STOCHASTIC MODEL FOR SIMULATING MAIZE YIELD

E. R. Detomini, D. Dourado Neto, J. A. Frizzone, A. Doherty, H. Meinke,
K. Reichardt, C. T. S. Dias, M. G. Figueiredo

ABSTRACT. *Maize is one of the most important crops in the world. The products generated from this crop are largely used in the starch industry, the animal and human nutrition sector, and biomass energy production and refineries. For these reasons, there is much interest in figuring the potential grain yield of maize genotypes in relation to the environment in which they will be grown, as the productivity directly affects agribusiness or farm profitability. Questions like these can be investigated with ecophysiological crop models, which can be organized according to different philosophies and structures. The main objective of this work is to conceptualize a stochastic model for predicting maize grain yield and productivity under different conditions of water supply while considering the uncertainties of daily climate data. Therefore, one focus is to explain the model construction in detail, and the other is to present some results in light of the philosophy adopted. A deterministic model was built as the basis for the stochastic model. The former performed well in terms of the curve shape of the above-ground dry matter over time as well as the grain yield under full and moderate water deficit conditions. Through the use of a triangular distribution for the harvest index and a bivariate normal distribution of the averaged daily solar radiation and air temperature, the stochastic model satisfactorily simulated grain productivity, i.e., it was found that 10,604 kg ha^{-1} is the most likely grain productivity, very similar to the productivity simulated by the deterministic model and for the real conditions based on a field experiment.*

Keywords. *Bivariate normal distribution, Corn, Crop modeling, Depleted productivity, Grain productivity, Triangular distribution.*

The advent of crop models implemented on computers can be traced back to groundbreaking work in the 1950s, such as the study by Monsi and Saeki (1953) on light interception and de Wit's (1958) classic "Transpiration and Crop Yields" that also draws on some of Penman's early work (Penman, 1948). These and similar publications constructed the framework for the emerging formalism of system analysis (Zadoks and Rabbinge, 1985). Phrasing physiological processes in mathematical terms and collating them to meteorological variables led to today's proliferation of computer simulation models that have been developed and used in agriculture.

Submitted for review in August 2009 as manuscript number BE 8178; approved for publication by the Biological Engineering Division of ASABE in May 2012.

The authors are **Euro Roberto Detomini,** PhD, Department of Biosystems Engineering, University of São Paulo, Brazil; **Durval Dourado Neto,** Full Professor, Department of Crop Sciences, University of São Paulo, Brazil; **José Antonio Frizzone,** Full Professor, Department of Biosystems Engineering, University of São Paulo, Brazil; **Alastair Doherty,** Computer Programmer, Department of Primary Industries and Fisheries, Toowoomba, Queensland, Australia; **Holger Meinke,** Adjunct Professor, Department of Land and Water, Wageningen Rural University, Wageningen, The Netherlands; **Klaus Reichardt,** Soil Physics Researcher, Centre of Nuclear Energy in Agriculture, University of São Paulo, Brazil; **Carlos Tadeu dos Santos Dias,** Full Professor, Department of Exact Sciences, University of São Paulo, Brazil; and **Margarida Garcia de Figueiredo,** Professor Adjunto, Agribusiness and Regional Development Postgraduate Program, Federal University of Mato Grosso, Brazil. **Corresponding author:** Euro Roberto Detomini, ESALQ/USP, Departamento de Engenharia de Biossistemas, Av. Pádua Dias 11, Piracicaba - SP, CEP 13418-900, Brazil; phone:+55-19-3429-4217; e-mail: erdetomini@hotmail.com.

Simulation models of agricultural plants, crops, and cropping systems are becoming commonplace. Traditionally, they have been used as knowledge depositories by scientists to describe an area of interest. Once available, interest quickly shifted from curiosity about the underlying principles to the use of models either in a predictive capacity (e.g., to develop scenarios or to support decisions) or in an explanatory capacity to investigate interactions between processes studied in an isolated manner. This manner of studying models initiated a debate about the appropriateness of mathematically describing biological relationships and the level of details needed to achieve a "good" model. Defining this goodness, by clearly stating the objectives of every modeling endeavor, could make much of that debate redundant (Meinke, 1996).

