



HAL
open science

Reducing a model of sugar metabolism in peach to catch different patterns among genotypes

Hussein Kanso, Bénédicte Quilot-Turion, Mohamed-Mahmoud Memah,
Olivier Bernard, Jean-Luc Gouzé, Valentina Baldazzi

► To cite this version:

Hussein Kanso, Bénédicte Quilot-Turion, Mohamed-Mahmoud Memah, Olivier Bernard, Jean-Luc Gouzé, et al.. Reducing a model of sugar metabolism in peach to catch different patterns among genotypes. *Mathematical Biosciences*, 2020, 321, pp.108321. 10.1016/j.mbs.2020.108321 . hal-02922359

HAL Id: hal-02922359

<https://hal.inrae.fr/hal-02922359v1>

Submitted on 26 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Reducing a model of sugar metabolism in peach to catch different patterns among genotypes

Hussein Kanso^{a,b}, Bénédicte Quilot-Turion^a, Mohamed-Mahmoud Memah^b,
Olivier Bernard^d, Jean-Luc Gouzé^d, Valentina Baldazzi^{b,c,d,*}

^a*INRAE, GAFL, F-84143, Montfavet, France*

^b*INRAE, PSH, F-84914, Avignon, France*

^c*Université Côte d'Azur, INRAE, CNRS, ISA, Sophia-Antipolis, France*

^d*Université Côte d'Azur, Inria, INRAE, Sorbonne Université, BIOCORE, Sophia-Antipolis, France*

Abstract

Several studies have been conducted to understand the dynamic of primary metabolisms in fruit by translating them into mathematics models. An ODE kinetic model of sugar metabolism has been developed by Desnoues et al. [1] to simulate the accumulation of different sugars during peach fruit development. Two major drawbacks of this model are (a) the number of parameters to calibrate and (b) its integration time that can be long due to non-linearity and time-dependent input functions. Together, these issues hamper the use of the model for a large panel of genotypes, for which few data are available. In this paper, we present a model reduction scheme that explicitly addresses the specificity of genetic studies in that: i) it yields a reduced model that is adapted to the whole expected genetic diversity ii) it maintains network structure and variable identity, in order to facilitate biological interpretation. The proposed approach is based on the combination and the systematic evaluation of different reduction methods. Thus, we combined multivariate sensitivity analysis, structural simplification and timescale-based approaches to simplify the number and the structure of ordinary differential equations of the model. The original and reduced models were compared based on three criteria, namely the corrected Akaike Information Criterion (AIC_C), the calibration time and the expected error of the reduced model over a progeny of virtual genotypes. The resulting reduced model not only reproduces the

*Corresponding author

predictions of the original one but presents many advantages including a reduced number of parameters to be estimated and shorter calibration time, opening new promising perspectives for genetic studies and virtual breeding. The validity of the reduced model was further evaluated by calibration on 30 additional genotypes of an inter-specific peach progeny for which few data were available.

Keywords: model reduction, sensitivity analysis, structural simplification, quasi-steady-state, peach fruit, kinetic model, model calibration, gene-to-phenotype.

1. Introduction

Plants are sessile organisms endowed with the capacity to alter their development, physiology, and morphology depending on the context. Plant phenotype is the result of the interaction between the environment, cultural practices and plant's genetic background (genotype). In the context of agronomy, increasing efforts have been made to select varieties that better meet consumers' expectations. Today it is clear that future breeding should account for complex plant phenotypes, responding to a large panel of criteria, including increased yield, abiotic and biotic stress tolerance, and quality of food products.

Genotype-phenotype models have been considered as the tools of the future to design new genotypes since they can help to test the performance of new genotypes (G) under different Environments (E) x Management (M) conditions. The challenge is to build ecophysiological models that integrate genetic information associated to specific processes (traits). In general, genotypes are defined by a set of parameters, which depends on gene expression or allelic combination, depending on the genetic complexity of the considered trait as well as the available information [2]. Genetic-improved ecophysiological models can then be used to capture GxExM interactions. They can also be used to design "ideotypes" i.e. real or virtual plant cultivars expressing an ideal phenotype adapted to a particular biophysical environment, crop management, and end-use [3, 4]. For this, it is necessary to combine the genetic-improved ecophysiological model with a multi-objective optimization algorithm to identify the best genotypes for specific conditions [5].

Construction of gene-to-phenotype models is challenging. First, the approach requires that a sole and unique model can reproduce the behavior of

27 all genotypes, in multiple environments, the diversity observed being sup-
28 ported by different sets of parameters. Second, calibration of the models for
29 a large number of genotypes is generally difficult, due to a large number of
30 parameters (typically from 50 to 200 in whole-plant ecophysiological mod-
31 els) along with a restricted number of observations [6, 7]. Due to the model
32 complexity and non-linearities, evolutionary and bio-inspired algorithms are
33 increasingly used both for parameter estimation and ideotype design. These
34 methods can explore high-dimensional parameter space efficiently but they
35 rely on a large number of model evaluations, that can rapidly increase the
36 computational time required to find a solution. Third, the genetic architec-
37 ture of complex traits can be very complex, due to epistatic and pleiotropic
38 effects. In this sense, the presence of biologically-meaningful parameters can
39 considerably help the interpretation of the resulting genetic architecture, fa-
40 cilitating the breeding process. Ideally, most the model is close to omics data,
41 the easier the linkage between the parameters and the underlying physiolog-
42 ical processes.

43 Kinetic modeling has been successfully applied to several metabolic path-
44 ways in plants [8, 9, 10]. In this spirit, a kinetic model of sugar metabolism
45 has been developed in [1] to simulate the accumulation of different sugars
46 during peach fruit development. The model correctly accounts for annual
47 variability and the genotypic variations observed in ten genotypes derived
48 from a larger progeny of inter-specific peach cross. At term, the objective
49 of the research is to integrate the genetic control of sugar metabolism in
50 this kinetic model and develop a methodology to design ideotypes by vir-
51 tual breeding. To achieve this, it is necessary to estimate accurately the
52 values of the influential parameters of the model for the whole progeny of
53 106 genotypes for which few data are available. Unfortunately, the size of
54 the parameter space and the non-linearity of the reaction rates make the
55 calibration of the model unreliable and time-consuming.

56 One way to face these weaknesses is to reduce the complexity of the model
57 [11]. Several reductions and approximation approaches exist in the literature,
58 each one addressing a specific aspect of model complexity [12, 13]. A number
59 of methods, such the lumping method [14, 15] or the classical quasi-steady-
60 state (QSS) approaches, aim at reducing the number of variables based on
61 chemical or time-scale considerations [16, 17]. Methods from sensitivity anal-
62 ysis may help to reduce the parameter space by identifying non-influential
63 parameters, whose values can be fixed by broad literature data [18, 19, 20, 21].
64 Last but not least, the structure of the model itself can be simplified. Meth-

65 ods for model decomposition [22, 23, 24] aim to separate the system into
66 sub-networks or sub-models, that are easier to analyze and parameterize.
67 The choice of reaction kinetics is also very important for model complexity.
68 In this perspective, the use of simplified enzyme kinetics [25, 26, 27] may be
69 useful to avoid the emergence of numerical and identifiability issues.

70 Different reduction methods can be combined together. In [28] for in-
71 stance, model decomposition is associated to variable transformation, re-
72 sulting in a low-dimensional description of the “exterior” part of the system,
73 whereas in [15] time scale analysis is used to identify a cluster of fast variables
74 to be lumped together.

75 In the work of Apri et al. [29] different reduction steps (parameter re-
76 moval, node removal, variable lumping) are sequentially tested following a
77 practical scheme: at each step, if the reduced model, after parameter re-
78 estimation, can reproduce some target outputs, the modification is selected,
79 and rejected otherwise. From the point of view of genetic applications, a
80 major drawback of the approach of Apri et al. [29] is that the selection of
81 acceptable reduction results depends on the specific target dynamics.

82 As a consequence, different target outputs (i.e. genotypes) can give rise
83 to reduced models with different structures or parameters number, making
84 their comparison difficult in the perspective of genetic studies.

85 The objective of this work was to provide a method to build a reduced
86 model that is adapted to the specificity of genetic studies in that: i) it yields
87 a reduced model that is adapted to the whole expected genetic diversity ii) it
88 maintains network structure and variable identity, in order to facilitate the
89 biological interpretation of the reduced model.

90 Similarly to the approach of Apri et al. [29], our reduction strategy tests
91 different methods in several *parallel* steps that, if retained, are combined
92 together into a final reduced model (Fig. 1).

93 First, multivariate sensitivity analysis was attempted to reduce the pa-
94 rameter space [30]. Second, we tried to simplify the structure of the model
95 by reducing non-linearity and time-dependent forcing, and finally, a quasi-
96 steady-state approximation based on time-scale separation was tested to re-
97 duce the size of the system. Particular attention was devoted to the system-
98 atic evaluation of the different reduction methods. Three main criteria were
99 used to assess the interest of the reduction: i) the corrected AIC value, eval-
100 uating the relative gain between model simplification and loss of accuracy
101 over an experimental dataset, ii) the calibration time, as a measure of model
102 efficiency, iii) the expected error between the original and the reduced model

103 over a population of virtual genotypes, as a measure of the reliability of the
104 simplification scheme.

105 As a case study, the proposed reduction scheme was applied to the model
106 of sugar metabolism proposed by Desnoues et al. [1]. The resulting reduced
107 model correctly reproduces data on the original 10 genotypes with only 9 es-
108 timated parameters (out of 14 in the original model) and a gain in calibration
109 time over 40%. In addition, the reduced model was successfully calibrated on
110 30 new genotypes of the same inter-specific peach progeny, for which fewer
111 data points were available.

112 The paper is organized as follows. In the next section, we briefly present
113 the original model of sugar metabolism developed by Desnoues et al. [1].
114 Section 3 is devoted to the description of the individual reduction methods,
115 whereas Sections 4 and 5 present, respectively, the datasets and the numerical
116 methods used for the assessment of the proposed model reduction. The
117 results of the application of our reduction scheme to the model of sugar
118 metabolism are reported in section 6. A general discussion on the advantages
119 and limitations of our approach closes the paper.

120 **2. Description of the peach sugar model**

121 The model developed by Desnoues et al. [1] describes the accumulation
122 of four different sugars (sucrose, glucose, fructose, and sorbitol) in peach
123 fruit during its development over a progeny of ten peach genotypes with
124 contrasting sugar composition. The fruit was assumed to behave as a single
125 big cell with two intra-cellular compartments, namely the cytosol and the
126 vacuole. Carbon enters the fruit from the plant sap which is transformed by a
127 metabolic network, including enzymatic reactions and transport mechanisms
128 between the cytosol and the vacuole.

129 The developed dynamical model made explicit use of experimental data to
130 describe the evolution of the sub-cellular compartment (due to fruit growth)
131 and enzyme activities (due to fruit developmental program) over time. To
132 this aim, measured fruit dry and fresh masses and enzyme activities were
133 represented by genotype-specific temporal functions and provided as input
134 to the model.

From a mathematical point of view, the model can be described as a set

of parametric ordinary differential equations:

$$\frac{dx}{dt} = f(x(t), I(t), v(t), p), \quad (1)$$

$$x(t_0) = x_0, \quad (2)$$

135 where t is the independent time variable in days after bloom (DAB); $x \in \mathbb{R}^{10}$
136 is the concentration vector of metabolites in the corresponding intra-cellular
137 compartment and $x_0 \in \mathbb{R}^{10}$ in Eq.(2) is the vector of the corresponding ini-
138 tial values. $I \in \mathbb{R}$ is the time-dependent input of carbon from the plant
139 and $v \in \mathbb{R}^7$ is the vector of time-dependent measured enzymatic activities;
140 $p = (p_1, \dots, p_{23})$ is the vector of parameters defining the rate reactions where
141 p_1, \dots, p_{14} have to be estimated and p_{15}, \dots, p_{23} are fixed from literature
142 data. $f(x(t), I(t), v(t), p)$ of Eq.(1) describes the change in compounds con-
143 centrations. Equations of the reduced and original model are introduced in
144 Appendix Appendix A.1.

145 3. Model reduction methods

146 In this section, we present a reduction scheme explicitly dedicated to
147 genetic studies that combines different methods in several parallel steps as
148 shown in (Fig. 1) and explained in the next subsections.

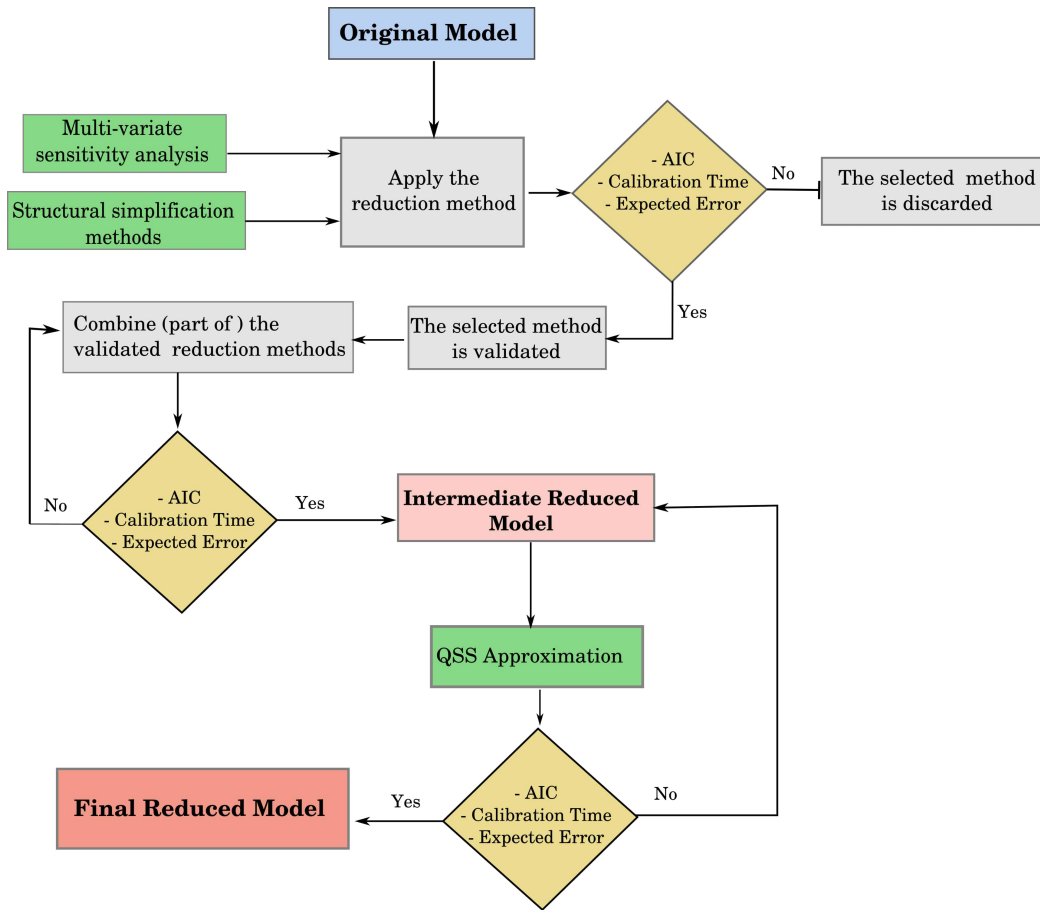


Figure 1: Graphical representation of the proposed model reduction scheme. Yellow diamonds represent model evaluation steps by means of our 3 criteria: the corrected AIC value, calibration time and expected error over a virtual population. The tested reduction methods are indicated in green. Multivariate sensitivity analysis and three structural simplification methods are independently applied to the original model and evaluated. The validated methods are then combined into an intermediate reduced model whose performances are again submitted to evaluation. Finally, the application of a QSS approximation over the intermediate reduced model is tested to yield the final reduced model.

149 *3.1. Multivariate sensitivity analysis*

150 Generally, in the case of complex models, estimating parameters requires
 151 a lot of effort and is known to be a difficult and challenging task. In par-
 152 ticular, it is tricky to determine which parameters can be fixed. The global
 153 sensitivity analysis methods allow to explore the influence of each parameter
 154 on model outputs and thus to identify the key parameters that affect model

155 performance and play important roles in model parameterization, calibra-
 156 tion and optimization [21]. Multivariate sensitivity is a method developed
 157 by Lamboni et al. [30] that allows the application of global sensitivity anal-
 158 ysis to models having a multivariate (eg. dynamic) output. The idea is to
 159 perform a principal components analysis on the outputs, and then compute
 160 the sensitivity indexes for each principal component. The results are sum-
 161 marized by the generalized sensitivity indices (GSI) that provide a unique
 162 ranking of the parameters over the whole output.

