceedings of the Pacific Symposium on Biocomputing, vol. 4. World Publishing Co., Singapore, pp. 65–76.
- Gradmann, H., 1928. Untersuchungen über die Wasserverhältnisse des Bodens als Grundlage des Pflanzenwachstums. *Jahrbucher für Wissenschaftliche Botanik* 69, 1–100.
- Halliday, K.J., Salter, M.G., Thingnaes, E., Whitelam, G.C., 2003. Phytochrome control of flowering is temperature sensitive and correlates with expression of the floral integrator FT. *Plant J.* 33, 875–885.
- Hammer, G.L., 1998. Crop modeling: current status and opportunities to advance. *Acta Hort.* 456, 27–36.
- Hartemink, A.J., Gifford, D.K., Jaakkola, T.S., Young, R.A., 2001. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks Proceedings of the Pacific Symposium on Biocomputing, vol. 6. World Publishing Co., Singapore, pp. 422–433.
- Hillel, D., 1971. *Soil and Water. Physical Principles and Processes*. Academic Press, New York, p. 288.
- Hillel, D., 1998. *Environmental Soil Physics*. Academic Press, New York, p. 771.
- Horn, P.J., Peterson, C.L., 2002. Chromatin higher order folding: wrapping up transcription. *Science* 297, 1824–1827.
- Hsiao, T.C., 2000. Leaf and root growth in relation to water status. *HortScience* 35, 1051–1058.
- Ideker, T.E., Thorsson, V., Karp, R.M., 2000. Discovery of regulatory interactions through perturbation: Inference and experimental design Proceedings of the Pacific Symposium on Biocomputing, vol. 5. World Publishing Co., Singapore, pp. 302–313.
- Irmak, A., Jones, J.W., Mavromatis, T., Welch, S.M., Boote, K.J., Wilkerson, G.G., 2000. Evaluating methods for simulating soybean cultivar responses using cross-validation. *Agron. J.* 92, 1140–1149.
- Johnson, I.R., Thornley, J.H.M., 1985. Temperature dependence of plant and crop processes. *Ann. Bot.* 55, 1–24.
- Jones, H.G., 1978. Modelling diurnal trends of leaf water potential in transpiring wheat. *J. Appl. Ecol.* 15, 613–626.
- Jones, H.G., 1992. *Plants and Microclimate*, second ed. Cambridge University Press, Cambridge, p. 428.
- Kirkham, M.B., 2002. The concept of the soil-plant-atmosphere continuum and applications. In: Smiles, P.A.C., Raats, P.A.C., Warrick, A.W. (Eds.), *Heat and Mass Transfer in the Natural Environment*, Geophysical Monograph Series. Am. Geophysical Union, Washington, DC, pp. 327–335.

- Kitano, H., 2002. Systems biology: A brief overview. *Science* 295, 1662–1664.
- Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araaki, T., Yano, M., 2002. Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant Cell Physiol.* 43, 1096–1105.
- Koornneef, M., Alonso-Blanco, C., Peeters, A.J.M., Soppe, W., 1998. Genetic control of flowering time in Arabidopsis. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 49, 345–370.
- Kouzarides, T., 2002. Histone methylation in transcriptional control. *Curr. Opin. Genet. Dev.* 21, 198–209.
- Lemmon, H., Chuk, N., 1997. Object-oriented design of a cotton crop model. *Ecol. Model.* 94, 45–51.
- Lewin, B., 2000. *Genes VII*. Oxford University Press and Cell Press, New York.
- Liang, S., Fuhrman, S., Somogyi, R., 1998. REVEAL: A general reverse engineering algorithm for inference of genetic network architecture. *Proceedings of the Pacific Symposium on Biocomputing*, vol. 3. World Publishing Co., Singapore, pp. 18–29.
- Linden, H., Mancino, G., 1997. White-collar-2, a partner in blue-light signal transduction, controlling expression of light regulated genes in *Neurospora crassa*. *EMBO J.* 16, 98–109.
- Lindenmayer, A., 1968. Mathematical models for cellular interaction in development, Parts I and II. *J. Theo. Biol.* 18, 280–315.
- Liu, Y., Mellow, M., Loros, J.J., Dunlap, J.C., 1998. How temperature changes reset a circadian oscillator. *Science* 281, 825–829.
- Loros, J.J., Dunlap, J.C., 2001. Genetic and molecular analysis of circadian rhythms in *Neurospora*. *Ann. Rev. Physiol.* 63, 757–794.