Arguments about the right way to build crop models have largely concentrated on the level of empiricism acceptable when representing such sequences mathematically. Passioura (1996) asserted that the purpose of scientific models is to improve our understanding of physiology and environmental interactions, while engineering models utilize robust, empirical relationships to obtain results. This separation would constitute a traditional reductionist paradigm because it would reinforce the disassociation of scientific and engineering modeling rather than allow for a synthesis of the different approaches. Rather than separating engineering from science and alienating many professionals in the process, it might be more useful to view this differentiation as the pragmatic end of a continuous quest for knowledge and a solution to the problems. Used constructively, this polarity should advance future model developments (Meinke, 1996).

Thus, models are simplified representations of a system that is a part of the real world and contains related components inside predefined boundaries. This system can be affected by the surroundings, but the surroundings cannot affect it significantly. The definition of scale is very important to ensure that the conclusions of a system will often be based on the performance of the low-hierarchy components. The main roles of models, in our understanding, are to: (1) organize information; (2) highlight gaps in the various research areas of knowledge; (3) visualize a robust idea about the potentialities, limitations, or eventually magnitudes of a given variable of interest; and (4) simulate impossible and difficult scenarios (e.g., CO_2 injection on earth, insect biology studies). Moreover, the existence of a coincidence does not necessarily imply a cause-effect relationship.

The applicability of crop models emerges when one needs to optimize the use of resources such as land and water under given boundary conditions. For example, if an economist intends to calculate the highest possible profitability under certain resource constraints (i.e., land and water) over the course of an agricultural project, the analysis regarding how much water and land area could be available for cropping during a couple of years will depend on refined and well-established crop models. This analysis certainly justifies modeling research efforts meant to improve simulation outputs. It is convenient, for instance, to classify models only as deterministic and stochastic, regardless of other non-straightforward distinctions. Although deterministic models can output a solution through simple mathematical implementation, they are limited because they provide only a single outcome. In contrast, stochastic models, which use statistical methods or stochastic components, provide a range of results, with each result associated with its corresponding probability of occurrence.

Maize is one of the most important crops in the world. The products generated from this crop are largely used in the starch industry, the animal and human nutrition sector, and biomass energy production and refineries. For these reasons, there is frequently a significant interest in knowing the potential grain yield of maize genotypes in relation to the environment in which they will be subjected for cropping, as productivity directly affects agribusiness or farm profitability. The main objective of this work is to conceptualize a stochastic model for predicting maize grain yield and productivity under different conditions of water supply while considering the uncertainties of daily climate data. A deterministic model and some outputs are also analyzed to evaluate the time course of the above-ground dry matter growth of maize.

Materials and Methods

Model Description and a Brief Parameterization

The proposed model includes concepts from both generic (i.e., the family of Dutch models) and maize-specific (i.e., CERES-Maize; Jones and Kiniry, 1986) approaches. Our maize model includes assimilation processes and depends on few empirical input parameters; it does not predict any crop phenology except the physiological maturity point as a function of thermal time. Dobermann et al. (2003) provided a good comparison between each family of models by pointing out a number of advantages and disadvantages of each family.

From dimensional analysis, Detomini (2008) derived the following mechanistic equation for estimating the potential above-ground dry matter (DM) on a daily basis:

$$Wpa = 0.1498 \cdot GP \cdot \lambda_I \cdot RUE \quad (1)$$

where *Wpa* is the potential above-ground dry matter (kg DM ha^{-1} d^{-1}) estimated for a given day, *GP* is the gross photosynthesis rate (kg CH_2O ha^{-1} d^{-1}), λ_I is the fraction of solar radiation intercepted by the canopy, and *RUE* is the radiation use efficiency (g above-ground dry matter MJ^{-1} intercepted photosynthetically active radiation).

The *GP* function was also deducted by collating Clapeyron's law with dimensional analysis according to Detomini (2008):

$$GP = \frac{36.5854 \cdot GAR \cdot LAI \cdot H \cdot P}{T + 273} \quad (2)$$

where *GAR* is the gross assimilation rate (μL CO_2 cm^{-2} leaf area h^{-1}), *LAI* is the leaf area index, *H* is the day length (h d^{-1}), *P* is the atmospheric pressure (atm) that relies on altitude (*Alt*, m), and *T* is the daily average air temperature (°C). Clapeyron's law is generally used to convert the volume of a given substance into its corresponding mass, leading to a justification of *GAR* units in a volume basis. Substituting equation 2 into equation 1 yields:

$$Wpa = 5.48 \frac{GAR \cdot LAI \cdot H \cdot P}{T + 273} \lambda_I RUE \quad (3)$$

