REVIEW PAPER

Under what circumstances can process-based simulation models link genotype to phenotype for complex traits? Case-study of fruit and grain quality traits

Nadia Bertin1,*, Pierre Martre2,3, Michel Génard1, Bénédicte Quilot4 and Christophe Salon5

1 INRA, UR1115 Plantes et Systèmes de Culture Horticoles, F-84 914 Avignon, France
2 INRA, UMR1095 Génétique, Diversité et Ecophysiologie des Céréales, F-63 100 Clermont-Ferrand, France
3 Université Blaise Pascal, UMR1095 Génétique, Diversité et Ecophysiologie des Céréales, F-63 100 Clermont-Ferrand, France
4 INRA, UR1052 Génétique et Amélioration des Fruits et Légumes, F-84 143 Avignon, France
5 INRA, UMR102 Génétique et Ecophysiologie des Légumineuses, F-21 065 Dijon, France

* To whom correspondence should be addressed: E-mail: nadia.bertin@avignon.inra.fr

Received 4 November 2009; Accepted 3 December 2009

Abstract

Detailed information has arisen from research at gene and cell levels, but it is still incomplete in the context of a quantitative understanding of whole plant physiology. Because of their integrative nature, process-based simulation models can help to bridge the gap between genotype and phenotype and assist in deconvoluting genotype-by-environment (G×E) interactions for complex traits. Indeed, G×E interactions are emergent properties of simulation models, i.e. unexpected properties generated by complex interconnections between subsystem components and biological processes. They co-occur in the system with synergistic or antagonistic effects. In this work, different kinds of G×E interactions are illustrated. Approaches to link model parameters to genes or quantitative trait loci (QTL) are briefly reviewed. Then the analysis of G×E interactions through simulation models is illustrated with an integrated model simulation of peach (Prunus persica (L.) Batsch) fruit mass and sweetness, and with a model of wheat (Triticum aestivum L.) grain yield and protein concentration. This paper suggests that the management of complex traits such as fruit and grain quality may become possible, thanks to the increasing knowledge concerning the genetic and environmental regulation of organ size and composition and to the development of models simulating the complex aspects of metabolism and biophysical behaviours at the plant and organ levels.

Key words: Fruit, gene-based model, genotype-by-environment interaction, genotypic parameter, grain, QTL-based model, quality, simulation model.

Introduction

Like many quantitative crop traits, the quality of harvested organs is a complex issue, which results from many overlapping physiological processes, genetically and environmentally controlled during grain, seed or fruit development. Over the last 50 years, yield for various crops has continuously increased due to genetic and crop management progresses (Calderini and Slafer, 1998; Cassman, 1999, 2001), but, at the same time, quality attributes have levelled off or even decreased for numerous products such as cereals (Oury et al., 2003), grain legumes (Weber and Salon, 2002; Graham and Vance, 2003), oilseeds (Triboi and Triboi-Blondel, 2002), fruits for processing (Grandillo et al., 1999) or fresh fruits (Causse et al., 2003). Thus, a critical question for the future is how to manage crop quality while maintaining yield, by finding the best combinations of genetic resources and cultural practices adapted to, and respectful of specific environments.

Despite the identification of numerous quantitative trait loci (QTL) for quality traits, and the identification of genes involved in their control for different species, the genetic
improvement of crop quality is still a complicated and rather slow process. Up to now, few physiological functions have been clearly ascribed to known gene sequences and, to date, the huge progress in gene discovery has only weakly aided genetic selection (Miflin, 2000; Sinclair et al., 2004). This results at first, from the complexity of most of the traits of interest and of their sensitivity to the environment. Secondly, the processes involved are controlled by multiple interacting genes, which themselves interact with the environment and crop management (Causse et al., 2007). For instance, in tomato (Solanum lycopersicum L.) fruit, more than 100 genes located in 16 regions of the genome, are associated with fruit composition, mainly sugar and acid contents (Causse et al., 2004; Bermúdez et al., 2008). Consequently, QTLs for a given trait usually explain only low proportions of the observed trait variations. Moreover, most of these QTLs depend on the environment and on the genetic background (Börner et al., 1993; Blanco et al., 2002; Chaib et al., 2006; Dudley et al., 2007). This usually results in strong genotype-by-environment (G×E) or genotype-by-management interactions, which renders the research in genetics and their applications for selection complex. As a consequence, in order to analyse the genetic and environmental determinants of crop attributes, agronomists and geneticists often have to perform extensive experiments over several years at different sites or under different environmental conditions. Although this approach is still useful to evaluate QTL stability (Prudent et al., 2009), it is laborious, expensive, and time-consuming and, thus, it is often limited to the comparison of a low number of genotypes and traits conducted under restricted environmental conditions.