163 This method was applied to the 23 parameters of the original model
 164 and to the measured enzymatic activities v . Each parameter was studied at
 165 three levels, corresponding to 0.05, 0.5 and 0.95 quantiles of the previously
 166 estimated 14 parameters values [1] and to a variation of -20% and $+20\%$
 167 of the fixed values for the other parameters. For time-dependent enzyme
 168 activities, the same -20% and $+20\%$ variation was applied on their average
 169 values over the whole dynamics.

170 In order to evaluate the impact of the genotype choice on the results of
 171 the sensitivity analysis, simulations were performed according to a factorial
 172 design, following the ANOVA model $genotypes \times (p_1 + \dots + p_{23} + v_1 + \dots + v_7)^2$.
 173 The package "Planor" in **R** (R Development Core Team 2015) was used.
 174 The minimum resolution of the plan was fixed by using the tool MinT [31]
 175 to test all main effects and interactions. The factorial design resulted in
 176 $10 \times 3^9 = 196\ 830$ simulations.

177 Multivariate sensitivity analysis was performed independently on the dy-
 178 namics of the four output sugars (*i.e* sucrose, glucose, fructose, and sorbitol)
 179 that compose peach fruit. In order to determine the least sensitive paramet-
 180 ers, the whole sugar phenotype has to be taken into account, with respect to
 181 the relative proportions of each sugar. For this aim, an aggregate generalized
 182 sensitivity index ($aGSI$) was constructed for each parameter as

$$aGSI = \sum_{i=1}^4 GSI_i \beta_i \quad (3)$$

183 where GSI is the generalized sensitivity indice computed for the sugar i and
 184 β_i the relative proportion of sugar i in the fruit. $\beta = (0.72, 0.13, 0.09, 0.05)$
 185 for sucrose, glucose, fructose, and sorbitol, respectively.

186 3.2. Structural simplification methods

187 This section aims to simplify the structure of the model in terms of net-
 188 work and reaction rates while preserving its predictive ability. The structural

189 simplification includes the three following strategies:

190 *3.2.1. Simplifying the description of enzymatic capacities*

191 Seven enzymatic capacities V_{max} are represented in the original model.
 192 Some of these capacities were assumed to vary over time (temporal effect)
 193 and/or to depend on the phenotypic group (phenotype effect), according to
 194 experimental evidences [32]. The characteristics of enzyme capacities are
 195 summarized in Table 1. In order to simplify the model, we systematically
 196 tested the impact of the suppression of the phenotype and/or the tempo-
 197 ral effect on each single capacity. Depending on the characteristics of the
 198 considered enzyme (Table 1), the procedure is slightly different:

$$\text{Phenotype effect : } \begin{cases} V_{max}^1 \\ V_{max}^2 \end{cases} \rightarrow \frac{V_{max}^1 + V_{max}^2}{2} \quad (4)$$

$$\text{Temporal effect : } V_{max}(t) \rightarrow \langle V_{max}(t) \rangle_t \quad (5)$$

$$\text{Double effect : } (4) \text{ then } (5) \text{ applied} \quad (6)$$

199 where $\langle . \rangle_t$ stands for temporal average over the whole dynamics.

Table 1: Characteristics of enzymatic activities in [1]

V_{max}	Phenotype effect	Temporal effect
v_1	No	No
v_2	No	Yes
v_3	Yes	No
v_4	No	Yes
v_5	No	Yes
v_6	Yes	Yes
v_7	Yes	Yes

200 *3.2.2. Rate simplification*

201 In the original model, enzymatic reactions were represented by an irre-
 202 versible Michaelis-Menten (MM) equation:

$$u(x, t) = V_{max} \frac{x(t)}{K_m + x(t)} \quad (7)$$

203 where V_{max} is the enzymatic capacity. K_m is the affinity of the enzyme for
204 the substrate, $x(t)$ is the concentration of the substrate at time t .

205 The objective here is to simplify Eq.(7) in order to improve the efficiency
206 of the numerical simulation. Depending on the relative levels of the substrate
207 concentration and the MM equation affinity, two simplifications of the flows'
208 equations can be made:

209 **Case 1:** if $x(t) \ll K_m$

210 Substrate concentration is small compared to the affinity of the enzyme for
211 the substrate then we can write: $u(x, t) = \frac{V_{max}}{K_m} x(t)$.

212 **Case 2:** if $x(t) \gg K_m$

213 Substrate concentration exceeds the affinity of the enzyme for the substrate,
214 so that the enzyme can be supposed close to saturation: $u(x, t) = V_{max}$.

215 3.2.3. Futile cycle removal

216 The presence of internal cycles within a metabolic network can lead to
217 the appearance of thermodynamically unfeasible loops i.e. reactions that run
218 simultaneously in opposite directions (for example Fig. 2) and have no overall
219 effect on the exchange fluxes of the system. This is an undesirable situation
220 that causes numerical issues and makes the estimation of the corresponding
221 parameter values an ill-posed problem.

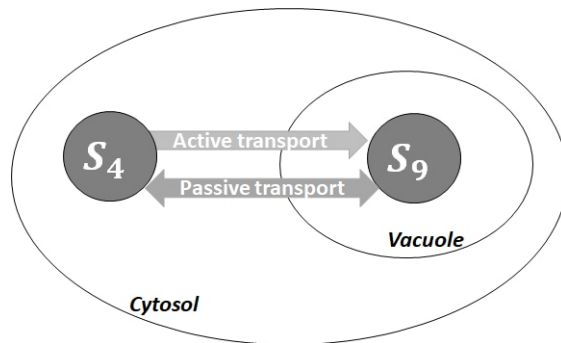


Figure 2: S_4 is the glucose in the cytosol transported to the vacuole as S_9 via an active (unidirectional transport) and passive (reversible transport).

222 In this context, our strategy was to remove each futile cycle by replacing
223 the antagonist reactions by a single effective reaction preserving the net ex-
224 change flux of the system. Different kinetics can be tested for the effective

225 reaction, as alternative reduction approaches. Consistently with the previous
 226 reduction method, we decided to test two linear reaction forms, namely

$$u(x, t) = k_i x_i - k_j x_j \quad (8)$$

227 and

$$u(x, t) = k_i (x_i - x_j) \quad (9)$$

228 where x_i, x_j are the variables involved in the futile cycle and k_i, k_j are the
 229 coefficients to be estimated.

230 3.3. Time-scale analysis and QSS approximation

231 Biological systems are often characterized by the presence of different
 232 time scales (seconds, hours, days). Following Heinrich and Schuster [17], an
 233 appropriate measure of the time scales involved is given by

$$\tau_i(t) = -\frac{1}{Re(\lambda_i(t))} \quad (10)$$

234 where $Re(\lambda_i)$ are real parts of the eigenvalues λ_i of the Jacobian matrix of the
 235 system, along a given trajectory. The presence of fast modes in the system
 236 allows the reduction of the number of variables based on a quasi-steady-state
 237 assumption.

238 Based on the above information and on the analysis of time-series of the
 239 full model, variables can be divided into two groups $x = (x^{(1)}, x^{(2)})$, where
 240 $x^{(1)}$ and $x^{(2)}$ correspond respectively to the slow and fast variables of the
 241 system [17, 33].

242 Application of the QSS approximation states that

$$\frac{dx^{(2)}}{dt} = f_2(x^{(1)}, x^{(2)}, I(t), v(t), p) = 0 \quad \rightarrow \quad x_{ss}^{(2)} = g(x^{(1)}) \quad (11)$$

243 It follows that, after a relaxation period, the system can be approximated by
 244 the reduced model:

$$\frac{dx^{(1)}}{dt} = f_1((x^{(1)}, g(x^{(1)}), I(t), v(t), p) \quad (12)$$

245 of lower dimension.

246 4. Experimental and artificial data

247 4.1. Experimental data

248 The 106 peach genotypes used in this study come from an inter-specific
249 progeny obtained by two subsequent back-crosses between *Prunus davidi-*
250 *ana* (Carr.) P1908 and *Prunus persica* (L.) Batsch ‘Summergrand’ and then
251 ‘Zephyr’ [34]. They were planted in 2001 in a completely randomized design
252 in the orchard of the INRAE Research Centre of Avignon (southern France).
253 Experimental monitoring of peach fruit growth and quality has been con-
254 ducted in 2012, as described in [32]. The concentration of different metabo-
255 lites, namely sucrose, glucose, fructose, sorbitol, and hexoses phosphates, the
256 fruit flesh fresh weight and dry matter content were measured at different
257 time points during fruit development, for all genotypes. In addition, the
258 temporal evolution of enzymatic capacities (maximal activity) of the twelve
259 enzymes involved in sugar metabolism was measured over the whole popu-
260 lation [32]. The resulting dynamic patterns were analyzed and compared
261 by means of a generalized mixed linear-effect model (GLMM). Accordingly,
262 some enzyme activities were shown to vary over time and/or depend on the
263 phenotypic group [32].

264 *Training set*

265 The 10 genotypes already used by Desnoues et al. [1] were selected as
266 the training set for our reduction strategies. They include five genotypes
267 having a ‘standard phenotype’, namely a balanced fructose-to-glucose ratio
268 at maturity between 0.6 and 0.9, and five considered to have a ‘low fructose
269 phenotype’ due to the lower proportion of fructose compared with glucose
270 based on their sugar composition at maturity [1]. For these 10 genotypes, 3
271 biological measurements are available at 6 dates after bloom.

272 The training set was used to test each reduction method individually as
273 well as their combination, based on the AIC_C value and the calibration time
274 (see section 5.3).

275 *Validation set*

276 The quality of the final reduced model was evaluated by calibration on a
277 validation set for which fewer data points were available (one single biological
278 measurement at 6 dates). The idea was to select 30 additional genotypes of
279 the inter-specific peach progeny, which in complement to the training set,
280 represented the greatest diversity in terms of growth rate and duration. For

281 this aim, experimentally measured growth curves were interpolated with a
282 smoothing spline algorithm [35] with 16.4 degrees of freedom in **R** (R Devel-
283 opment Core Team 2015) and the maximum and average growth rate quan-
284 tified as the maximum and the average of the growth curve’s derivative over
285 fruit development. A principal component analysis (PCA) was performed on
286 growth rate and growth duration for the whole progeny of 106 genotypes us-
287 ing the **R ADE4** library. The first two principal components accounted for
288 more than 90% of the genetic diversity. The first axis was mainly related to
289 the growth rate whereas the second one reflected the duration of growth. As
290 shown in Fig. 3, the ten genotypes of the original study provided a good rep-
291 resentation of the observed diversity in growth rate. However, their growth
292 duration was relatively short, compared to the existing variability. As a con-
293 sequence, most of the new genotypes have been selected in the upper-left
294 panel of the plan, in order to capture the greatest genetic diversity in terms
295 of fruit development. An equal proportion of the two phenotypic groups was
296 maintained.

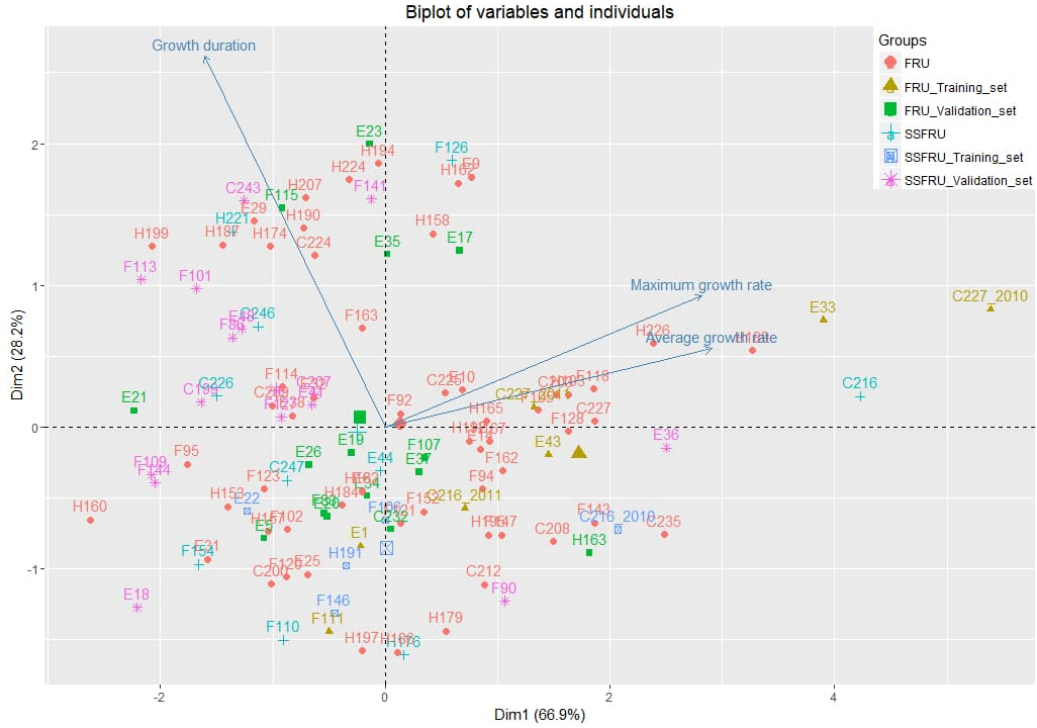


Figure 3: Principal component analysis (PCA) for the whole progeny of 106 genotypes. It represents the projection on the Dim1 and Dim2 of the growth duration and growth rate obtained with growth curves.

297 *4.2. Virtual genotypes*

298 In addition to the training set, a virtual experiment was performed to
 299 evaluate the reliability of the reduction methods to variations in parameter
 300 values, initial conditions, and input functions, expected in large genetic pop-
 301 ulations. For this aim, 20 000 virtual genotypes were generated by randomly
 302 assigning model parameters and inputs, based on data from the 10 profiles
 303 used in [1].

304 The values of the parameters p were taken randomly using a uniform
 305 distribution between the minimum and the maximum of the previously es-
 306 timated values over the set of 10 genotypes [1]. Initial conditions, such as
 307 *initial fruit weight*, and *initial sugar concentration* were assigned ran-
 308 domly using a uniform distribution within the range of observed values plus
 309 a variation of 40%.

310 Given the high correlation among parameters describing fruit growth curves

311 [36], model inputs, such as *fruit weight*, were randomly assigned using a uni-
 312 form distribution picking one of the observed growth dynamics and adding
 313 an overall random variation between zero and 10% on fruit weight. Finally,
 314 shifts in the duration of fruit development among genotypes were also consid-
 315 ered. The maturity date was chosen randomly using a uniform distribution
 316 within the range of observed dates broaden of 40%.

317 5. Numerical methods

318 5.1. Mathematical notations

- 319 • $x(t, p^{(k)})$: original model associated to parameters $p^{(k)}$ (i.e. genotype
 320 k)
- 321 • $\tilde{x}(t, \tilde{p}^{(k)})$: reduced model for the genotype k .

322 Note that the notation $\tilde{x}(t, \tilde{p}^{(k)})$ can apply to different versions of the reduced
 323 model, depending on the considered reduction method.

- 324 • $\mathcal{T}_S^{(k)}$: set of the N_S simulation times for the genotype k
- 325 • $\mathcal{T}_M^{(k)}$: set of the N_M measurement times for the genotype k
- 326 • $X^{(k)}(t_j)$: N experimental observations for the genotype k , with $t_j \in$
 327 $\mathcal{T}_M^{(k)}$. Note that $N = 4 \times N_M \times r$, where r is the number of replicates
 328 at time t_j , for the 4 different sugars (sucrose, glucose, fructose and
 329 sorbitol). $r = 3$ for the training set and $r = 1$ for the validation set.