- Maki, Y., Tominaga, D., Okamoto, M., Watanabe, S., Eguchi, Y., 2001. Development of a system for the inference of large scale genetic networks. *Proceedings of the Pacific Symposium on Biocomputing*, vol. 6. World Publishing Co., Singapore, pp. 446–458.
- Mandelbrot, B.B., 1983. *The Fractal Geometry of Nature*. W.H. Freeman, San Francisco, p. 486.
- Marnellos, G., Deblandre, G.A., Mjolsness, E., Kintner, C., 2000. Delta-notch lateral inhibitory patterning in the emergence of ciliated cells in *Xenopus*: Experimental observations and a gene network model. *Proceedings of the Pacific Symposium on Biocomputing*, vol. 5. World Publishing Co., Singapore, pp. 326–337.
- Martinez-Zapater, J.M., Coupland, G., Dean, C., Koornneef, M., 1994. The transition to flowering in Arabidopsis. In: Meyerowitz, E.M., Somerville, C.R. (Eds.), *Arabidopsis*. Cold Spring Harbor Lab. Press, Cold Spring Harbor, NY, pp. 403–433.
- Martre, P., Cochard, H., Durand, J.L., 2001. Hydraulic architecture and water flow in growing grass tillers (*Festuca arundinacea* Schreb.). *Plant Cell Environ.* 24, 65–76.
- Matsuno, H., Doi, A., Nagasaki, M., Miyano, S., 2000. Hybrid Petri net representation of gene regulatory network. *Proceedings of the Pacific Symposium on Biocomputing*, vol. 5. World Publishing Co., Singapore, pp. 338–349.
- McCown, R.L., Hammer, G., Hargreaves, J., Holzworth, D., Freebairn, D., 1996. APSIM: A novel software system for model development, model testing, and simulation in agricultural systems research. *Agron. Syst.* 50, 255–271.
- Mech, R., Prusinkiewicz, P., 1996. Visual models of plants interacting with their environment. In: *Proceedings of SIGGRAPH 96*, New Orleans, Louisiana, August 4–9. Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, 1996, pp. 397–410.
- Mendoza, L., Alvarez-Buylla, E.R., 1998. Dynamics of the genetic regulatory network for *Arabidopsis thaliana* flower morphogenesis. *J. Theo. Biol.* 193, 307–319.
- Mendoza, L., Alvarez-Buylla, E.R., 2000. Genetic regulation of root hair development in *Arabidopsis thaliana*: A network model. *J. Theo. Biol.* 204, 311–326.
- Messina, C.D., 2003. Gene-based systems approach to simulate soybean growth and development and application to ideotype design in target environments. Ph.D. dissertation, University of Florida.
- Monteith, J.L., 1996. The quest for balance in crop modeling. *Agron. J.* 88, 695–697.
- Mouradov, A., Cremer, F., Coupland, G., 2002. Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell (Suppl.)*, S111–S130.
- O’Neil, P.V., 1991. *Advanced engineering mathematics*. Wadsworth Publishing Co., New York, p. 1594.

- Passioura, J.B., 1996. Simulation models: science, snake oil, education, or engineering. *Agron. J.* 88, 690–694.
- Pearlmutter, B.A., 1995. Gradient calculations for dynamic recurrent neural networks: A survey. *IEEE Trans. Neural Networks* 6, 1212–1228.
- Peterson, C.L., 2002. Chromatin remodeling enzymes: taming the machines. *EMBO Rep.* 31, 319–322.
- Phelps, T.J., Palumbo, A.V., Beliaev, A.S., 2002. Metabolomics and microarrays for improved understanding of phenotypic characteristics controlled by both genomics and environmental constraints. *Curr. Opin. Biotechnol.* 13, 20–24.
- Philip, J.R., 1966. Plant water relations: Some physical aspects. *Ann. Rev. Plant Physiol.* 17, 245–268.
- Pittendrigh, C.S., 1972. Circadian surfaces and the diversity of possible roles of circadian organization in photoperiodic induction. *Proc. Natl. Acad. Sci. USA* 69, 2734–2737.
- Podlich, D., Cooper, M., 1998. QU-GENE: a simulation platform for quantitative analysis of genetic models. *Bioinformatics* 14, 632–653.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. *Numerical Recipes in C*. Cambridge University Press, Cambridge UK.
- Prusinkiewicz, P., 1998. Modeling of spatial structure and development of plants: a review. *Sci. Hortic.* 74, 113–149.
- Prusinkiewicz, P., 1999. A look at the visual modeling of plants using L-systems. *Agronomie* 19, 211–224.