To overcome these difficulties, several authors have proposed the use of ecophysiological process-based simulation models for analysing QTL or genotype effects on different processes and for different species. For example, this has been done for yield in barley (Hordeum vulgare L.; Yin et al., 2000), phenological development in soybean [Glycine max (L.) Merr.; Stewart et al., 2003], leaf elongation rate in maize (Zea mays L.; Reymond et al., 2003), or fruit quality in peach [Pruunus persica (L.) Batsch; Quilot et al., 2005b]. Different approaches have been proposed to introduce genetic information into simulation models; they have been applied successfully to predict the parameter values of specific gene or allele combinations (Shorter et al., 1991; White and Hoogenboom, 1996; Yin et al., 1999), to design new ideotypes adapted to target environments (Kropff et al., 1995; Tardieu, 2003; Letort et al., 2008), and to analyse G×E interactions (Yin et al., 2005).

All of these studies outlined the possibility of process-based simulation models for predicting G×E interactions under a wide range of conditions. They also pointed out the need for upgraded models, in particular to predict complex traits, and highlighted the necessity to link model parameters with easily measurable physiological traits and known QTLs or genes (Yin et al., 2004; Struik et al., 2007). Objectives of this paper were to illustrate how process-based simulation models can help in bridging the gap between genotype and phenotype, due to their intrinsic capacity to mimic complex systems and to integrate multi-scale levels of controls. How G×E interactions emerge from process-based simulation models is first outlined, and the model structure necessary for such analysis is briefly discussed. Ways to introduce genetic information in these models are subsequently presented and are illustrated using fruit size and sweetness, and cereal yield and grain protein concentration, as two examples.

Genotype-by-environment interactions are emergent properties of process-based simulation models

Process-based simulation models are mathematical surrogates, and one of their functions is to describe the interconnections and feedback regulations between subsystem components (e.g. organs or tissues) and biological processes (e.g. photosynthesis or protein synthesis). The notion of a single limiting factor is thereby replaced by the idea of a sequence and/or network of different limitations operating through the plant’s life cycle. These interconnections and feedback regulations among the system components generate unexpected global system properties, called emergent properties, which do not appear when the subsystems are individually considered (Trewawas, 2006). G×E interactions are emergent properties of the whole system in which several processes interact, though they can also operate at the process level.

At the process level, G×E interactions occur when expression of the genotypic variation of a process depends on the environment. Three types of G×E interactions can be observed (Fig. 1). The first type arises from genotype-dependent responses to an environmental variable (Fig. 1a). This is the case of fruit or grain demand for carbon, which is driven by temperature, but the temperature response curve is genotype specific (Lescourret et al., 1998). A more complex situation, although very frequent in plants, is when a process does not depend directly on environmental variables, but on plant variables which themselves depend on environmental variables. In that case, the response of the process to the plant variable may be unique whatever the genotype (Fig. 1b). This is the case of light-saturated photosynthesis which depends on leaf carbohydrate reserve through a unique response curve, as shown for peach trees (Pruunus persica L.; Quilot et al., 2002). In such a case, a G×E interaction arises if the plant variable depends both on the genotype and on the environment, and the process intensity varies with the plant variable (Fig. 1b). The third type of interaction at the process level arises when the response curve of the process to the plant variable also depends on the genotype, which allows interactions for the same reason as in the previous case, but also because the response of the studied process to the plant variable depends on the genotype (Fig. 1c). This is the case of leaf photosynthesis response to leaf nitrogen content in rice which is genotype dependent, as illustrated in Yin and
Struik (2008). These three types of interactions can act together in the whole system allowing either the emergence of strong interactions at the whole system level or the loss of interactions when several processes respond in an opposite way to genotype or environmental variations.