330 5.2. Parameter estimation

In this section, we aim to estimate the parameters of the models to fit
 our observations i.e. our measured sugars concentrations. For this purpose,
 we note $X^{(k)} = (X_1^{(k)}, \dots, X_N^{(k)})$ the vector of the experimental observations
 at several times for the genotype k and suppose that:

$$\mathbb{E}(X_i^{(k)}) = \mathcal{M}_{p^{(k)}}(x_i^{(k)})$$

331 where $x_i^{(k)} = (x^{(k)}(t_i))$ is the set of system variables at $(t_i)_{i \in [1, N]}$, $p^{(k)}$ is
 332 the vector of parameters to be estimated and $\mathcal{M}_{p^{(k)}}$ is the mathematical
 333 function relying the considered model to the data (see Appendix A for more

334 information). Here, the observations $X^{(k)}$ are assumed to follow a Gaussian
 335 law $\mathcal{N}(\mathcal{M}_{p^{(k)}}(x^{(k)}), \sigma_k^2)$ with constant variance σ_k^2 .

336 The estimation of our parameters can be performed through the maxi-
 337 mization of the likelihood. We note $\ell(p^{(k)}, \sigma_k^2)$ the log-likelihood function for
 338 the genotype k .

Under the assumption of observation independence, the log-likelihood can be defined as follows:

$$\ell(p^{(k)}, \sigma_k^2) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma_k^2) - \frac{1}{2\sigma_k^2} \sum_{i=1}^N (X_i^{(k)} - \mathcal{M}_{p^{(k)}}(x_i^{(k)}))^2 \quad (13)$$

339 A maximum log-likelihood estimator $(\hat{p}^{(k)}, \hat{\sigma}_k^2)$ of $(p^{(k)}, \sigma_k^2)$ is a solution to
 340 the maximization problem:

$$(\hat{p}^{(k)}, \hat{\sigma}_k^2) = \arg \max_{p^{(k)}, \sigma_k^2} \ell(p^{(k)}, \sigma_k^2) \quad (14)$$

341 In this Gaussian case, the maximum log-likelihood estimator is thus equiv-
 342 alent to the ordinary least-square estimator:

$$\hat{p}^{(k)} = \arg \max_{p^{(k)}} \sum_{i=1}^N (X_i^{(k)} - \mathcal{M}_{p^{(k)}}(x_i^{(k)}))^2 \quad (15)$$

$$\hat{\sigma}_k^2 = \frac{1}{N} \sum_{i=1}^N (X_i^{(k)} - \mathcal{M}_{\hat{p}^{(k)}}(x_i^{(k)}))^2 \quad (16)$$

343 Matlab software (MATLAB R2018a, The MathWorks Inc., Natick, MA)
 344 was used for model integration (solver ode23tb [37]) and calibration. A ge-
 345 netic algorithm (function `ga` [38] of Global Optimisation Toolbox) was used
 346 for maximization of Eq. (15). The population size, the maximum number
 347 of generations, and the crossover probability have been respectively set at
 348 200, 300, and 0.7. For each reduced version of the model (individual or com-
 349 bined reduction methods), free parameters were numerically re-estimated.
 350 The fitting process was considered at convergence when the average relative
 351 change in the best-cost function, i.e. the sum of squared errors, value over
 352 generations was less than 10^{-6} . For each genotype k and reduced model,
 353 estimations procedure has been repeated ten times to take into account the
 354 stochastic nature of the genetic algorithm and to ensure the good exploration
 355 of the parameters' space. The solution having the best score was kept for
 356 subsequent analyses.

357 *5.3. Model selection*

358 Individual and combined reduction methods were evaluated according to
 359 three criteria of major importance for our application: the corrected Akaike
 360 Information Criterion (AIC_C), the gain in calibration time (%) and the ex-
 361 pected error (%) between the original and reduced models.

362 *Akaike Information Criterion*

363 The AIC gives information on the likelihood of the proposed model based
 364 on available experimental data and weighted by the number of free paramete-
 365 rs: [39]:

$$AIC(p) = -2\ell(p, \sigma^2) + 2n_p \quad (17)$$

366 where n_p is the number of estimated parameters p and $\ell(p, \sigma^2)$ is the max-
 367 imum log-likelihood. In this paper, we used the corrected AIC as we deal
 368 with a small set of observations and a considerable number of parameters.

$$AIC_C(p) = AIC(p) + \frac{2n_p(n_p + 1)}{N - n_p - 1} \quad (18)$$

369 where N is the number of observations. For genotype k and for each reduction
 370 method, we defined

$$\Delta_{AIC_C}^{(k)}(\tilde{p}^{(k)}, p^{(k)}) = AIC_{C_{reduced}}(\tilde{p}^{(k)}) - AIC_{C_{original}}(p^{(k)}) \quad (19)$$

371 as the AIC_C difference between the reduced and the original model. Note
 372 that Δ_{AIC_C} is always computed using the best estimated parameter solution
 373 for the considered model. Whenever the average over the 10 genotypes ($<$
 374 $\Delta_{AIC_C} >_G$) was negative, the reduction method was validated.

375 *Gain in calibration time*

376 We used the calibration of a specific genotype ($E43$) as a proxy of the
 377 maximum expected calibration time on the population. Genotype $E43$ was
 378 selected because it required a long calibration time on the original model
 379 proposed by Desnoues et al. [1] (approximately 11 hours on average on a
 380 3.1GHz Intel(R) Xeon(R) processor) but it did not suffer from numerical
 381 instabilities, that could complicate the calibration process. Note that the
 382 overall calibration time of a model depends both on the integration time of
 383 each evaluation step and on the convergence of the cost function that sets
 384 the actual number of generations performed by the algorithm. Both aspects
 385 may be affected by the model reduction.

386 To evaluate the gain in calibration time due to model reduction, pa-
 387 rameter estimation was performed for each reduction method, following the
 388 general procedure (see section 5.2), and compared to the calibration time
 389 obtained for the original model. An initial population \mathcal{P}_0 was randomly se-
 390 lected assuming a uniform distribution in the parameter range and then kept
 391 fixed for all calibration processes (both original and reduced models). For
 392 models having a reduced number of parameters, the initial population was
 393 directly derived from \mathcal{P}_0 .

394 The gain (G_t) was defined as the gain (in %) in calibration time T between
 395 the original and the reduced model:

$$G_T = \frac{T_{original} - T_{reduced}}{T_{original}} \times 100$$

396 *Expected error*

397 Simulations of the original and reduced models were compared by the
 398 Normalized Root Mean Square Error over the 10 model variables :

$$J_i(p^{(k)}, \tilde{p}^{(k)}) = \frac{\sqrt{\frac{1}{N_S} \sum_{j=1}^{N_S} (x_i(t_j, p^{(k)}) - \tilde{x}_i(t_j, \tilde{p}^{(k)}))^2}}{\max_j(x_i(t_j, p^{(k)})) - \min_j(x_i(t_j, p^{(k)}))} \quad \forall i \in \{1, \dots, 10\} \quad (20)$$

399 where $x(t, p_k)$ and $\tilde{x}(t, \tilde{p}_k)$ are the concentration predicted by the original and
 400 reduced model, respectively. Parameters for the reduced model were derived
 401 from the values of the corresponding parameters in the original model.

402 The quality of the QSS approximation was assessed by computing J_i for
 403 each variable in the model, over the whole dynamics.

404 In the context of the virtual experiment, the Expected Error (%) of the
 405 reduced model was defined as the average distance J over the virtual popu-
 406 lation:

$$\text{Expected Error} = \frac{1}{N_{VG}} \sum_{k=1}^{N_{VG}} \langle J_i(p^{(k)}, \tilde{p}^{(k)}) \rangle \times 100 \quad (21)$$

with

$$\langle J_i(p^{(k)}, \tilde{p}^{(k)}) \rangle = \frac{1}{10} \sum_{i=1}^{10} J_i(p^{(k)}, \tilde{p}^{(k)})$$

407 where N_{VG} is the number of virtual genotypes and 10 is the number of
 408 variables. In our case, $N_{VG} = 20\,000$. The Expected Error was used to
 409 quantify the reliability of the reduction.

410 **6. Results**

411 *6.1. Strategy 1: Identification of low sensitive parameters*

412 The objective of the sensitivity analysis was to identify parameters having
 413 a significant influence on the outputs of the model, over the whole dynamics
 414 and for all tested genotypes. A multivariate sensitivity analysis [30] was used
 415 for this purpose. The aggregate generalized sensitivity indices (aGSI) (see
 416 section 3.1) shown in Fig. 4 give a common ranking of model parameters
 417 according to their influence on the whole sugar phenotype, as it is made up
 418 by the four output sugars (sucrose, sorbitol, glucose, and fructose).

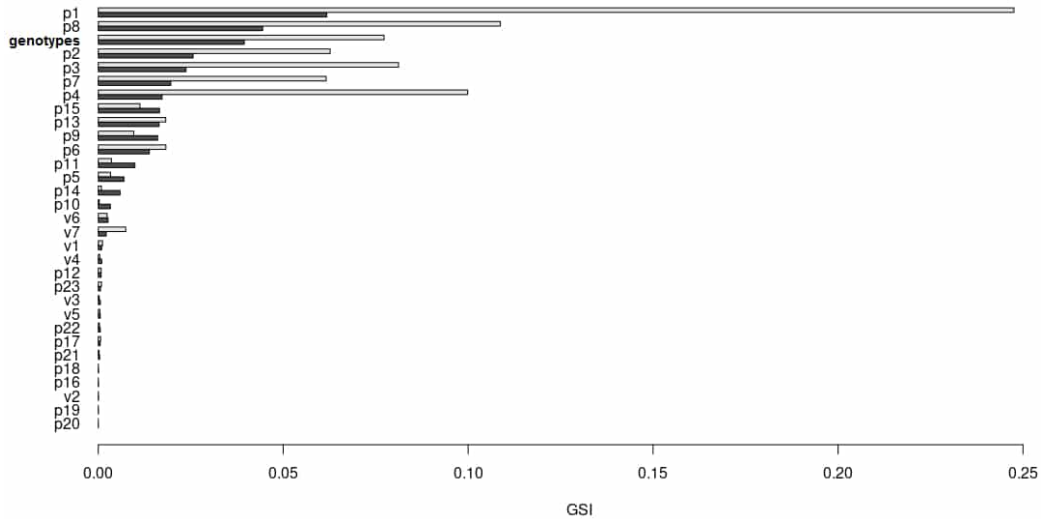


Figure 4: Aggregate Generalized sensitivity indices (aGSI) for the parameters of the model and genotypes (the training set) on four outputs (Sucrose, Sorbitol, Glucose and Fructose) of the sugar model. The main sensitivity indices are in dark bars and interaction ones are in grey bars.

419 Parameter (p_1) related to the action of cell-wall invertase in fruit apoplasm
 420 and the coefficient of sucrose import (p_8) are the most important parameters,
 421 followed by the activities of acid invertase (p_2), the activities of Fructokinase
 422 (p_3), Hexokinase (p_4) and the resynthesis rate of sucrose from hexose phos-
 423 phate (p_7). Indeed, p_1 , p_3 , and p_4 parameters are the most sensitive param-
 424 eters for sucrose, fructose and glucose concentrations respectively (see Fig.
 425 B.2).

426 Interestingly, the **genotype** factor is ranked third, meaning that it does
 427 not affect parameters' sensitivity as much as expected. A closer look at the

428 results shows that the choice of the genotype essentially affects the second
 429 principal component, via the definition of the initial conditions of the model
 430 (see the supplemental information Fig. B.1).

431 Among the 14 parameters estimated (p_1, \dots, p_{14}) in the original model,
 432 four parameters, namely p_5 , p_{10} , p_{12} and p_{14} , have a negligible effect on
 433 the four outputs, independently of the peach genotype. Accordingly, these
 434 parameters can be fixed to their nominal values i.e. their average value over
 435 the ten genotypes, without affecting the quality of predictions. The validity
 436 of such a reduction strategy was tested on the ten genotypes of the training
 437 set. The difference in Akaike criterion (Δ_{AIC_C}) between the reduced and
 438 the original models was computed for each genotype. Results presented in
 439 Table 2 show that such a reduction in the number of parameters is strongly
 440 beneficial for nine out of the ten genotypes with largely negative Δ_{AIC_C}
 441 values, and roughly neutral for one genotype ($\Delta_{AIC_C} \sim 0$). The gain in
 442 calibration time, however, is important (25%) and the expected error over
 443 the progeny of virtual genotypes is low, demonstrating a good reliability of
 444 the proposed simplification. For these reasons, the model with 10 parameters
 445 to be estimated was selected.

Table 2: Δ_{AIC_C} calculated between reduced and original models for the training set and the gain in calibration time (%) for E43. The Expected error \pm standard deviation (Std) between original and different reduced models for 20 000 virtual genotypes.

Simplification method	Δ_{AIC_C}											Calibration Time gain %	Expected Error Virtual genotypes	
	E1	E33	E43	F111	E22	F106	F146	H191	C216	C227	$< \Delta_{AIC_C} >_G$			
Low sensitive parameters fixed	-11.5	-6.4	-0.9	-14.04	-13.2	-28.3	-13.5	-14.3	-18.7	-87.7	-20.8	25.8	4.9 \pm 6.5	
V_{max} Type effect removed	v_3	-1.01	-5.9	-4.15	-4.2	1.1	-2.3	-0.3	-6.1	-6.1	-72.02	-7.9	22.4	0.5 \pm 1.3
	v_6	-0.1	-4.3	0.06	-3.9	0.7	-5.5	-0.3	-5.4	-6.1	-87.7	-11.3	26.6	1.7 \pm 1.6
	v_7	-0.7	-36.4	0.2	-6.02	1.4	-5.6	1.9	-5.2	-4.9	-94.5	-14.9	33.9	2.9 \pm 4.5
V_{max} Temporal effect removed	v_2	-0.1	-3.1	0.06	-3.7	0.7	-5.1	-0.3	-0.3	-6.3	-83.4	-10.1	31.6	0.3 \pm 0.7
	v_4	-0.8	-8.2	0.7	-5.6	-5.03	-2.5	-2.5	-5.1	-6.1	-90.3	-12.3	19.4	2.9 \pm 2.5
	v_5	0.2	-6.8	0.5	-4.8	1.8	-5.8	2.03	-2.9	-6.1	-91.1	-11.3	20.3	5.5 \pm 5.7
	v_6	-0.3	-0.4	-0.1	-27.04	1.7	-5.5	-0.2	-5.3	-5.7	-84.9	-12.7	30.5	4.1 \pm 3.1
	v_7	8.6	-25.1	21.1	11.02	19.6	20.01	29.05	12.4	15.6	-97.5	1.5	24.2	6.8 \pm 4.5
Rate simplification		-17.2	-53.4	8.9	-35.4	-2.9	2.7	-14.7	-22.9	-5.7	-71.04	-21.1	6.7	18.6 \pm 9.7
Futile cycle removal	Eq. (8)	2.5	-0.9	15.6	-1.6	-0.01	-2.3	-0.6	-1.5	-5.9	-43.23	-3.8	23.6	12.7 \pm 14.7
	Eq. (9)	0.7	-56.7	-5.6	-37.1	-9.02	-10.5	-6.7	-35.5	-12.2	-70.7	-24.3	24.1	11.5 \pm 9.9
Intermediate reduced model		-32.7	-18.6	-3.7	-24.5	-11.8	-24.04	-16.5	-20.3	-18.8	-43.1	-21.4	30.5	22.5 \pm 8.4
Final reduced model		-32.5	-19.1	-4.3	-25.1	-12.7	-1.01	-16.4	-20.4	-18.8	-43.3	-18.5	43.3	22.5 \pm 8.5

446 6.2. Strategy 2: Structural simplification of the model

447 Structural simplification methods are another way to reduce the complex-
 448 ity of dynamic systems by improving the generality of the model and the
 449 numerical integration of the ordinary differential equations.

450 Firstly, we tried to remove the temporal and the phenotype effects in the
 451 enzyme activities, v_2, \dots, v_7 (v_1 has neither phenotype nor temporal effects).
 452 The results of this simplification are shown in Table 2. The elimination of
 453 the phenotype effect for v_3, v_6 and v_7 resulted in a decrease of the AIC_C
 454 value for nine genotypes, neutral for one genotype, and was thus selected for
 455 the final reduction. The elimination of the temporal effect for v_2, v_4, v_5, v_7
 456 was also advantageous on the corrected AIC results for all ten genotypes.
 457 Nevertheless, when we tried to eliminate the temporal effect of v_7 , the result-
 458 ing Δ_{AIC_C} was positive for most genotypes. This is in line with the results
 459 of multi-variate sensitivity analysis according to which v_2, \dots, v_6 have a low
 460 sensitivity on the four outputs of the model, whereas v_7 has a non-negligible
 461 effect on the dynamics of glucose concentration. According to these results,
 462 the elimination of the temporal effect was validated only for v_2, v_4, v_5, v_6 .
 463 In support of this choice, the test with the virtual genotypes shows that the
 464 expected error between the reduced and the original model is small (Table
 465 2).