Bouger-Lambert's law states that photosynthetically active radiation transmitted (PAR_t, MJ m^{-2} d^{-1}) vertically through a canopy can be derived from Beer's law (Monsi and Saeki, 1953):

$$PAR_t = PAR_0 \left(e^{-k \cdot LAI} \right) \quad (4)$$

where PAR_0 refers to the incident photosynthetically active radiation flow (MJ m^{-2} d^{-1}) on top of the canopy, and k is the light extinction coefficient. The ratio PAR_t/PAR_0 defines transmittance (τ) so that the complementary fraction (1 – τ) defines the intercepted fraction (λ_I). For practical purposes, λ_I could also be understood as a canopy covering fraction if canopy leaves are randomly oriented and spread (Loomis and Connor, 1992, p. 274).

The basic difference between Bouger-Lambert's law and Beer's law is that the latter assumes a homogeneous mean, which does not occur in plant populations (Loomis and Connor, 1992, p. 36). Thus, k should not be constant through either the canopy profile or the crop cycle. The light extinction coefficient rapidly decreases as *LAI* values increase during the initial stages of crop development, but it is likely to assume a constant value if the canopy closes fast (i.e., under irrigated field conditions). Because of this, the initial variations of *k* can be neglected for crop model purposes in non-limiting conditions of water (Meinke, 1996).

It is important to highlight that equation 3 explains the daily above-ground dry matter as a function of plant variables, such as *RUE*, *k* (implicit, see eq. 4), *GAR*, and *LAI*, and as a function of seasonal and climate variables, such as day length (*H*), atmospheric pressure (*P*), temperature (*T*), and absorbable radiation flow. This latter variable can be explicitly stated by expressing *GAR* (eq. 7) as a function of equation 5, which relies on both air temperature and solar radiation flow (implicit, see eq. 6). By analyzing data from the graphs shown by Heemst (1986) for C4 plants, Detomini (2008) presented an empirical generic function for describing the potential gross assimilation rate (GAR_p, $\mu L\ CO_2\ cm^{-2}$ leaf area h^{-1}) under controlled conditions:

$$GAR_p = \frac{A_0 + A_1 R_{ab} + A_2 R_{ab}^2 + A_3 R_{ab}^3 + A_4 \ln(T)}{1 + A_5 R_{ab} + A_6 R_{ab}^2 + A_7 R_{ab}^3 + A_8 \ln(T) + A_9 \left[\ln(T)\right]^2} \tag{5}$$

where T is the air temperature ($0 < T < 40°C$), and R_{ab} is the absorbable photosynthetically active radiation flow ($0 < R_{ab} < 0.35$ cal cm^{-2} leaf area min^{-1}). The empirical, non-user-defined parameters are: A_0 = 1.566792, A_1 = 53515909, A_2 = -221.805971, A_3 = 310.191491, A_4 = -0.491961, A_5 = -0.190506, A_6 = 0.373910, A_7 = -0.088166, A_8 = -0.554728, and A_9 = 0.080398.

In fact, the absorbable photosynthetically active radiation flow was derived from a dimensional analysis by Detomini (2008) as a function of solar radiation flow (*Rg*, MJ m^{-2} d^{-1}) and day length (*H*, h d^{-1}) with some corrections:

$$R_{ab} = 0.3987 \frac{Rg}{H} \lambda_{PAR} \lambda_{ab} \left(1 - \lambda_r\right) \tag{6}$$

According to Sinclair and Muchow (1999), representative values for λ_{PAR} and λ_{ab} would be 0.5 (MJ photosynthetically active radiation MJ^{-1} incident solar radiation) and 0.85 (MJ absorbable radiation MJ^{-1} photosynthetically active radiation), respectively. Oguntunde and van de Giesen (2004) suggested a value of λ_r = 0.23 (MJ soil-plant reflected radiation MJ^{-1} incident solar radiation) for maize crop albedo.