- Prusinkiewicz, P., Hanan, J., 1989. *Lindenmayer Systems, Fractals, and Plants*, Lecture Notes in Biomathematics, vol. 79. Springer, New York, p. 120.
- Prusinkiewicz, P., Remphrey, W., Davidson, C., Hammel, M., 1994. Modeling the architecture of expanding *Fraximus pennsylvanica* shoots using L-systems. *Can. J. Bot.* 72, 701–714.
- Reinitz, J., Sharp, D.H., 1995. Mechanism of formation of eve stripes. *Mech. Dev.* 49, 133–158.
- Reyes, J.C., Hennig, L., Gruissem, W., 2002. Chromatin-remodeling and memory factors: new regulators of plant development. *Plant Physiol.* 130, 10901101.
- Rock, C.D., 2000. Pathways to abscisic acid-regulated gene expression. *New Phytol.* 148, 357–396.
- Roden, L.C., Song, H., Jackson, S.D., Morris, K., Carre, I.A., 2002. Floral responses to photoperiod are correlated with the timing of rhythmic expression relative to dawn and dusk in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 99, 13313–13318.
- Roderick, M.L., 2001. On the use of thermodynamic methods to describe water relations in plants and soil. *Aust. J. Plant Physiol.* 28, 729–742.
- Roldan, M., Gomez-Mena, C., Ruiz-Garcia, L., Salinas, J., Martinez-Zapater, J.M., 1999. Sucrose availability on the aerial part of the plant promotes morphogenesis and flowering of *Arabidopsis* in the dark. *Plant. J.* 20, 581–590.
- Room, P., Hanan, J., Prusinkiewicz, P., 1996. Virtual plants: new perspectives for ecologists, pathologists, and agricultural scientists. *Trends Plant Sci.* 1, 33–38.
- Samach, A., Coupland, G., 2000. Time measurement and the control of flowering in plants. *BioEssays* 22, 38–47.
- Samsonova, M.G., Serov, V.N., 1999. NetWork: An interactive interface to the tools for analysis of genetic network structure and dynamics. *Proceedings of the Pacific Symposium on Biocomputing*, vol. 4. World Publishing Co., Singapore, pp. 102–111.
- Sellers, P.J., Minz, Y., Sud, Y.C., Dalcher, A., 1986. A simple biosphere model (SiB) for use within general circulation models. *J. Atmos. Sci.* 43, 505–531.
- Sellers, P.J., Randall, D.A., Collatz, G.J., Berry, J.A., Field, C.B., Dazlich, D.A., Zhang, C., Collelo, G.D., Bounoua, L.A., 1996a. A revised land surface parameterization (SiB2) for atmospheric GCMs. I. Model formulation. *J. Climate* 9, 676–705.
- Sellers, P.J., Los, S.O., Tucker, C.J., Justice, C.O., Dazlich, D.A., Collatz, G.J., Randall, D.A., 1996b. A revised land surface parameterization (SiB2) for atmospheric GCMs. II. The generation of global fields of terrestrial biophysical parameters from satellite data. *J. Climate* 9 (4), 706–737.
- Sequeira, R.A., Olson, R.L., MCKinion, J.M., 1997. Implementing Generic, Object-Oriented Models in *Biology* vol. 94, 17–31 (Plus other papers in this Special Issue).

- SIBS, 2000. Functional genomics and the virtual plant: A blueprint for understanding how plants are built and how to improve them. Salk Institute for Biol. Studies, La Jolla, CA, January 13–14, 2000. Available from: <<http://www.arabidopsis.org/workshop1.html>>.
- Simpson, G.G., Gendall, A.R., Dean, C., 1999. When to switch to flowering. *Ann. Rev. Cell Dev. Biol.* 99, 519–550.
- Sinclair, T.R., Seligman, N.G., 1996. Crop modeling: from infancy to maturity. *Agron. J.* 88, 698–704.
- Slichter, C.S., 1899. Theoretical investigation of the motion of ground-water. US Dep. Interior Geol. Survey Ann. Rep. 19, 295–384.
- Smith, A.R., 1984. Plants, fractals, and formal languages. *Comp. Graph.* 18, 1–10.
- Suarez-Lopez, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., Coupland, G., 2001. CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* 410, 1116–1120.
- Suraweera, F., 1989. A fast algorithm for the minimum spanning tree. *Comput. Ind.* 13, 181–185.