466 In the second phase, we tested the possibility of simplifying the enzymatic
 467 reaction rates (Eq.(7)). For each reaction in the model, Fig. 5 compares the
 468 order of magnitude of the substrate $x(t)$ to the corresponding affinity K_m .
 469 The boxplots show that (**Case 2**, see section 3.2.2) simplification strategy
 470 can be applied only for the reaction rates u_5 and u_7 . Therefore, their reaction
 471 rates can be written as $u = V_{max}$. All other flows verify the (**Case 1**, see
 472 section 3.2.2) and can therefore be expressed as $u = \frac{V_{max}}{K_m}x(t)$. The rates
 473 simplification improves the corrected AIC for eight genotypes and yields a
 474 substantial gain in the calibration time. The expected error over the virtual
 475 progeny is higher than in the previous reduction steps, but still in the range
 476 of accuracy of the original model [1]. According to these observations, the
 477 enzymatic reaction rates simplification strategy was validated.

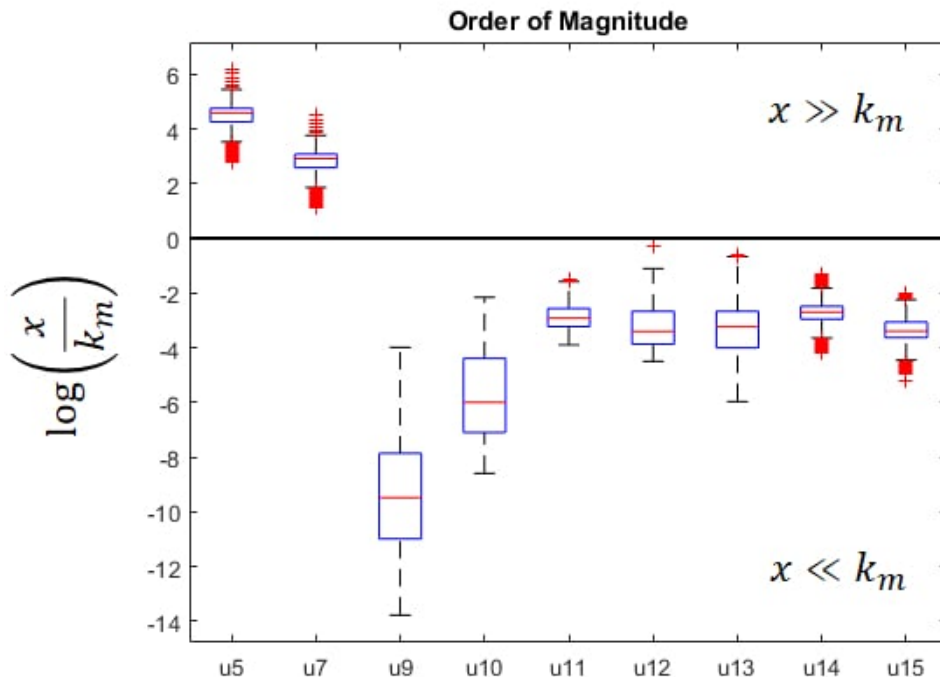


Figure 5: Differences in order of magnitude between enzyme affinity (K_m) and substrate concentration (x) calculated over the whole dynamics and the training set for each reaction rate u_i , $i \in \{5, 7, 9 \dots 15\}$.

478 Eventually, futile cycles were detected to reduce the full system. In the
 479 original model, glucose, and fructose sugars can be transported to the vacuole
 480 via two possible mechanisms: an active, unidirectional transport (u_5, u_7)
 481 and passive reversible transport (u_6, u_8). Simulations showed that, whenever
 482 the genotype, the net flux mostly pointed in the direction of an export for
 483 both fructose and glucose from the vacuole to the cytosol [1]. However,
 484 futile cycles occurred due to the presence of the active transport mechanism,
 485 that continually brings glucose and fructose back into the vacuole. Indeed,
 486 u_5 and u_6 (respectively u_7 and u_8) had the same evolution over the whole
 487 dynamics for all ten genotypes (Fig. 6): the active and passive transport ran
 488 simultaneously in two opposite directions.

489 According to our strategy (section 3.2.3), we tried to remove futile cycles
 490 by replacing reactions (u_5, u_6) (respectively (u_7, u_8)) with an effective reac-
 491 tion rate of the form $p_{10} x_9 - p_{11} x_4$ (respectively $p_9 x_8 - p_{12} x_3$) preserving the

492 net export flux from vacuole to the cytosol. We compared the performance of
493 the reduced model with respect to the original one (Table 2). The corrected
494 AIC values were generally slightly negative, with the exception of genotypes
495 E_1 and E_{43} , suggesting an overall improvement of the model structure. No-
496 tice that the present strategy did not reduce the total parameters number
497 but decreased model complexity and improved the calibration time.
498 As a further simplification, we then tried to use a special case of the above
499 mentioned reaction rate with $p_{10} = p_{11}$ (respectively $p_9 = p_{12}$). This time,
500 the simplification was fully validated by the corrected AIC on all genotypes
501 (Table 2, Eq.(9)). The expected error over the virtual genotypes was esti-
502 mated to 13% and the calibration time was lowered by 24% with respect to
503 the original model, thanks to structural simplification and the reduction of
504 the number of parameters to be estimated. Accordingly to these results, the
505 simplification by Eq.(9) was validated.

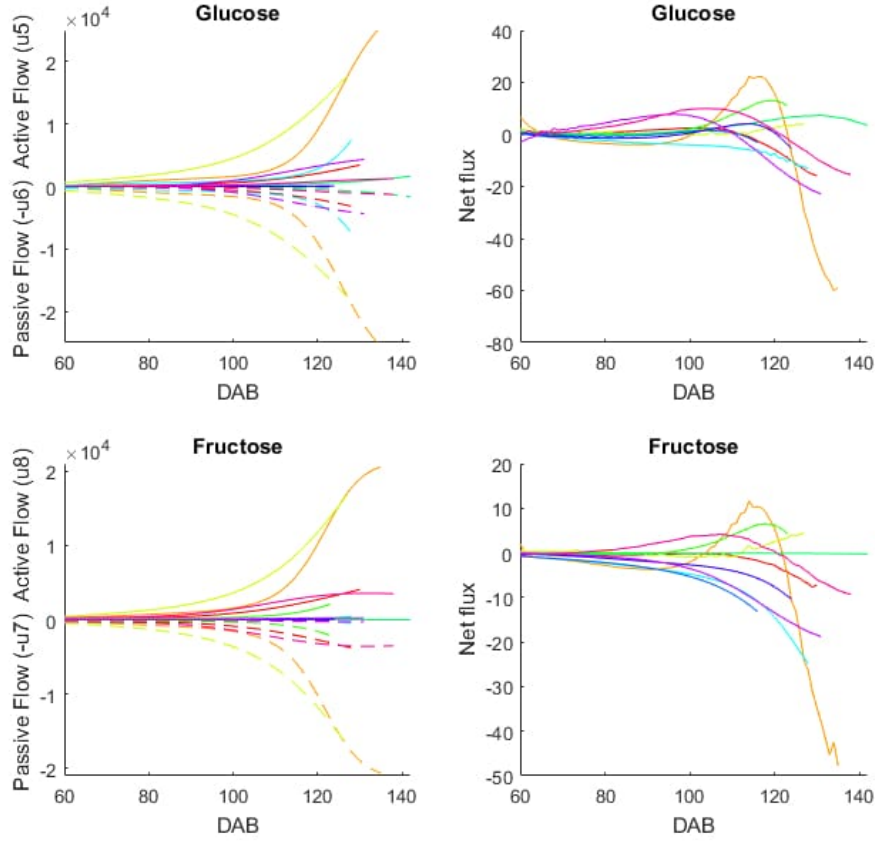


Figure 6: Evolution of the active flux (solid lines) and passive transport (dashed lines) for glucose (respectively fructose) and net flux during fruit development (DAB, day after bloom) for the ten genotypes of the training set (different colors).

506 *6.3. Strategy 3: Time-scale analysis and QSSA*

507 Results from the reduction strategies 1 and 2 were combined into an inter-
 508 mediate reduced model. This model had only 9 parameters to be estimated,
 509 linear flows and only one temporal enzymatic capacity, common to all geno-
 510 types. Improvement in AIC_C with respect to the original model confirmed
 511 a strong benefice for all ten genotypes (Table 2). The expected error over
 512 a large progeny was estimated around 20%, close to the performance of the
 513 original model.

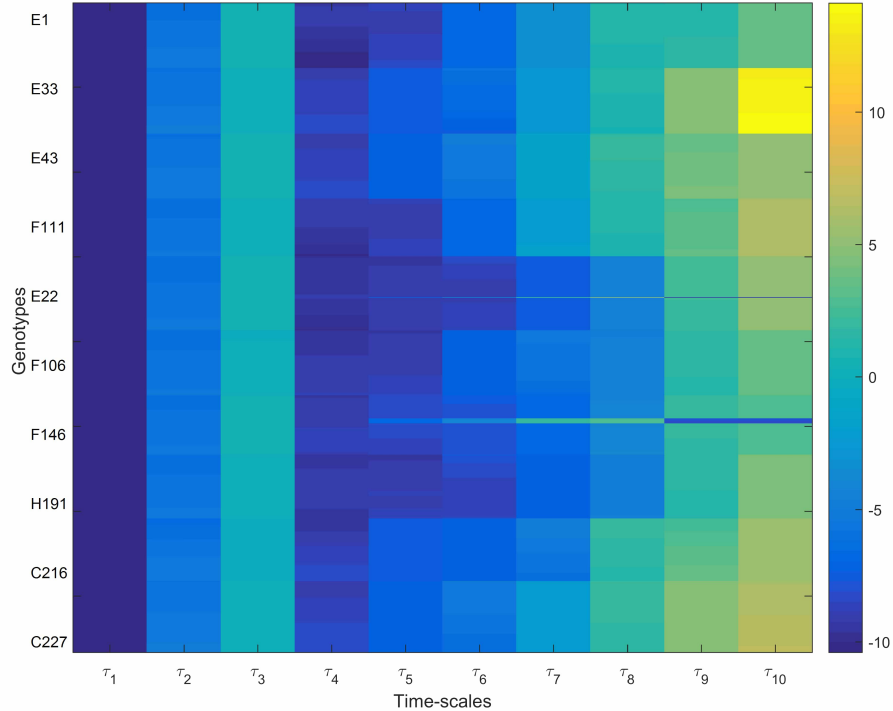


Figure 7: Order of magnitude of time scales τ_i along fruit development (DAB, days after bloom) for the 10 genotypes of the training set.

514 On the basis of this intermediate reduced model, time scale analysis was
 515 performed to detect the possible presence of fast modes in the system. The
 516 analysis of the Jacobian matrix, indeed, confirmed the presence of different
 517 modes, with typical time scales spanning a few seconds up to days, for all
 518 tested genotypes (Fig. 7).

519 A fast transient dynamics, followed by a slow one, was observable in the
 520 numerical simulations of the original and intermediate reduced models for
 521 the hexose phosphates concentration (variable x_5 , see supplemental infor-
 522 mation, Fig. D.4). In addition, following the method proposed in [33, 17],
 523 we analyzed the predicted concentration of sugars in both intracellular com-
 524 partments, for all genotypes. The concentration of the hexose phosphate
 525 (x_5) was systematically lower than the concentrations of the other variables
 526 in the system, as expected for the fast components of the system (Fig. 8).

527 Accordingly, x_5 was assumed to be at quasi-steady-state and its equation was
 528 replaced by an algebraic function of the slow variables.

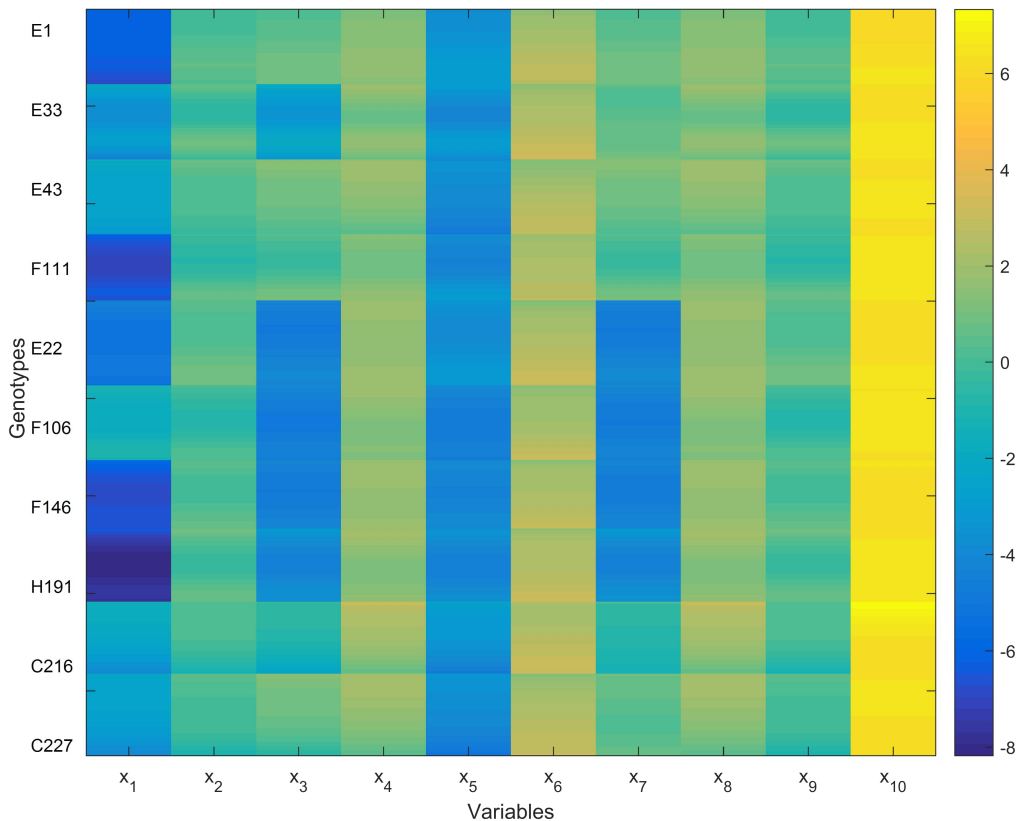


Figure 8: Order of magnitude of the predicted sugars concentrations (mg gFW^{-1}) in the cytosol (x_1 : Sucrose, x_2 : Sorbitol, x_3 : Fructose, x_4 : Glucose, x_5 : Hexose Phosphate, x_{10} : Other compounds) and vacuole (x_6 : Sucrose, x_7 : Fructose, x_8 : Glucose, x_9 : Sorbitol), along fruit development (DAB, days after bloom) for the ten genotypes of the training set.

529 We compared the intermediate reduced model with its QSS approxima-
 530 tion by calculating J_i (Eq.(20)) as explained previously. J_i was very low, less
 531 than 1%, over the whole dynamics for all variables (Fig. 9). This result was
 532 validated also on the virtual genotypes simulated with QQS approximation
 533 (see the supplemental information Fig. D.5). In addition the QSS assump-
 534 tion, further increased the performance of the model, leading to a gain in the
 535 calibration time of 40% with respect to the original model.