Interactions can also emerge at the level of the whole system. This can best be illustrated with a simple theoretical model having just two processes, each of them varying linearly with an environmental variable without any G×E interactions (i.e. the slope of the relationship between the process and the environmental variable is independent of the genotype; Fig. 2a, b), and where an output (or intermediate) variable is the product of these two processes. If, for at least one of these two processes, the y-intercept of the relationship depends on the genotype, then the output (or intermediate) variable varies non-linearly in response to environmental variations, and as a consequence, a G×E interaction emerges at the whole system level (Fig. 2c).

Which models and which parameters to link genotype to phenotype?

Some mechanistic models to integrate physiological knowledge

Most growth simulation models currently used were originally developed for agronomic applications. These models were constructed using empirical response curves or laws describing the relationship between plant growth and environmental conditions and management practices (e.g. N dilution curve; Lemaire and Gastal, 1997). Although these process-control oriented models are robust, they are less plastic than real plants, hence restricting their capabilities to accurately represent the wide range of plant responses to environmental and genetic variations.

Predicting complex issues such as product quality traits in relation to G×E interactions, requires the design of mechanistic models: such models have to describe physiological processes and their response to variations in environmental conditions, to allow physiological feedback features and the integration of information from different organizational levels, for instance from cell to organ (Struik et al., 2005; Génard et al., 2007). According to Chapman et al. (2003), models need to produce emergent properties, i.e. they should be able to handle perturbations to any process and self-correct, as do real plants. As stated by these authors, this philosophy of modelling the principles of responses and feedbacks infers that models should be able to express complex behaviours. Such models of fruit and seed or grain quality have been developed over the last decade. Their current state has recently been reviewed (van Ittersum and Donatelli, 2003; Bertin et al., 2006; Martre et al., 2009): for
wheat and seeds, they focus on size and protein concentration and composition (e.g. starch, gluten-forming proteins, albumin proteins), which are the most important criterion determining the end-use value of the product. For fresh fruits, models of quality describe the main processes controlling fruit size, and sugar and acid composition, which are largely involved in flavour perception. In these models, the environment is characterized through the measurement of environmental variables (temperature, light, humidity, mineral nutrient availability etc), and the plant is characterized by developmental variables (date of flowering and number of flowering nodes etc), growth variables (biomass, amount of retrieved nitrogen etc) and metabolic variables (respiration, sugar synthesis etc).

In such mechanistic models, it is now possible to link model parameters with physiological traits or process and to link them with loci or genes. Below, the constraints on model parameters are defined. The relatively low number of parameters (a few tens to a couple of hundreds in most simulation models) in comparison with the 20 000 to 40 000 genes of a plant can be explained by the co-ordinated action of groups of genes, and the lower effect of a gene expression when it is replaced in its metabolic pathway, at the cell, organ or plant level (Salon and Vance, 2004, Fernie et al., 2005). The set of interconnected processes controlled by such a group of genes was defined by Tardieu (2003) as ‘meta-mechanism’.

Genotypic and generic parameters of simulation models

Though plant traits are generally dependent on genotype, environment, and management, the parameters of the equations describing these meta-mechanisms are, ideally, independent of the environment and management. One can distinguish two types of parameters in a process-based simulation model: the genotypic parameters and the generic parameters. Generic parameters do not significantly vary among genotypes, or even among species. As such, in the peach growth model (Quilot et al., 2002), the light-saturated photosynthesis which determines the maximum dry matter accumulation is a generic parameter since it does not vary among peach genotypes and even seems stable in the Prunus genus. By contrast, genotypic parameters (also called genetic coefficients) are model parameters, (i) of which values show a significant range of variation among the studied genotypes, and (ii) which have significant influences on model outputs (Boote et al., 2001), and thus are likely to induce changes in important emergent properties. The set of genotypic parameters defining a particular genotype represents a phenotypic fingerprint of this genotype. To be considered as a ‘good’ genotypic parameter, a model parameter must be precisely estimated and, ideally, with low labour cost to be estimated on a large number of genotypes. It should be process-based and, ideally, the availability of mutants for this parameter would allow the validation of its theoretical variations in the model. Examples of ‘good’ genotypic parameters are illustrated later.