- Swamy, M.N.S., Thulasiraman, K., 1981. *Graphs, Networks, and Algorithms*. Wiley Interscience, New York, p. 592.
- Szallasi, Z., Liang, S., 1998. Modeling the normal and neoplastic cell cycle with realistic Boolean genetic networks: Their application for understanding carcinogenesis and assessing therapeutic strategies. *Proceedings of the Pacific Symposium on Biocomputing*, vol. 3. World Publishing Co., Singapore, pp. 54–65.
- Thornley, J.H.M., Johnson, I.R., 2000. *Plant and Crop Modelling*. The Blackburn Press, Caldwell, NJ, p. 669.
- Tominaga, D., Okamoto, M., Maki, Y., Watanabe, S., Eguchi, Y., 1999. Nonlinear numerical optimization technique based on genetic algorithm for inverse problem: Towards the inference of genetic networks. In: *Computer Science and Biology, Proceedings of the German Conference on Bioinformatics*, vol. 4. Hanover, Germany, pp. 127–140.
- Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T.S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, K., Yugi, K., Venter, J.C., Hutchison III, C.A., 1999. E-CELL: software environment for whole-cell simulation. *Bioinformatics* 15, 72–84.
- Tsuji, G.Y., Uehara, G., Balas, S., eds. 1994. *Decision Support System for Agro-technology Transfer (DSSAT)*, Version 3. University of Hawaii, Honolulu, HI.
- Turner, F.J., 1893. The significance of the frontier in American history. Chicago Worlds Fair, July 12. Available from: <<http://xroads.virginia.edu/~HYPER/TURNER/chapter1.html>>.
- van den Honert, T.H., 1948. Water transport in plants as a catenary process. *Disc. Faraday Soc.* 3, 146–153.
- Waage, P. Guldberg C.M., 1864. *Studies concerning affinity*. Forhandler: Videnskabs-Selskabet i Christiania, 35 (In Norwegian). Engl. trans. by H.I Abrash at <http://dbhs.wvusd.k12.ca.us/webdocs/Chem-History/Concerning-Affinity.html>.
- Ward, J.K., Antonovics, J., Thomas, R.B., Strain, B.R., 2000. Is atmospheric CO<sub>2</sub> a selective agent on model C<sub>3</sub> annuals. *Oecologia* 123, 330–341.
- Weaver, D.C., Workman, C.T., Stormo, G.D., 1999. Modeling regulatory network with weight matrices. *Proceedings of the Pacific Symposium on Biocomputing*, vol. 4. World Publishing Co., Singapore, pp. 112–123.
- Weiss, A., 2003. Introduction. *Agron. J.* 95, 1-3 *et seq.* Collected papers from the “Crop Modeling and Genomics” Symposium Nov 7, 2000. ASA Annual Meeting, Minneapolis, MN.
- Welch, S.M., Roe, J.L., Dong, Z., 2003. A genetic neural network model of flowering time control in *Arabidopsis thaliana*. *Agron. J.* 95, 71–81.
- Welch, S.M., Wilkerson, G., Whiting, K., Sun, N., Vagts, T., Buol, G., Mavromatis, T., 2002. Estimating soybean model genetic coefficients from private-sector variety performance trial data. *Trans. ASAE* 45, 1163–1175.
- White, J., Hoogenboom, G., 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. *Agron. J.* 88, 416–422.
- Wolf, D.M., Eeckman, F.H., 1998. On the relationship between genomic regulatory element organization and gene regulatory dynamics. *J. Theo. Biol.* 195, 167–186.

- Wraith, J., Or, D., 1998. Nonlinear parameter estimation using spreadsheet software. *J. Nat. Res. Life Sci. Ed.* 27, 13–19.
- Yanovsky, M.J., Kay, S.A., 2002. Molecular basis of seasonal time measurement in *Arabidopsis*. *Nature* 419, 308–312.
- Yin, X., Kropff, M.J., McLaren, G., Visperas, R.M., 1995. A nonlinear model for crop development as a function of temperature. *Agric. Forest Meteorol.* 77, 1–16.
- Yin, X., Kropff, M.J., Nakagawa, H., Horie, T., Goudriaan, J., 1997a. A model for photothermal responses of flowering in rice I. Modevaluation. *Field Crops Res.* 51, 201–211.
- Yin, X., Kropff, M.J., Horie, T., Nakagawa, H., Centeno, H.G.S., Zhu, D., Goudriaan, J., 1997b. A model for photothermal responses of flowering in rice I. Model description and parameterization. *Field Crops Res.* 51, 189–200.