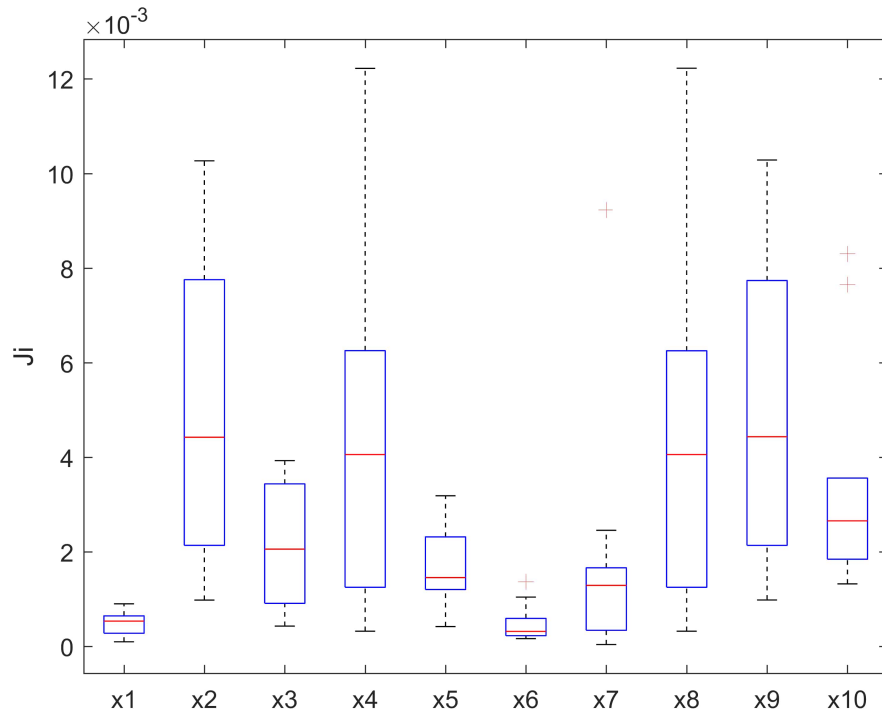


Figure 9: Normalized Root Mean Square Errors $J_i, i \in \{1, \dots, 10\}$ between the intermediate and reduced models after application of the QSSA to x_5 . The boxplot shows the variability of J_i over the training set

536 *6.4. Evaluation of the reduced model*

537 The validity of the reduced model was verified on some new genotypes of
 538 the inter-specific peach progeny, for which few data were available.

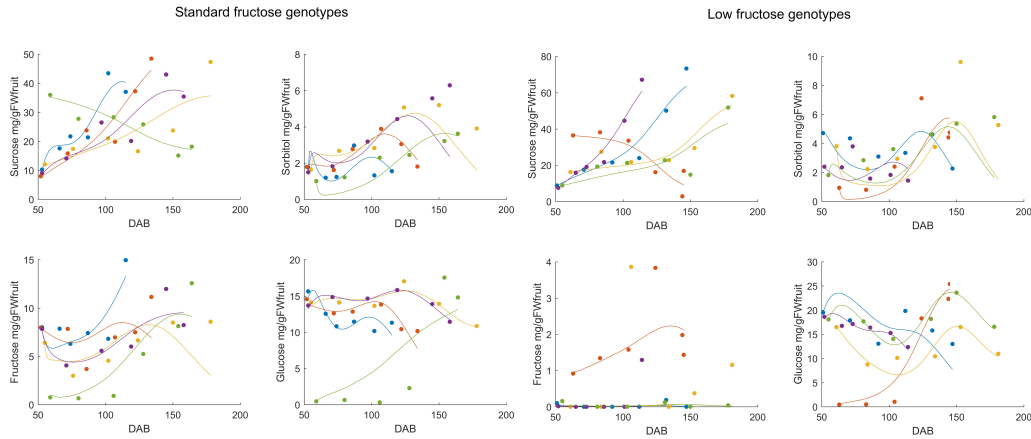


Figure 10: Evolution of the concentration ($mgFW^{-1}$) of sugars during fruit development (DAB, days after bloom) for ten representative genotypes of the validation set with standard (left) and low fructose (right) phenotypes. Dots represent experimental data and lines are model simulations.

539 The reduced model was then calibrated on the dynamics of sugar concen-
 540 tration of these selected genotypes, as described in section 5.2. The results
 541 presented in (Fig. 10) showed a satisfactory agreement between model and
 542 data, all over fruit development, for most genotypes. The average *NRMSE*
 543 (Table D.6) ranged from 10% to 30% for the main sugars, in good agreement
 544 with estimations over the virtual progeny. These results confirmed that the
 545 reduced model offered a quality of prediction close to the original one with
 546 fewer parameters to be estimated and shorter integration time.

547 From a biological perspective, an important prediction of the model de-
 548 veloped by Desnoues et al. [1] was that a difference in fructokinase affinity
 549 could be at the origin of the phenotypic difference observed between standard
 550 and low fructose genotypes.

551 We checked if the estimations obtained with the reduced model still sup-
 552 ported this hypothesis. Fig. 11 shows a significant difference of estimated
 553 fructokinase affinity between the two phenotypic groups, in agreement with
 554 the original model based on the Student t-test (p-value $< 2.0187e^{-9}$)

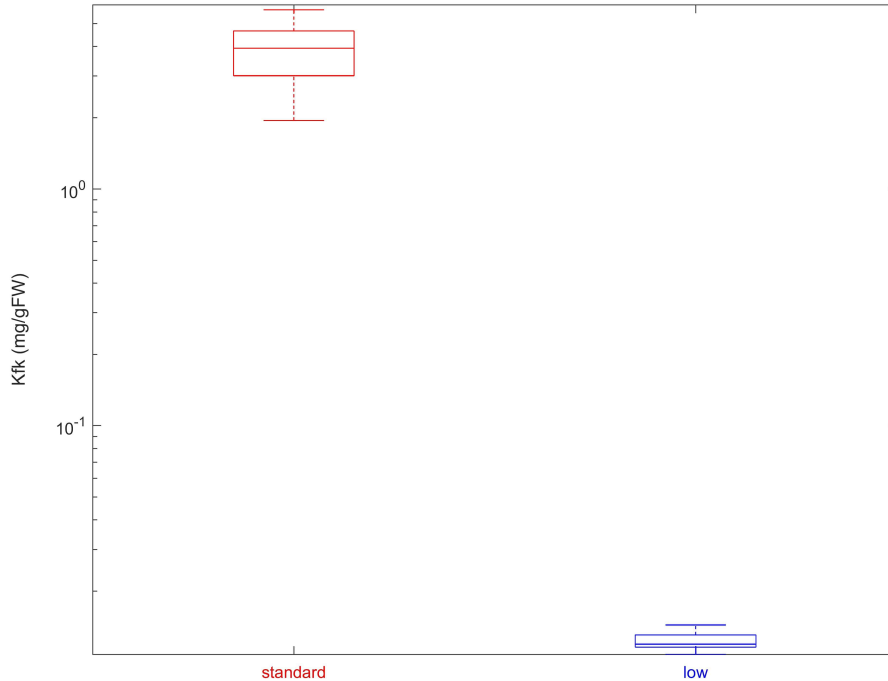


Figure 11: Difference in the estimated fructokinase affinity between standard and low fructose phenotypes, for forty genotypes (training and validation sets). The difference is significant with a p-value $< 2.0187e^{-9}$.

555 7. Discussion

556 Models of metabolic systems are usually very complex. Complexity stems
 557 from the number of components and the high degree of non-linearity included
 558 in both the network structure and the individual reaction rates. As a conse-
 559 quence, metabolic models usually suffer from numerical and identifiability is-
 560 sues that seriously hamper their application in the context of genetic studies,
 561 especially when they have to be calibrated for hundreds of genotypes. In this
 562 paper, we present a reduction scheme that explicitly accounts for genomic
 563 diversity. Our approach is based on the systematic evaluation of different
 564 reduction methods, that, if successful, are then combined together to yield
 565 the final reduced model. When applied to the model of sugar metabolism
 566 developed by Desnoues et al. [1] our approach led to a reduced model that

567 could be efficiently calibrated on a large diversity of genotypes, for which few
568 data are available. The reduced model showed comparable predictions and
569 biological interpretation as the original model, with only a limited number of
570 estimated parameters. Indeed, calibration time was reduced by 40%, a con-
571 siderable improvement when considering that the calibration of the original
572 model could span up to 30 hours for a single genotype. Moreover, mitigation
573 of model non-linearities can help limiting numerical issues and increase the
574 reliability of estimated parameters, an important aspect in the context of
575 genetic studies, where large genetic populations have to be calibrated.

576 The proposed reduction scheme is especially suitable for dynamical mod-
577 els of metabolic and biochemical networks, in which a large number of chem-
578 ical reactions interact with similar non-linear kinetics. In these systems,
579 indeed, the connectivity properties of the network usually prime over the
580 precise description of the individual rate laws [40]. The presence of satu-
581 rating kinetic functions (like the classical Michaelis-Menten), in particular,
582 allows the simplification of the rate function depending on the substrate
583 range whereas the presence of redundant or opposite reactions opens the
584 way to structural simplification of the system. The extension of these reduc-
585 tion steps to another kind of models is less straightforward. Crop models for
586 instance can involve a large variety of process kinetics, one for each described
587 physiological process. The complexity of the cellular network is replaced by
588 the interaction of a comparatively small number of processes but described
589 by complicated, ad-hoc kinetic functions that can involve several model com-
590 ponents as well as external environmental variables (temperature, humidity,
591 light). The simplification of individual rate laws is still possible but it in-
592 volves case-by-case study.

593 Although the application of specific reduction methods is tailored to
594 model structure, the proposed evaluation strategy is pretty generic and easily
595 adaptable to a large range of biological models. The main objective of this
596 work was to provide a method to build a reduced model that is adapted to
597 the application to a large panel of genotypes. In this sense, we do not look
598 for the best model for a given genotype but rather for the best *compromise*
599 in terms of accuracy and efficiency over a large genetic diversity. The ques-
600 tion recalls the one of "model validation domain" i.e. the ability for a given
601 model to describe data obtained in conditions different from those in which
602 the model itself was calibrated [41]. Here it is about selecting for a reduced
603 model having a large validation domain and able to cope with changes in
604 model's inputs, parameter values, and initial conditions.

605 For this aim, we proposed a criterium based on the simulation of a large
606 number of virtual genotypes and the systematic comparison of the expected
607 distance between the original and the reduced models. Virtual genotypes are
608 built based on the variability observed in a sub-sample of the population,
609 plus a basal variability, expressed as a random effect, to limit the bias due to
610 the choice of the initial sample and to assure a minimal diversity across the
611 virtual population. A few remarks are needed. First, the above method tests
612 the reliability of the reduction, assuming that the original model is valid. In
613 this sense, the amplitude of the basal random effect should be subject to an
614 expert knowledge so to avoid biologically unreasonable situations, that fall
615 outside the conditions of applicability of the model. Second, it is worth to
616 notice that, given the virtual nature of our comparison, the reduced model
617 is parameterized using parameter values that are directly derived from the
618 parameters of the original model, to which it is compared. In this sense, the
619 'expected NRMSE error' of the reduced model represents an upper bound of
620 its actual accuracy over an experimental dataset, as parameter re-calibration
621 can significantly improve the performances of the reduced model on real ge-
622 netic populations.

623 Ultimately, the existence of a reduced model will considerably speed up
624 the integration of genetic control into ecophysiological models. Currently,
625 most genetic-improved ecophysiological models make use of Quantitative
626 Trait loci (QTL) to describe the genetic architecture of specific model param-
627 eters. Basically, each parameter has a specific distribution in the population
628 of genotypes and QTL analyses can be performed for each parameter to deci-
629 pher the architecture of its genetic control (QTL number and effects, linkage).
630 However, a major drawback of this approach is the difficulty in the calibra-
631 tion of the models for a large number of genotypes (due to a large number
632 of parameters along with restricted number of observations) [6, 7]. Indeed,
633 the statistical power of QTL analyses strongly depends on the size of the
634 population and on the QTL effects i.e. their contribution to the variation of
635 the trait they are associated with [42]. So, in order to be of interest, genetic
636 parameters have to vary among genotypes and be quantifiable with relevant
637 accuracy either experimentally or through numerical optimization.

638 In this perspective, a reduced model with a simpler structure will allow for
639 a better exploration of the parameter space and a more accurate estimation
640 of parameter values. Moreover, the improved calibration time opens the
641 possibility of exploring larger genetic populations so to get more robust QTLs
642 estimation. Finally, it will allow to do simulations over a large number of

643 environmental conditions and/or climatic scenarios.

644 This is an important step towards dealing with complex Genotype x Envi-
645 ronment x Management interactions issues expected in the near future. The
646 development of reliable gene-to-phenotype models will be an important lever
647 to optimize farming in the future climatic conditions.

648 **Acknowledgments**

649 HK was founded by a scholarship of the Lebanese government. We
650 would like to thank V. Signoret for her help in maintaining the inter-specific
651 peach progeny. We are grateful to the IE-EMMAH UMR1114 and IE-GAFL
652 UR1052 teams for taking care of the experimental orchard, and to Dr Olivier
653 Martin for his help in statistics.

654 **References**

- 655 [1] E. Desnoues, M. Génard, B. Quilot-Turion, V. Baldazzi, A kinetic model
656 of sugar metabolism in peach fruit reveals a functional hypothesis of a
657 markedly low fructose-to-glucose ratio phenotype, *The Plant Journal*
658 94 (4) (2018) 685–698.
- 659 [2] J. W. White, G. Hoogenboom, Gene-Based Approaches to Crop Simu-
660 lation, *Agron. J.* 95 (1) (2003) 52–64.
- 661 [3] V. Letort, P.-H. Cournède, A. Mathieu, P. De Reffye, T. Constant,
662 Parametric identification of a functional–structural tree growth model
663 and application to beech trees (*fagus sylvatica*), *Functional plant biology*
664 35 (10) (2008) 951–963.
- 665 [4] F. Tardieu, Virtual plants: modelling as a tool for the genomics of tol-
666 erance to water deficit, *Trends in plant Science* 8 (1) (2003) 9–14.
- 667 [5] B. Quilot-Turion, M. Génard, P. Valsesia, M.-M. Memmah, Optimiza-
668 tion of allelic combinations controlling parameters of a peach quality
669 model, *Frontiers in plant science* 7 (2016) 1873.
- 670 [6] P. Martre, B. Quilot-Turion, D. Luquet, M.-M. O.-S. Memmah,
671 K. Chenu, P. Debaeke, Chapter 14 - model-assisted phenotyping and
672 ideotype design, in: V. O. Sadras, D. F. Calderini (Eds.), *Crop Phys-*
673 *iology (Second Edition)*, second edition Edition, Academic Press, San
674 Diego, 2015, pp. 349 – 373.

- 675 [7] N. Bertin, P. Martre, M. Genard, B. Quilot-Turion, C. Salon, Under
676 what circumstances can process-based simulation models link genotype
677 to phenotype for complex traits? case-study of fruit and grain quality
678 traits, *Journal of experimental botany* 61 (2010) 955–67.
- 679 [8] N. Curien, F. Moreau, The music industry in the digital era: Toward
680 new contracts, *Journal of Media Economics* 22 (2009) 102–113.
- 681 [9] T. Nägele, W. Weckwerth, Mathematical modeling reveals that
682 metabolic feedback regulation of *snrk1* and hexokinase is sufficient to
683 control sugar homeostasis from energy depletion to full recovery, *Frontiers in plant science* 5 (2014) 365.
684
- 685 [10] B. P. Beauvoit, S. Colombié, A. Monier, M. . H. Andrieu, B. Biais,
686 C. Bénard, C. Chéniclet, M. Dieuaide-Noubhani, C. Nazaret, J. . P.
687 Mazat, Y. Gibon, Model-assisted analysis of sugar metabolism through-
688 out tomato fruit development reveals enzyme and carrier properties in
689 relation to vacuole expansion, *Plant Cell* 26 (2014).
- 690 [11] M. S. Okino, M. L. Mavrovouniotis, Simplification of mathematical mod-
691 els of chemical reaction systems, *Chemical reviews* 98 (2) (1998) 391–
692 408.
- 693 [12] A. N. Gorban, N. K. Kazantzis, I. G. Kevrekidis, H. C. Öttinger,
694 C. Theodoropoulos, *Model reduction and coarse-graining approaches for
695 multiscale phenomena*, Springer, 2006.
- 696 [13] T. J. Snowden, P. H. van der Graaf, M. J. Tindall, Methods of model
697 reduction for large-scale biological systems: a survey of current methods
698 and trends, *Bulletin of mathematical biology* 79 (7) (2017) 1449–1486.
- 699 [14] J. Wei, J. C. W. Kuo, Lumping analysis in monomolecular reaction sys-
700 tems. analysis of the exactly lumpable system, *Industrial & Engineering
701 Chemistry Fundamentals* 8 (1) (1969) 114–123.
- 702 [15] M. Sunnåker, G. Cedersund, M. Jirstrand, A method for zooming of
703 nonlinear models of biochemical systems, *BMC systems biology* 5 (1)
704 (2011) 140.