Fig. 2. Schematic representation of G×E interactions at the whole system level. A simple system with two processes, f (a) and h (b), is considered. These two processes vary linearly with the environment (E), and depend on the genotype without any G×E interaction. The variations of $f$ and $h$ are described by the following equations: $f(E, G) = aE + g_1(G)$ and $h(E, G) = bE + g_2(G)$ where $a$ and $b$ are two generic (constant) parameters, and $g_1(G)$ and $g_2(G)$ are two parameters which depend on the genotype. A quality variable ($y$) which depends on these two processes according to the equation: $y = f(E, G) \times h(E, G)$, is considered (c). $y$ responds non-linearly to variations of the environmental variable and a G×E interaction emerges at the whole system level.
Sensitivity analysis of the model to its parameters can help in identifying important genotypic parameters, and their putative effects under different climate and management practices. Sensitivity analysis of the peach growth model followed by the analysis of the variation among genotypes in the values of the most important parameters, showed that among the 40 parameters of this model, only 10 are genotypic key parameters (Quilot et al., 2005a).

**From genotype to phenotype**

Making links between model parameters and genes or QTLs, implies that the model captures sufficient details and physiological functionalities, necessary to simulate the expression of single genes or a gene network. Several attempts and approaches have been proposed to integrate genetic information into process-based models and to go from genome to organ (Reymond et al., 2003), plant (Quilot et al., 2005b) or crop (White and Hoogenboom, 1996; Chapman et al., 2003). These approaches are briefly reviewed below.

**Gene-based modelling**

Because genotypic parameters are independent of the environment, one can theoretically predict their value knowing the genotype. On this basis, genetic information can be integrated into simulation models. The phenotype can then be simulated *in silico* under various environmental and management conditions. This approach was pioneered by White and Hoogenboom (1996) and Hoogenboom and White (2003) who replaced genotypic parameters of the BEANGRO simulation model for common bean (*Phaseolus vulgaris* L.) by linear functions describing the effect of eight genes affecting phenology (*ppd, Hr*, and *Tip*), growth habit (*Fin* and *Fd*) and seed size (*Ssz-1, Ssz-2*, and *Ssz-3*). The genotypes of 30 common bean cultivars were determined for these genes, with two alleles for each gene, one dominant (coded 1) and the other one recessive (coded 0). The new model, GeneGro, simulated growth and development as well as BEANGRO, and could even simulate new *G*×*E* interactions, providing major simplifications since the 30 genotypic parameters of the BEANGRO model were replaced by only eight binary coefficients, the eight loci. As pointed out by these authors, the genotypes can usually be determined more precisely by these coefficients than by field-determined genotypic parameters, so gene-based models should also reduce uncertainties in the calibration of simulation models and facilitate model calibration for new genotypes. This approach has recently been included into the soybean simulation model CROPGRO-soybean to characterize the effect of six loci on growth and development, using a set of isogenic lines (Messina et al., 2006).

These studies suggest that only a few genes need to be characterized in order to simulate genetic variations and *G*×*E* interactions for complex traits such as seed size, challenging the current view that quantitative traits have polygenic inheritance. However, the genes included in the BEANGRO and CROPGRO-soybean models may represent the action of co-regulated groups of genes and thus they are similar to estimates of QTL. Indeed the three hypothetical genes controlling seed size for common bean introduced in GenGro were mainly inferred from QTL studies. The next steps would be to simulate the regulation of gene expression(s) and the effects of genes at the process level rather than through cultivar coefficients.

When a trait is controlled by a low number of major genes, *bottom-up* modelling of a gene network can also be attempted. Such an approach has been successfully used to model flowering time (Welch et al., 2003, 2004) and cell cycle and expansion in leaves (Beemster et al., 2006) for *Arabidopsis* (*Arabidopsis thaliana* (L.) Heynh.), but it has not been applied to fruit or grain quality traits yet. Although gene networks have the potential for becoming overly complex, there is probably no need to model full networks, as it should be possible to extract simple rules which could capture the effects of major genes involved in the network.

Currently, the strongest limitation to develop a gene-based model for complex traits, is the lack of knowledge and characterization of specific genes or loci controlling these traits, including epistatic interactions and pleiotropic effects, to define the *phenotypic fingerprint* of cultivars for genotypic parameters. Moreover, detailed studies to quantify the environmental effects on gene expression and gene action are also required.

**QTL-based modelling**

In the absence of information on specific genes or loci, QTL analysis can be performed on model parameters considered as quantitative physiological traits. Then, for each genotype of a mapping population, the values of genotypic parameters can thus be predicted based on the allelic composition at the molecular markers flanking the QTLs, taking into account interactions between alleles and among QTLs (dominance, additivity, and epistasis). This approach was pioneered by Yin et al. (1999, 2000), who recalculated the value of 10 genotypic parameters of the SYP-BL simulation model for barley, related to crop growth. The major weakness of this approach was the ability of the original model to simulate observed variations. This work has been extended to barley (Yin et al., 2005) and rice (*Oryza sativa* L.; Nakagawa et al., 2005) phenology. The QTL-based models were able to simulate the phenology of recombinant inbred lines in new environments. QTL-based models were also developed to analyse the genetic variability of leaf elongation rate for maize in response to temperature and soil water deficit (Reymond et al., 2003, 2004). In these studies, a simple static model based on response curves of leaf elongation rate to temperature, vapour pressure deficit, and soil water potential was used. Thirteen maize lines grown under six contrasted environments were used as material for validating the model, which accounted for 74% of the genetic and environmental variations of leaf elongation rate (Reymond et al., 2004).

More recently, this approach was extended to peach fruit quality. Based on the sensitivity analysis of the virtual
peach fruit model and the analysis of the variation among 139 genotypes in the values of the most important parameters (Quilot et al., 2005a), 10 genotypic parameters which strongly affect fruit growth and sugar accumulation were selected among the 40 parameters of the model, for a QTL analysis (Quilot et al., 2005b). These genotypic parameters were substituted in the simulation model by the sum of QTL effects. The model was then able to account for a large part of the genetic and environmental variations in fruit size (observed and predicted values of fruit dry mass showed a correlation coefficient of 0.55). In this example, the QTL analysis of the genotypic parameters gave some insight on the processes that control quality traits, as they co-localized on the genetic map with QTLs for fruit size and sugar content. This suggests putative physiological interpretations of the functions of genes under these QTLs. For instance on the linkage groups 1 and 7 of the peach genetic map, QTLs for fruit fresh mass were located at the same regions as QTLs for genetic coefficients involved in sugar metabolism, pulp carbon growth, and in and out water fluxes in the fruit (Quilot et al., 2005b).

Analysis of G×E interactions through process-based simulation models: case-study of fruit and grain quality traits

The significance of process-based simulation models of quality traits can be exemplified through the analysis of both genotypic and environmental effects, using two integrated models of quality, one for peach fruit and the other for wheat grain.

Predicting peach fruit size and sweetness index in response to tree management and genetic variations under variable environmental conditions

The virtual peach fruit model of Lescourret and Genard (2005) predicts the fresh mass, dry matter content, and sugar composition of fruit, in response to environmental fluctuations, by coupling three main sub-models (Fig. 3): (i) a carbon sub-model which calculates the daily carbon availability (assimilation and remobilization from reserves) and allocates carbon among vegetative and reproductive organs, based on organ demand and priority rules. It thus determines the daily carbon flux to any average fruit of the