- 705 [16] M. Schauer, R. Heinrich, Quasi-steady-state approximation in the math-
706 ematical modeling of biochemical reaction networks, *Mathematical bio-*
707 *sciences* 65 (2) (1983) 155–170.
- 708 [17] R. Heinrich, S. Schuster, *The regulation of cellular systems*, Chapman
709 and Hall, 1996.
- 710 [18] T. Turányi, Sensitivity analysis of complex kinetic systems. tools and
711 applications, *Journal of mathematical chemistry* 5 (3) (1990) 203–248.
- 712 [19] J. Cariboni, D. Gatelli, R. Liska, A. Saltelli, The role of sensitivity
713 analysis in ecological modelling, *Ecological Modelling* 203 (1) (2007)
714 167 – 182.
- 715 [20] E. Vanuytrecht, D. Raes, P. Willems, Global sensitivity analysis of yield
716 output from the water productivity model, *Environmental Modelling*
717 *Software* 51 (2014) 323 – 332.
- 718 [21] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli,
719 M. Saisana, S. Tarantola, *Global sensitivity analysis: the primer*, John
720 Wiley & Sons, 2008.
- 721 [22] P. Holme, M. Huss, H. Jeong, Subnetwork hierarchies of biochemical
722 pathways, *Bioinformatics* 19 (4) (2003) 532–538.
- 723 [23] J. Anderson, Y.-C. Chang, A. Papachristodoulou, Model decomposition
724 and reduction tools for large-scale networks in systems biology, *Auto-*
725 *matica* 47 (6) (2011) 1165 – 1174, special Issue on Systems Biology.
- 726 [24] X. Sun, M. Medvedovic, Model reduction and parameter estimation of
727 non-linear dynamical biochemical reaction networks, *IET Systems Biol-*
728 *ogy* 10 (2016) 10–16(6).
- 729 [25] F.-S. Wang, C.-L. Ko, E. O. Voit, Kinetic modeling using S-systems and
730 lin-log approaches, *Biochem. Eng. J.* 33 (3) (2007) 238–247.
- 731 [26] I. E. Nikerel, W. A. van Winden, P. J. Verheijen, J. J. Heijnen, Model
732 reduction and a priori kinetic parameter identifiability analysis using
733 metabolome time series for metabolic reaction networks with linlog ki-
734 netics, *Metab. Eng.* 11 (1) (2009) 20–30.

- 735 [27] H. Schmidt, M. F. Madsen, S. Danø, G. Cedersund, Complexity re-
736 duction of biochemical rate expressions, *Bioinformatics* 24 (6) (2008)
737 848–854.
- 738 [28] W. Liebermeister, U. Baur, E. Klipp, Biochemical network models sim-
739 plified by balanced truncation, *The FEBS Journal* 272 (16) (2005) 4034–
740 4043.
- 741 [29] M. Apri, M. de Gee, J. Molenaar, Complexity reduction preserving dy-
742 namical behavior of biochemical networks, *Journal of theoretical biology*
743 304 (2012) 16–26.
- 744 [30] M. Lamboni, D. Makowski, S. Lehuger, B. Gabrielle, H. Monod, Multi-
745 variate global sensitivity analysis for dynamic crop models, *Field Crops*
746 *Research* 113 (3) (2009) 312–320.
- 747 [31] R. Schürer, W. C. Schmid, Mint: A database for optimal net parameters,
748 in: H. Niederreiter, D. Talay (Eds.), *Monte Carlo and Quasi-Monte*
749 *Carlo Methods 2004*, Springer Berlin Heidelberg, Berlin, Heidelberg,
750 2006, pp. 457–469.
- 751 [32] E. Desnoues, Y. Gibon, V. Baldazzi, V. Signoret, M. Génard, B. Quilot-
752 Turion, Profiling sugar metabolism during fruit development in a peach
753 progeny with different fructose-to-glucose ratios, *BMC Plant Biology*
754 14 (1) (2014) 336.
- 755 [33] C. López Zazueta, O. Bernard, J.-L. Gouzé, Analytical reduction of
756 nonlinear metabolic networks accounting for dynamics in enzymatic re-
757 actions, *Complexity* 2018 (2018).
- 758 [34] B. Quilot, M. Génard, J. Kervella, F. Lescouret, Analysis of genotypic
759 variation in fruit flesh total sugar content via an ecophysiological model
760 applied to peach, *Theoretical and Applied Genetics* 109 (2) (2004) 440–
761 449.
- 762 [35] J. M. Chambers, T. J. Hastie, et al., *Statistical models in S*, Vol. 251,
763 Wadsworth & Brooks/Cole Advanced Books & Software Pacific Grove,
764 CA, 1992.

- 765 [36] C. Barrasso, M.-M. Memah, M. Génard, B. Quilot-Turion, Model-based
766 qtl detection is sensitive to slight modifications in model formulation,
767 PloS one 14 (10) (2019).
- 768 [37] M. Hosea, L. Shampine, Analysis and implementation of tr-bdf2, Ap-
769 plied Numerical Mathematics 20 (1-2) (1996) 21–37.
- 770 [38] D. E. Goldberg, Genetic algorithms in search, Optimization, and Ma-
771 chineLearning (1989).
- 772 [39] K. P. Burnham, D. R. Anderson, Model selection and multimodel infer-
773 ence : a practical information-theoretic approach, Springer, New York,
774 2002.
- 775 [40] A.-L. Barabási, Z. N. Z. N. Oltvai, Network biology: understanding the
776 cell’s functional organization, Nat. Rev. Genet. 5 (2) (2004) 101–113.
- 777 [41] F. Mairet, O. Bernard, Twelve quick tips for designing sound dynamical
778 models for bioprocesses, PLoS Comput Biol in press (2019) 1–15.
- 779 [42] T. Mackay, E. Stone, J. Ayroles, The genetics of quantitative traits:
780 challenges and prospects., Nat. Rev. Genet. 10 (8) (2009) 565–77.
- 781 [43] M. Génard, N. Bertin, H. Gautier, F. Lescourret, B. Quilot, Virtual
782 profiling: a new way to analyse phenotypes, The Plant Journal 62 (2)
783 (2010) 344–355.
- 784 [44] K. Shiratake, Y. Kanayama, S. Yamaki, Characterization of hexose
785 transporter for facilitated diffusion of the tonoplast vesicles from pear
786 fruit, Plant and cell physiology 38 (8) (1997) 910–916.

787 **Appendices**

788 **Appendix A. Model description**

789 *Appendix A.1. Model equations*

790 The original model [1] was written in terms of species *carbon* quantities
791 $C(t)$. Here, we decided to rewrite the system as a function of species concen-
792 tration $x_i(t)$, for a better readability. The quantity of carbon as a sugar i (C_i)
793 depends on the concentration of i (x_i) according to the following equation:

$$C_i = \sigma_i x_i V_j \tag{A.1}$$

794 where σ_i is the carbon concentration of sugar i and V_j is the volume of the
795 intracellular compartment (cytosol or vacuole) in which species i is located.
796 The carbon content σ_i for the different sugar molecules is reported in Table
797 A.1. Table A.2 specifies variable location within the cell's compartments.
798 Differentiation of Equation (Eq. (A.1)) leads to:

$$\frac{dx_i}{dt} = \frac{1}{\sigma_i V_j} \frac{dC_i}{dt} - \frac{1}{V_j} x_i \frac{dV_j}{dt} \quad (\text{A.2})$$

799 Accordingly, for variables $1, \dots, 5, 10$, $C_i = \sigma_i x_i V_1$ whereas $C_i = \sigma_i x_i V_2$ for
800 $i \in [6, 9]$. For simplicity, we assume $\frac{V_1}{V_2} = \alpha$. This leads to $\mu(t) = \frac{1}{V_1} \frac{dV_1}{dt} =$
801 $\frac{1}{V_2} \frac{dV_2}{dt}$.
802

Table A.1: Carbon content of each sugar

σ	Sugar	Value
σ_1, σ_6	Sucrose	0.421
σ_3, σ_8	Fructose	0.4
σ_2, σ_7	Sorbitol	0.39
σ_4, σ_9	Glucose	0.4
σ_5	Hexose phosphate	0.27
σ_{10}	Other compounds	0.44

Table A.2: Model variables and location

S_1	Sucrose	Cytosol
S_2	Sorbitol	Cytosol
S_3	Fructose	Cytosol
S_4	Glucose	Cytosol
S_5	Hexose phosphate	Cytosol
S_6	Sucrose	Vacuole
S_7	Sorbitol	Vacuole
S_8	Fructose	Vacuole
S_9	Glucose	Vacuole
S_{10}	Other compounds	Cytosol

Table A.3: Reaction rates of the original and reduced models

Equations	Original Model	Reduced Model
Input flows	$I(t) = \sigma_f \frac{dDW}{dt} + R(t) = (\sigma_f + q_g) \frac{dDW}{dt} + q_m DW Q_{10}^{\frac{(T-20)}{10}}$ $R(t) = q_m DW Q_{10}^{\frac{(T-20)}{10}} + q_g \frac{dDW}{dt}$ $DW = DW(t_0) + w_1(1 - e^{-w_2 t}) + \frac{w_3}{1 + e^{-w_4(t-t_0)}}$ $u_1(I) = \frac{1}{\sigma_1 V_1} \lambda \lambda_{suc}(t) I(t)$ $\lambda_{suc}(t) = \frac{p_1 t}{t_{max}}$ where t_{max} corresponds to the maturation time $u_2(I) = \frac{1}{\sigma_2 V_1} (1 - \lambda) I(t)$ $u_3(I) = \frac{1}{\sigma_3 V_1} \frac{\lambda}{2} (1 - \lambda_{suc}(t)) I(t)$	
Metabolism	$u_9(v_2, x_1) = \frac{v_2(t)}{p_5 + x_1} x_1(t)$ $u_{10}(x_1) = \frac{v_3}{p_{21} + x_1} x_1(t)$ $u_{11}(v_4, x_2) = \frac{v_4(t)}{p_{22} + x_2} x_2(t)$ $u_{12}(v_5, x_2) = \frac{v_5(t)}{p_{13} + x_2} x_2(t)$ $u_{13}(x_6, x_8, x_9) = \frac{v_1}{(1 + \frac{x_8 + x_9}{p_2}) p_{23} + x_6} x_6(t)$ $u_{14}(v_6, x_3) = \frac{v_6(t)}{p_3 + x_3} x_3(t)$ $u_{15}(v_7, x_4) = \frac{v_7(t)}{p_4 + x_4} x_4(t)$ $u_{16}(x_5) = p_7 x_5(t)$ $u_{17}(x_5) = p_6 x_5(t)$ $u_{18}(R) = R(t)$	$u_9(x_1) = \frac{v_2}{p_5} x_1(t) = r_1 x_1(t)$ $u_{10}(x_1) = \frac{v_3}{p_{21}} x_1(t) = r_2 x_1(t)$ $u_{11}(x_2) = \frac{v_4}{p_{22}} x_2 = r_3 x_2(t)$ $u_{12}(x_2) = \frac{v_5}{p_{13}} x_2(t) = r_4 x_2(t)$ $u_{13}(x_6) = r_5 x_6(t)$ $u_{14}(x_3) = \frac{v_6}{p_3} x_3(t) = r_6 x_3(t)$ $u_{15}(v_7, x_4) = \frac{v_7(t)}{p_4} x_4(t)$ $u_{16}(x_5) = p_7 x_5(t)$ $u_{17}(x_5) = p_6 x_5(t)$ $u_{18}(R) = R(t)$
Transport processes	$u_4(S, x_1) = p_8 x_1(t) S(t)$ $u_5(S, x_3, x_4) = \frac{p_{11}}{p_{19} + x_3 + x_4} x_4(t) S(t)$ $u_6(S, x_4, x_9) = (x_9 - x_4) p_{10} S(t)$ $u_7(S, x_3, x_4) = \frac{p_{12}}{p_{20} + x_3 + x_4} x_3(t) S(t)$ $u_8(S, x_3, x_8) = (x_8 - x_3) p_9 S(t)$ $u_{19}(S, x_2, x_7) = p_{14} (x_7 - x_2) S(t)$	$u_4(S, x_1) = p_8 x_1(t) S(t)$ $u_5 = 0$ $u_6(S, x_4, x_9) = (x_9 - x_4) p_{10} S(t)$ $u_7 = 0$ $u_8(S, x_3, x_8) = (x_8 - x_3) p_9 S(t)$ $u_{19}(S, x_2, x_7) = p_{14} (x_7 - x_2) S(t)$

804 of 19 reactions and one input function $I(t)$. The latter described the carbon
805 supply from the mother plant to the fruit and it was estimated as the sum
806 of the carbon used for fruit dry mass (DW) increase and the carbon lost by
807 respiration ($R(t)$). Two parameters λ and λ_{suc} described the fraction of the
808 input flow that is converted into the different forms of sugars. Fruit respira-
809 tion was computed following the growth-maintenance paradigm, as described
810 in [1].

811 Reaction rates are reported in Table A.3. Enzymatic reactions were generally
812 described using an irreversible Michaelis-Menten kinetics, with experimentally-
813 measured capacities $v_i(t)$. Transport processes between cytosol and vacuole
814 were assumed proportional to the vacuole surface (hypothesis of constant
815 density of transporters) computed from vacuole fresh mass (proxy of the vol-
816 ume) supposing the vacuole as a sphere of surface $S(t) = (4\pi)^{\frac{1}{3}}(V_2)^{\frac{2}{3}}$ (see [1]
817 for more information). Both active and passive transport mechanisms were
818 considered for fructose and glucose.

819 Model equations are reported in Table A.4, for both the original and the
820 reduced model.

Table A.4: System of original and reduced models

System of original model	System of reduced model
$\frac{dx_1}{dt} = u_1 + \frac{\sigma_5}{\sigma_1}u_{16} - u_{10} - u_4 - \mu(t)x_1$	
$\frac{dx_2}{dt} = u_2 - u_{11} - u_{12} + \frac{1}{\sigma_2 V_1}u_{19} - \mu(t)x_2$	
$\frac{dx_3}{dt} = u_3 + \frac{1}{\sigma_3 V_1}u_8 + \frac{1}{2} \frac{\sigma_1}{\sigma_3}u_9 + \frac{1}{2} \frac{\sigma_1}{\sigma_3}u_{10} + \frac{\sigma_2}{\sigma_3}u_{11} - u_7 - u_{14} - \mu(t)x_3$	$\frac{dx_3}{dt} = u_3 + \frac{1}{\sigma_3 V_1}u_8 + \frac{1}{2} \frac{\sigma_1}{\sigma_3}u_9 + \frac{1}{2} \frac{\sigma_1}{\sigma_3}u_{10} + \frac{\sigma_2}{\sigma_3}u_{11} - u_{14} - \mu(t)x_3$
$\frac{dx_4}{dt} = u_3 + \frac{1}{\sigma_4 V_1}u_6 + \frac{1}{2} \frac{\sigma_1}{\sigma_4}u_{10} + \frac{\sigma_2}{\sigma_4}u_{12} - u_5 - u_{15} - \mu(t)x_4$	$\frac{dx_4}{dt} = u_3 + \frac{1}{\sigma_4 V_1}u_6 + \frac{1}{2} \frac{\sigma_1}{\sigma_4}u_{10} + \frac{\sigma_2}{\sigma_4}u_{12} - u_{15} - \mu(t)x_4$
$\frac{dx_5}{dt} = \frac{1}{2} \frac{\sigma_1}{\sigma_5}u_9 + \frac{\sigma_3}{\sigma_5}u_{14} + \frac{\sigma_4}{\sigma_5}u_{15} - u_{17} - u_{16} - \frac{1}{\sigma_5 V_1}u_{18} - \mu(t)x_5$	$x_5 = \frac{1}{p_6 + p_7 + \mu(t)} \left(\frac{1}{2} \frac{\sigma_1}{\sigma_5}u_9 + \frac{\sigma_3}{\sigma_5}u_{14} + \frac{\sigma_4}{\sigma_5}u_{15} - \frac{1}{\sigma_5 V_1}u_{18} \right)$
$\frac{dx_6}{dt} = \alpha u_4 - u_{13} - \mu(t)x_6$	
$\frac{dx_7}{dt} = -\frac{1}{\sigma_7 V_2}u_{19} - \mu(t)x_7$	
$\frac{dx_8}{dt} = \alpha u_7 + \frac{1}{2} \frac{\sigma_6}{\sigma_8}u_{13} - \frac{1}{\sigma_8 V_2}u_8 - \mu(t)x_8$	$\frac{dx_8}{dt} = \frac{1}{2} \frac{\sigma_6}{\sigma_8}u_{13} - \frac{1}{\sigma_8 V_2}u_8 - \mu(t)x_8$
$\frac{dx_9}{dt} = \alpha u_5 + \frac{1}{2} \frac{\sigma_6}{\sigma_9}u_{13} - \frac{1}{\sigma_9 V_2}u_6 - \mu(t)x_9$	$\frac{dx_9}{dt} = \frac{1}{2} \frac{\sigma_6}{\sigma_9}u_{13} - \frac{1}{\sigma_9 V_2}u_6 - \mu(t)x_9$
$\frac{dx_{10}}{dt} = \frac{\sigma_5}{\sigma_{10}}u_{17} - \mu(t)x_{10}$	

821 *Appendix A.2. Model parameterization and initialization*

822 A total of 23 parameters are needed to fully define the reaction rates of
 823 Table A.3. Following [1], 9 of these parameters were fixed based on published
 824 data, which were obtained from research studies of peach or fruit. The re-
 825 maining 14 parameters were estimated numerically, as described in section
 826 5.2. In order to compare model and data, sugar concentrations at the fruit
 827 level have to be computed from model variables, describing the metabolite
 828 concentration within intra-cellular compartments. Assuming a constant pro-
 829 portion of vacuolar space in fruit cells, the concentration of each sugar j
 830 (sucrose, glucose, fructose, and sorbitol) at the fruit level is given by:

$$\mathbb{E}(X_j) = \mathcal{M}_p(x_i) = x_i^{vac} \frac{1}{\alpha + 1} + x_i^{cyt} \frac{\alpha}{\alpha + 1} \quad (\text{A.3})$$

831 where x_i^{vac} and x_i^{cyt} are respectively the variables located in the vacuole ($i \in$
 832 [6, 9]) and cytosol ($i \in [1, 5]$) (see Table A.2) and $\alpha = \frac{V_1}{V_2}$ is the intra-cellular
 833 volume ratio. The value of α was estimated by cytological analysis to 0.08 (
 834 see [1] for more information). Fruit fresh mass was assumed as a proxy for
 835 total volume $V_1 + V_2$.

For each genotype k , initial conditions $x_0^{(k)}$ were set using the concentra-
 tions $X^{(k)}(t_0)$ of sucrose, glucose, fructose, sorbitol, and hexose phosphates,
 measured at the fruit level at the first stage of development. The conversion
 between total and intra-cellular metabolite concentrations was performed
 based on metabolite localization at maturity. Accordingly, 98% of fructose,
 glucose, sucrose content and 90% of sorbitol content were assumed to be lo-
 cated in the vacuole, whereas the hexose phosphates were restricted to the
 cytosol. Accordingly, for sucrose, fructose, and glucose:

$$\begin{aligned} \text{cytosol : } & x_i^{(k)}(t_0) = 0.02 X^{(k)}(t_0) \frac{(1 + \alpha)}{\alpha} & i \in \{1, 3, 4\} \\ \text{vacuole : } & x_i^{(k)}(t_0) = 0.98 X^{(k)}(t_0) (1 + \alpha) & i \in \{6, 8, 9\} \end{aligned}$$

for sorbitol,

$$\begin{aligned} \text{cytosol : } & x_i^{(k)}(t_0) = 0.10 X^{(k)}(t_0) \frac{(1 + \alpha)}{\alpha} & i = 2 \\ \text{vacuole : } & x_i^{(k)}(t_0) = 0.90 X^{(k)}(t_0) (1 + \alpha) & i = 7 \end{aligned}$$

and for the hexoses phosphates

$$\text{cytosol: } x_i^{(k)} = X^{(k)}(t_0) \frac{(1 + \alpha)}{\alpha} \quad i = 10$$

Table A.5: Table of original and reduced models parameter description

p_i	Parameter	Corresponding model	Description	Reference	Value	Unit
p_1	λ_{Suc}	original and reduced	sucrose proportion hydrolyzed in the apoplasm	Estimated	0-1	
p_8	$TactifSuc$	original and reduced	coefficient of sucrose transport (active import) from cytosol to vacuole	Estimated	0-400	mg gFW ⁻¹ day ⁻¹
p_{10}	$TpassifGlu$	reduced	coefficient of glucose passive transport between cytosol and vacuole and in the opposite direction	Section. 3.1	112.1559	mg gFW ⁻¹ day ⁻¹
		original		Estimated	0-150	
p_9	$TpassifFru$	original and reduced	coefficient of fructose passive transport between cytosol and vacuole and in the opposite direction	Estimated	0-150	mg gFW ⁻¹ day ⁻¹
$r_1 = \frac{v_2}{p_5}$	$R_{susy} = \frac{V_{susy}}{K_{susy}}$	reduced	coefficient of the transfer function between sucrose and (fructose + hexoses phosphate) under action of sucrose synthase (susy) enzyme	Section. 3.2.2	1.8809	day ⁻¹
$r_2 = \frac{v_3}{p_{21}}$	$R_{ni} = \frac{V_{ni}}{K_{ni}}$	reduced	coefficient of the transfer function between sucrose and (glucose +fructose) under action of neutral invertase (ni) enzyme	Section. 3.2.2	95.5875	day ⁻¹
$r_3 = \frac{v_4}{p_{22}}$	$R_{sdh} = \frac{V_{sdh}}{K_{sdh}}$	reduced	coefficient of the transfer function between sorbitol and fructose under action of sorbitol dehydrogenase (sdh) enzyme	Section. 3.2.2	7.1592	day ⁻¹
$r_4 = \frac{v_5}{p_{13}}$	$R_{so} = \frac{V_{so}}{K_{so}}$	reduced	coefficient of the transfer function between sorbitol and glucose under action of sorbitol oxydase (so) enzyme	Estimated	0-10	day ⁻¹
$r_5 = \frac{v_1}{p_{23}}$	$R_{ai} = \frac{V_{ai}}{K_{ai}}$	reduced	coefficient of the transfer function between sucrose and (glucose +fructose) under action of acid invertase (ai) enzyme	Estimated	0-1	day ⁻¹
p_2	Ki_{Ai}	original	inhibitor constant of acid invertase	Estimated	0-10	mg gFW ⁻¹
$r_6 = \frac{v_6}{p_3}$	$R_{fk} = \frac{V_{fk}}{K_{fk}}$	reduced	coefficient of the transfer function between fructose and hexoses phosphate under action of fructokinase (fk) enzyme	Estimated	0-5000	day ⁻¹
v_7	$Vhk(t)$	reduced	hexokinase activity (hk) to transfer glucose to hexoses phosphate	Section. 3.2.1	$86.2 - 2.3t + 2e^{-2t^2} - 8.3e^{-5t^3}$	mg gFW ⁻¹ day ⁻¹
p_4	Khk	original and reduced	hexokinase affinity	Estimated	1-300	mg gFW ⁻¹
p_7	$ReSyntSuc$	original and reduced	coefficient of the transfer function between hexoses phosphate and sucrose	Estimated	0-300	day ⁻¹
p_6	$OthComp$	original and reduced	coefficient of the transfer function between hexoses phosphate and other compounds	Estimated	450-1500	day ⁻¹
p_{14}	$TpassifSor$	reduced	coefficient of sorbitol passive transport between cytosol and vacuole	Section. 3.1	4.1305	mg gFW ⁻¹ day ⁻¹
		original		Estimated	0-150	
σ_f	$PropCdw$	original and reduced	carbon concentration of the mesocarp	[43]	0.44	gC gDW ⁻¹
p_{17}	q_g	original and reduced	growth respiration coefficient	[43]	0.084	gC gDW ⁻¹
p_{18}	q_m	original and reduced	maintenance respiration coefficient	[43]	2.76e-4	gC gDW ⁻¹ day ⁻¹
p_{16}	Q_{10}	original and reduced	temperature ratio of maintenance respiration	[43]	1.9	
p_{15}	λ	original and reduced	sucrose sap proportion	[1]	0.65	
p_{11}	$VmtactifFru$	original	fructose active import (activity)	Estimated	0-150	mg gFW ⁻¹ day ⁻¹
p_{12}	$VmtactifGlu$	original	Glucose active import (activity)	Estimated	0-150	mg gFW ⁻¹ day ⁻¹
p_{19}	$KmtactifGlu$	original	Glucose active import (affinity)	[44]	0.054	mg gFW ⁻¹
p_{20}	$KmtactifFru$	original	fructose active import (affinity)	[44]	0.288	mg gFW ⁻¹

836 **Appendix B. Multi-variate sensitivity analysis**

837 *Appendix B.1. Multi-variate sensitivity analysis*

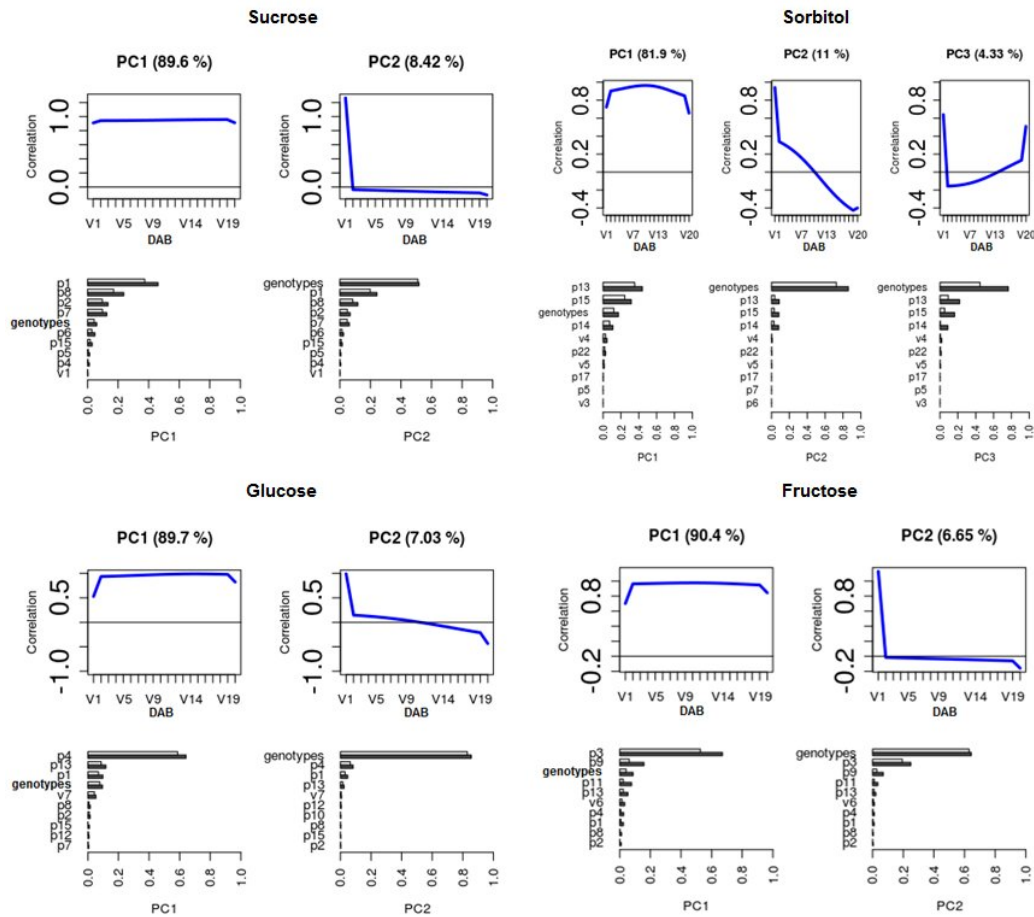


Figure B.1: PCA-based sensitivity analysis of the sugar model. Columns: principal components. Top row: correlation coefficients (y-axis) between the principal component and the output of each sugar during fruit development (DAB, days after bloom on the x-axis). Bottom row: first order sensitivity indices (dark bars) and total sensitivity indices (pale bars).

838 Multivariate sensitivity [30] was used to identify the influence of each
 839 parameter on the dynamic output $x(t)$ during fruit development. Where
 840 $x(t)$ is the sugar concentration (sucrose, glucose, fructose and glucose) and
 841 t is the independent time variable for 20 days after bloom ($t = (V1 =$

842 $\min(DAB), V2 = \max(DAB)/2 + 2, \dots, V19 = \max(DAB)/2 + 19, V20 =$
843 $\max(DAB))$. Results of the principal components and sensitivity principal
844 indices are presented in Fig. B.1. For sucrose, glucose and fructose, the first
845 two components explained 96% of the total inertia of the simulated sugar
846 dynamics. For sorbitol, the first three components explained 97%. The first
847 component was positively correlated with all time-points. Correlation val-
848 ues in Fig. B.1 show that the first principal component corresponds to the
849 average concentration of sugars (sucrose, glucose, fructose and sorbitol) pro-
850 duced during the whole fruit development. The second principal component
851 was positively correlated with sugar concentration at stage 1 and poorly or
852 slightly negatively correlated with simulated sugar during the second half
853 of fruit development. Thus, the second principal component corresponds to
854 the difference in sugar initialization values, that strongly depends on the
855 **genotype** factor. For sorbitol, the third principal component accounts for a
856 much smaller part of inertia, associated with the difference between the sor-
857 bitol produced in the middle of fruit development and the sorbitol produced
858 both very early and late. It was sensitive to the set of genotypes.

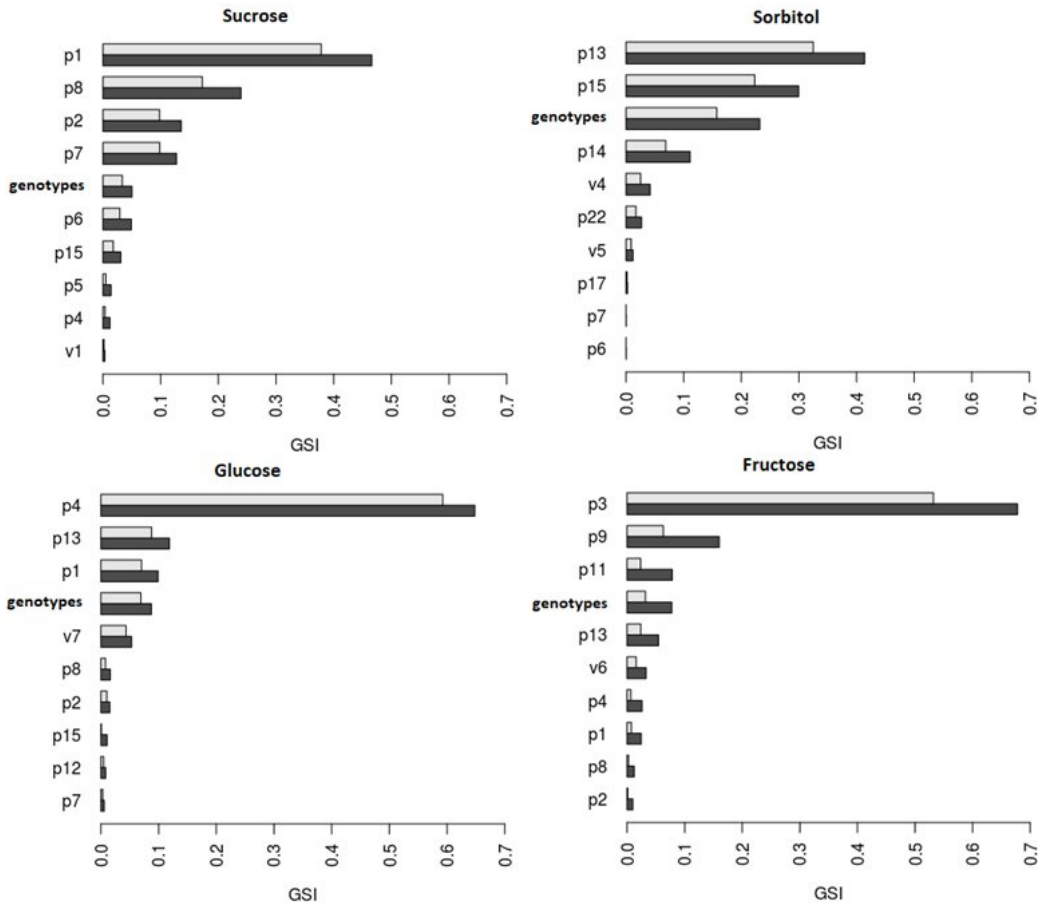


Figure B.2: Generalized sensitivity indices (GSI) for the first ten sensitive parameters (p_i) and ten genotypes (the training set) on four outputs (Sucrose, Sorbitol, Glucose, and Fructose) of the sugar model. The main sensitivity indices are in dark bars and interaction ones are in grey bars.