![Fig. 3. Schematic representation of the virtual peach fruit model (Lescourret and Génard, 2005). Main quality traits predicted by the model are flesh and stone dry and fresh masses, flesh dry matter and total sugar contents, and fruit sugar content. The model integrates three sub-models: a carbon sub-model simulates carbon partitioning based on organ demand and priority rules. Sugar accumulation is simulated by a metabolic sub-model that predicts the increase of total sugar content in the flesh during fruit growth. Fruit fresh mass accumulation is simulated by a water sub-model that predicts water fluxes as a function of the osmotic and turgor pressures, bio-rheological cell wall properties, and hydraulic conductivity.](http://jxb.oxfordjournals.org/content/doi/10.1093/jxb/eri242/21593034)
stem; (ii) the SUGAR model (Génard and Souty, 1996; Génard et al., 2003) which uses this daily influx of carbon as input to simulate the metabolic transformations among respired CO₂, individual sugar metabolism (sucrose, sorbitol, glucose, and fructose) and other compounds (structural carbohydrates); (iii) a water sub-model (Fishman and Génard, 1998) that predicts the water fluxes and tissue expansion as a function of the osmotic and turgor pressures, bio-rheological cell wall properties, and hydraulic conductivity. The fruit osmotic pressure is calculated from sugar content and composition simulated by the SUGAR submodel. The rules of communication among the three submodels are detailed in Lescourret and Genard (2005). The model predicts fairly well the fruit fresh and sweetness index in response to wide ranges of tree management practices (leaf to fruit ratio) and environmental conditions (year; Fig. 4a, b). As experimentally observed by Génard et al. (2003), under low fruit-to-leaf ratios the model predicts that fruit sugar content increases with fruit size. The simulated sweetness was slightly underestimated for the sweetest fruits, hypothetically because of an over-simplified description of carbon metabolism in the SUGAR model.

The ability of the virtual peach fruit model to simulate genetic variations of fruit quality has been investigated for a mapping population of 139 hybrid lines (Fig. 4c, d). Ten genotypic parameters have been assessed for each line (Quilot et al., 2005a). They allowed 95% and 52%, respectively, of the observed genetic variations in fruit fresh mass and sweetness index to be explained.

Considering the ability of such models to simulate the effects of crop management or genetic variations, they can be used to analyse virtual variations in the system. Let us consider ten virtual genotypes differing by a single genotypic trait involved in carbon allocation among sink organs, i.e. their potential fruit dry mass which ranged from 10 g to 100 g. This parameter is positively linked to the individual fruit demand (equation 1 in Lescourret and Génard, 2005). The environment was manipulated through plant management, i.e. fruit pruning, which varied from 1 to 10 fruits per branch. This factor affects the level of competition among growing sinks as the carbon available for growth has to be shared among individual fruits. The model was run for the 100 combinations of these two factors, and simulations of fruit fresh mass and sweetness at harvest were analysed. The

---

**Fig. 4.** Predicted values at maturity plotted against corresponding observed values for fresh fruit mass (a, c) and sweetness index (b, d) collected in different experiments on peach fruit exploring the crop management (a, b) and genetic (c, d) variability. (a, b) Data (cv. Suncrest) represent individual fruits sampled on experiments carried out over three years and at different leaf-to-fruit ratios. Simulations were performed with a single set of parameters given in Lescourret and Génard (2005). (c, d) Data represent mean fruit value over five fruits for 139 hybrid lines of a mapping population obtained from the second backcross of an interspecific Prunus persica L. (Batsch)×Prunus davidiana cross. Simulations were performed considering 10 genotypic parameters (Quilot et al., 2005a). Statistics concern the linear regression between observed and simulated data. RRMSE indicates the relative root mean squared error. Solid lines are y=x.
model predicted a significant management-by-genotype interaction (Fig. 5) emerging from multiple interactions among the processes involved in both traits, as more simply illustrated in Fig. 2 for two processes. These interactions induced a cascade of effects, including feedback effects which cannot be intuitively predicted. Finally, fruit fresh mass and fruit sweetness differently responded to fruit load according to the potential dry mass. In good agreement with observed data (Johnson and Handley, 1989), these simulations clearly show that accurate management of fruit load is essential for genotypes with high fruit size potential to avoid any detrimental effects on quality. Moreover, the behaviour of intermediate variables of the model could help us to understand the physiological causes of such interactions.