859 The generalized sensitivity indices (GSI) shown in Fig. B.2 gives a com-
 860 mon ranking of model parameters according to their influence on the four out-
 861 put sugars (Sucrose, Sorbitol, Glucose and Fructose), for all tested genotypes.
 862 Parameter p_1 related to the action of cell-wall invertase in fruit apoplasm is
 863 the most important parameter, for its effect on both sucrose (rank 1) and
 864 glucose (rank 3) dynamics. The activities of Fructokinase (p_3) and Hexok-
 865 inase (p_4) are the most sensitive parameters for fructose and glucose con-
 866 centrations, respectively, whereas the sorbitol oxydase affinity (p_{13}) and the
 867 proportion of sucrose in the plant sap (p_{15}) affect sorbitol content in the fruit.

868 Interestingly, the **genotype** factor is ranked only third to fifth, depending on
869 the sugar, meaning that it does not affect parameters' sensitivity as much as
870 expected. A closer look at the results shows that the choice of the genotype
871 essentially affects the second principal component, via the definition of the
872 initial conditions of the model (see the supplemental information Fig. B.1).

873 **Appendix C. Virtual experiment**

874 In order to evaluate the reliability of the proposed simplifications over a
875 larger diversity, a progeny of virtual genotypes was randomly created based
876 on a careful recombination, with noise, of the original 10 dynamics. This
877 included changes in parameters values, initial conditions and input functions.

878 We used the results from the principal component analysis (PCA) per-
879 formed on growth rate and growth duration for the whole progeny of 106
880 genotypes to verify the distribution of virtual genotypes. To this aim, growth
881 rates and durations of the 20 000 virtual genotypes were projected on the
882 PCA plan defined by the previous analysis. As shown in Fig. C.3, the virtual
883 genotypes provide a good representation of the diversity in growth rate and
884 growth duration observed in the real progeny.

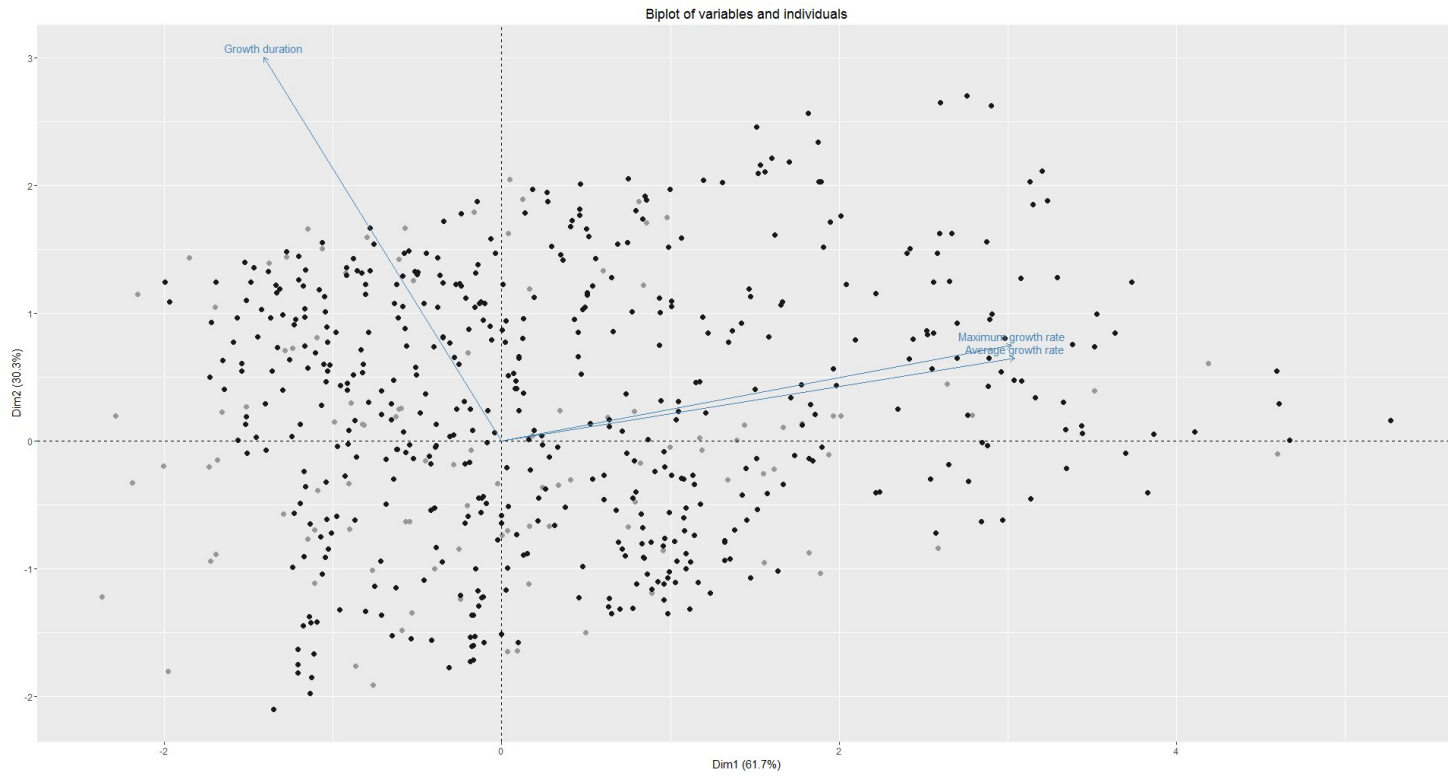


Figure C.3: Principal component analysis (PCA) for the whole progeny of 106 genotypes (grey) and 500, out of 20000, virtual genotypes (black). Represents the projection on the Dim1 and Dim2 of the growth duration and growth rate obtained with curves growth.

885 Appendix D. Time-scale analysis and QSSA

886 Timescale-based approaches and quasi-steady-state approximation [17]
887 were applied to reduce the number of ODEs of the model and to obtain
888 the final reduced model. The predicted concentrations of sugars in both in-
889 tracellular compartments were analyzed. A fast transient dynamics of the
890 concentration of the hexose phosphate (x_5), followed by a slow one, was
891 observable in the numerical simulations of the original and intermediate re-
892 duced model (Fig. D.4. Together with the analysis of the Jacobian matrix,
893 this observation led to the assumption of x_5 as a fast variable of the system.
894 Quasi-steady-state approximation on x_5 , indeed, strongly reduced the fast
895 transient dynamics in the final reduced model, for most genotypes. Notice
896 that a few fast modes (already pointed out by the analysis of the Jacobian
897 matrix) may nonetheless remain in the system. Their elimination would re-
898 quire a linear combination of the original variables, which is incompatible
899 with our objective to preserve the biological interpretation of the model. We
900 therefore decided not to push the simplification of the model further.

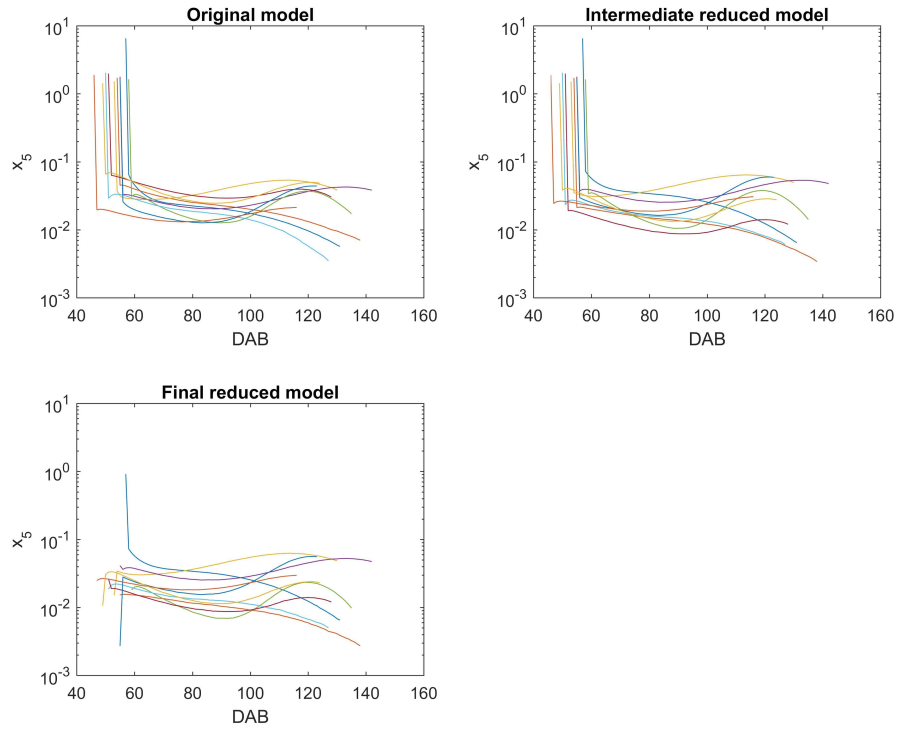


Figure D.4: Evolution of the concentration ($mgFW^{-1}$) of x_5 : *Hexose Phosphate* during fruit development (DAB, days after bloom) for ten genotypes for the original, intermediate reduced and final models.

901 *Appendix D.1. Results of quasi-steady-state approximation applied on the*
 902 *intermediate reduced model for the 20 000 virtual genotypes*

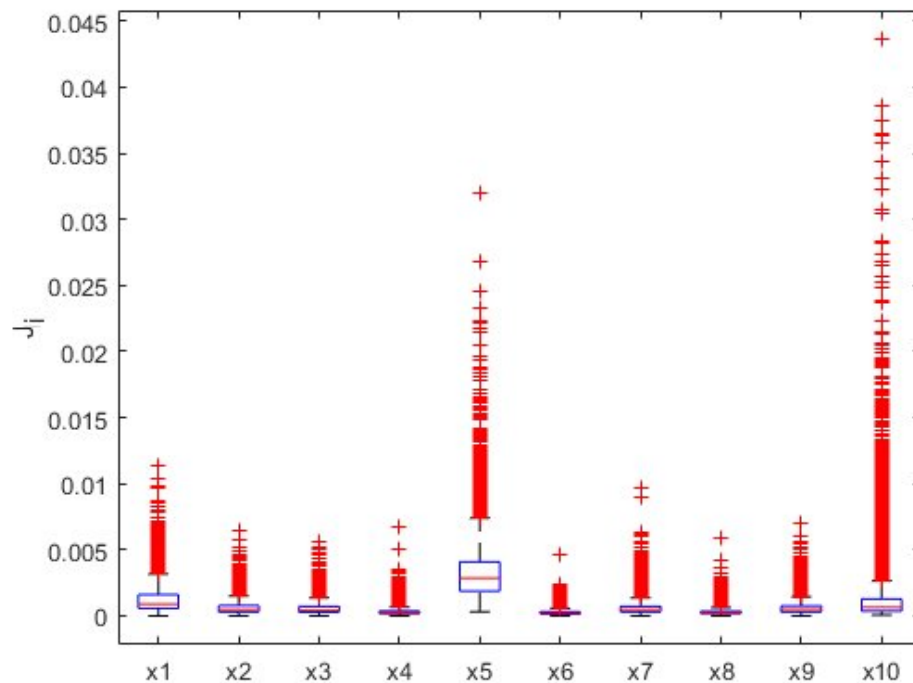


Figure D.5: Normalized Root Mean Square Errors $J_i, i \in \{1, \dots, 10\}$ between the intermediate and reduced model after application of the QSSA to x_5 . The boxplot shows the variability of J_i over the virtual genotypes

903 We compared the intermediate reduced model with its QSS approximation
 904 tion by calculating the Normalized Root Mean Square Error (J_i) on the 20
 905 000 virtual genotypes. All J_i was very low, less than 0.045, over the whole
 906 dynamics for all variables (Fig. D.5).

Table D.6: NRMSE between model simulation and experimental data. Calculated values of the normalized root mean squared error (NRMSE) are presented for each genotype, the four sugars separately.

	Genotype	Phenotype	Year	Sucrose	Sorbitol	Fructose	Glucose	Mean
Trainig set	E1	Standard	2012	0.09	0.14	0.16	0.26	0.16
	E33	Standard	2012	0.04	0.19	0.27	0.17	0.16
	E43	Standard	2012	0.07	0.11	0.24	0.16	0.14
	F111	Standard	2012	0.13	0.13	0.21	0.23	0.17
	C227	Standard	2011	0.07	0.28	0.16	0.13	0.16
	E22	Low	2012	0.11	0.09	0.21	0.18	0.14
	F106	Low	2012	0.07	0.7	0.14	0.15	0.26
	F146	Low	2012	0.05	0.14	0.02	0.17	0.10
	H191	Low	2012	0.07	0.15	0.18	0.16	0.14
	C216	Low	2011	0.08	0.26	0.25	0.11	0.17
Validation set	H163	Standard	2012	0.10	0.35	0.18	0.19	0.20
	F107	Standard	2012	0.11	0.11	0.24	0.25	0.17
	E23	Standard	2012	0.15	0.30	0.10	0.17	0.18
	E17	Standard	2012	0.14	0.37	0.19	0.05	0.18
	E21	Standard	2012	0.13	0.21	0.16	0.24	0.18
	E41	Low	2012	0.08	0.35	0.35	0.24	0.25
	E18	Low	2012	0.13	0.26	0.26	0.09	0.18
	F113	Low	2012	0.12	0.32	0.35	0.20	0.24
	F90	Low	2012	0.06	0.47	0.26	0.13	0.23
	C243	Low	2012	0.18	0.49	0.24	0.09	0.25
	C199	Low	2012	0.22	0.46	0.25	0.19	0.28
	C207	Low	2012	0.19	0.28	0.23	0.24	0.23
	E36	Low	2012	0.09	0.13	0.13	0.16	0.12
	E48	Low	2012	0.07	0.41	0.14	0.14	0.19
	F101	Low	2012	0.15	0.43	0.20	0.05	0.20
	F109	Low	2012	0.07	0.31	0.27	0.24	0.22
	F127	Low	2012	0.09	0.17	0.22	0.10	0.14
	F144	Low	2012	0.19	0.11	0.24	0.22	0.19
	F141	Low	2012	0.14	0.22	0.36	0.16	0.22
	F86	Low	2012	0.06	0.13	0.30	0.32	0.20
	C232	Standard	2012	0.05	0.30	0.25	0.13	0.18
	E5	Standard	2012	0.22	0.07	0.12	0.24	0.16
	E19	Standard	2012	0.05	0.14	0.24	0.03	0.11
	E20	Standard	2012	0.19	0.17	0.07	0.18	0.11
	E26	Standard	2012	0.05	0.20	0.18	0.38	0.20
	E34	Standard	2012	0.22	0.43	0.20	0.15	0.25
	E35	Standard	2012	0.11	0.27	0.19	0.07	0.16
	E37	Standard	2012	0.24	0.24	0.13	0.26	0.21
	F83	Standard	2012	0.09	0.19	0.06	0.27	0.15
	F115	Standard	2012	0.03	0.20	0.10	0.07	0.10

907 The NRMSE can defined as follows:

$$NRMSE(\tilde{p}^{(k)}) = \sum_{i=1}^4 J_i(\tilde{p}^{(k)})$$

908 with

$$J_i(\tilde{p}^{(k)}) = \frac{\sqrt{\frac{1}{N_M} \sum_{j=1}^{N_M} (\tilde{x}_i(t_j, \tilde{p}^{(k)}) - X_i^{(k)}(t_j))^2}}{\max_j(X_i^{(k)}(t_j)) - \min_j(X_i^{(k)}(t_j))} \quad (\text{D.1})$$

909 where N_M is the number of observations, $\tilde{x}(t, \tilde{p}^{(k)})$ are the concentrations
910 predicted by the model and $X^{(k)}t(t)$ are the experimental data and i is the
911 sugar index.