Both the fruit mass and fruit sweetness (deduced from individual sugar contents) can be affected by changes in the carbon and water balances. In the case of low potential genotypes, fruit mass was limited by the demand whatever

![Graph](http://jxb.oxfordjournals.org/)

**Fig. 5.** Changes in simulated fruit mass (a) and sweetness (b) in response to modifications of the number of fruits per shoot. The different lines represent model simulations for ten virtual genotypes with an increasing potential fruit dry mass from low (10 g) to high (100 g). Simulations were performed with the virtual peach fruit model described by Lescourret and Génard (2005) for representative climatic conditions for south-east France.

the number of fruits per shoot. The carbon flow, intermediate variable between the CARBON and SUGAR sub-models (Fig. 3) was unchanged, as well as the sugar accumulation, resulting osmotic pressure, and water influx. Thus, the fruit sweetness remained independent of fruit load. This was observed for genotypes with a potential fruit dry mass up to 30 g. For genotypes with higher fruit potential dry mass, both traits decreased as fruit load increased, and fruit growth was limited by the carbon supply. Consequently, the carbon flow to individual fruits and thus the accumulation of sugars decreased, as well as the osmotic pressure. This contributed to reduce the fruit mass in the WATER sub-model, which could be expected to attenuate the decrease in sweetness resulting from low sugar accumulation, by limiting the dilution effect. However, the absence of proportionality among all these processes and the numerous feedback effects made it difficult to analyse this response quantitatively. For instance, the slope break of the sweetness curve observed at a fruit load around 6–7 fruits per branch is difficult to explain, as many factors might be involved, such as transpiration, osmotic regulation, carbon partitioning between structural and soluble compounds etc.

This example emphasizes that capturing and unravelling the cascade of effects behind complex behaviours is not obvious, because numerous feedback effects and interactions among the physiological processes occur during fruit development. This example also illustrates how unexpected behaviours of complex systems can be predicted and analysed by process-based simulation models. Such simulation analysis can help in selecting the management (fruit thinning) best adapted to particular genotypes to meet specific objectives of yield and quality. Then, the main question is “can we trust the model?” The virtual fruit model and its sub-models have been published and evaluated before (Fishman and Génard, 1998; Génard et al., 1998, 2003; Lescourret et al., 1998; Quilot et al., 2002; Lescourret and Génard, 2005). Moreover, the predicted responses are in agreement with previous experimental observations (Johnson and Handley, 1989; Génard et al., 2003).

**Predicting wheat grain yield and protein concentration in response to crop management and genetic variations under variable environmental conditions**

The wheat simulation model *SiriusQuality*1 has been developed to analyse the responses of wheat crops to both environmental and genetic variations (Martre et al., 2006), it is based on the crop simulation model Sirius (Jamieson et al., 1998; Jamieson and Semenov, 2000). It consists of several sub-models describing soil water and nitrogen balances and crop development, canopy expansion, biomass, and N accumulation and partitioning, including responses to shortages in the supply of soil water and nitrogen (Fig. 6). Canopy development is simulated as a series of leaf layers associated with individual main stem leaves, and tiller production is simulated through the potential size of any layer. Each leaf layer within the
canopy intercepts light and uses it to produce biomass at an efficiency calculated from temperature, CO₂ concentration, soil water status, and the ratio of diffuse to direct radiation. Carbon and nitrogen allocation within the plant is then calculated as a function of resource (light, nitrogen, and water) availability using simple priority rules. After anthesis, the translocation of carbon and nitrogen to grains is first driven by the division of the endosperm cell, then, during the linear period of grain filling, it mainly results from the accumulation of starch and storage proteins (Martre et al., 2006). As illustrated in Fig. 6, the allocation of carbon and nitrogen to the different plant organs results from several interrelated feedback regulations. This model has been calibrated and evaluated for several modern wheat cultivars and tested in many environments and climates, including conditions of climate change (Jamieson et al., 1998, 2000; Jamieson and Semenov, 2000; Martre et al., 2006). The ability of SiriusQuality1 to simulate grain yield, nitrogen yield, and protein concentration variations for both bread wheat and durum wheat (T. turgidum L. subsp. durum (Desf.) Husn.) in response to crop management (e.g. sowing date and N fertilization) and environmental conditions (e.g. temperature and water supply) is illustrated in Fig. 7.

This model was used to analyse the effects on bread wheat (Triticum aestivum L.) grain yield and protein concentration of interactions between the environment and a genotypic trait, i.e. the maximum stem nitrogen concentration, which was decreased or increased by 30% compared to its default value. The nitrogen storage capacity of the stem is an important trait determining wheat nitrogen use efficiency and grain protein concentration (Foulkes et al., 1998). In a wheat crop, at anthesis approximately 50% of total crop nitrogen is stored in the stem and over 80% of this nitrogen is translocated to the grain during the grain-filling period. Environmental variations were taken into account by performing the analysis for 100 years of weather at three different sites representing the diversity of the climate in the European wheat growing areas and at two nitrogen supplies (Fig. 8).
At high nitrogen supply (Fig. 8a, c, e), a decrease in the stem nitrogen concentration lessened the protein concentration without reducing the grain yield, except for the driest years at Seville where water deficit severally reduced grain yield (Fig. 8e). Therefore, under high nitrogen inputs, changing the nitrogen storage capacity of the stem may significantly shift the negative relationship between grain yield and protein concentration under most European weather conditions (Martre et al., 2007). In contrast, at low nitrogen supply (Fig. 8b, d, f), changes in grain protein concentration in response to variations of the stem nitrogen storage capacity depended on the year, independently of the
grain yield, and, on average, it had no strong effect on grain protein concentration. The response of the grain protein concentration under low nitrogen supply clearly illustrates the power of a system analysis based on simulations, since such G×E interactions may be difficult to capture in the field under a restricted number of environmental conditions.

These two examples illustrate the potential of process-based simulation models to analyse (i) the effect of simple physiological parameters on complex traits using a system approach, and (ii) the interactions among the system components. However, some care is required when using crop simulation models to assess the effects of a particular trait, since the ability of a model to predict subtle G×E interactions depends on the simplifications and assumptions made in the model (Boote et al., 2001). On the other hand, simulation models allow us to focus on the most important aspects of the physiology and to reveal complex interactions that were not intuitive.

**Outlook**

Recent advances in our knowledge of the genetic, environmental, and management control of harvested organ size and composition have led to the suggestion that their phenotypic control will become a real possibility in the near future. Several process-based simulation models are now able to predict some quality traits of crop production as a function of genotype and environment, and thus they could be used to take a first step towards the analysis of G×E interactions. However, to go further in this analysis, it is still necessary to enlarge the ability of the model to simulate the complexity of plant and organ functioning. Indeed, most of the current models are restricted to the description of phenology, growth, and carbon–nitrogen–water balance at the plant or crop level. There is an urgent need for models that are able to simulate important aspects of metabolism and biophysical behaviour at the plant and organ levels (Struik et al., 2005, 2007; Génard et al., 2007).

Great attention should be paid to the uncertainty of model inputs, for instance, soil characteristics (including their spatial heterogeneity) or plant endogenous variables, when using simulation models to analyse the effect of genetic variations. Indeed, changes in yield or quality traits in response to genetic variations are often relatively small and the uncertainty in model inputs may limit the use of simulation models in predicting quantitatively the phenotype from the genotype.

The gap between detailed information emerging from sciences at the gene and cell levels and the quantitative understanding of the whole plant physiology is still large. Models can provide a platform that can accommodate new advances in this field science (Di Ventura et al., 2006). But, concomitantly to the development of process-based simulation models, more information is needed on the genetic control of the processes described in these models. In particular, quantitative data are still lacking to understand how gene actions are coordinated during plant development, and in response to environmental signals. Adequate data sets, with time series during fruit or grain filling and adequate description and characterization of the growing conditions and the genotypes are required. The production of isogenic lines or mutants, which may be experimentally described under well characterized environments, will play an important role for unravelling the physiological processes and G×E interactions involved in the control of quality traits.

**Acknowledgements**

This work was funded and carried out under the ‘Environmental determinism of harvested organ quality’ research network from INRA, Division of Environment and Agronomy.

**References**