GAR_p is corrected by cloudiness effects according to:

$$GAR = \left[F_{Nub} R_{Adc} + \left(1 - F_{Nub}\right)\right] GAR_p \tag{7}$$

where R_{Adc}, specific to genotype, is the relationship between gross assimilation rates under a clear sky and gross assimilation rates under an overcast sky; and F_{Nub}, specific to environment, is a cloudiness factor responsible for correcting the theoretical potential gross assimilation rate. The magnitude of the former was simulated to be around 0.2626 (Detomini, 2008), whereas the latter might be directly obtained if insolation data are available or, if not, estimated by implying that radiation flow during strongly overcast days accounts for 20% of the flow on very clear days, according to:

$$F_{Nub} = 1.25 \left(1 - \frac{Rg}{\left(a_{AP} + b_{AP}\right) R_T}\right) \tag{8}$$

where a_{AP} and b_{AP} are the Angström-Prescott coefficients, and R_T is the estimated radiation incident on top of the atmosphere once the Earth's eccentricity ($DRST^2$) and sunset hour angle (*Ahn*, degrees) are calculated:

$$R_T = 37.6 DRST^2 \left[Ahn \cdot \sin\left(\Phi \frac{\pi}{180}\right) \sin\left(\zeta \frac{\pi}{180}\right) + \cos\left(\Phi \frac{\pi}{180}\right) \cos\left(\zeta \frac{\pi}{180}\right) \sin\left(Ahn\right)\right] \tag{9}$$

$$DRST^2 = 1 + 0.033 \cos\left(DOY \frac{2\pi}{365}\right) \tag{10}$$

$$Ahn = \frac{\pi H}{24} \tag{11}$$

Note that equation 7 turns the generalized equation (eq. 5) into a specific condition by collating information from the plant (R_{Adc}), the climate (*Rg*), and the local (a_{AP}, b_{AP}, and R_T) conditions for a given day of the year (*DOY*). The resulting angle between the imaginary plane of the equator and the imaginary line that links earth to the sun defines the solar declination (ζ, degrees) and relies on DOY:

$$\zeta = 23.45 \sin\left[\frac{2\pi}{365}\left(DOY - 80\right)\right] \tag{12}$$

Thus, the length of a given day might be obtained as:

$$H = \frac{24}{\pi} \arccos\left[-\tan\left(\zeta \frac{\pi}{180}\right) \tan\left(-\Phi \frac{\pi}{180}\right)\right] \tag{13}$$

where Φ refers to the latitude (decimal values) of the place of interest. Negative values of Φ are conventionally set for locations in the southern hemisphere.

Because plant populations depend on both plant density of sowing (D_{sow}, plants m^{-1}) and spacing (S_e, m), these measurements are required for defining the leaf area index (*LAI*) in addition to the leaf area (*LA*, cm^2 $plant^{-1}$). At a farm-level view:

$$LAI = 10^{-4} LA \frac{D_{sow}}{S_e} \tag{14}$$

Several models have been proposed to approach plant leaf area. For example, Dwyer and Stewart (1986) introduced a bell-shaped function for estimating a single maize leaf as a function of the number of fully expanded leaves. This model was remarkable because it revealed many meaningful parameters, such as skewness, breadth, and the largest area and position of a single leaf (Valentinuz and Tollenar, 2006). However, there is still some empiricism for extrapolating this model for the whole plant because it requires empirical linear fittings for predicting the largest area and leaf position, in addition to the eventual leaf number. We suggest using a closed empirical Gauss model, which is somewhat similar to the mechanistic model presented by Yang and Alley (2005). The functional form of our Gauss model is:

$$LA\left(Dr\right)=\gamma_1 e^{-0.5\left(\frac{Dr-\gamma_2}{\gamma_3}\right)^2} \tag{15}$$

where Dr is the relative crop development, and γ_1, γ_2, and γ_3 are the empirical parameters.

The first parameter (γ_1) of equation 15 is biologically meaningful because it represents the maximum value of the entire plant leaf area (LA_{max}), equal to 8654.91 cm^2 in a model where Dr = γ_2 = 0.5758 (Detomini, 2008). The meaning of the third parameter (γ_3) is not well established, although it approximately a quarter of the entire Dr (γ_3 = 0.2473). Instead of considering time for crop development duration, this parameter was considered on a dimensionless basis (i.e., Dr) to allow better generalization for future comparisons with data obtained from other studies and experiments.

The relative crop development defines the cumulative thermal time (CTT_j, °C d) on the jth day after emergence in relation to the cumulative thermal time at a physiological maturity point (CTT_{mpp}, °C d), consistent with the following approach:

$$Dr_j=\frac{CTT_j}{CTT_{mpp}} \tag{16}$$

The value of CTT_{mpp} was found as 1392.82°C d from field experimental data (Detomini et al., 2008). However, the variable is user-defined in the model and varies according to maize genotype. The CTT is the sum of single daily thermal times (TT_j, °C):

$$CTT_j=\sum_{j=1}^{n}TT_j \tag{17}$$

If the relative development rate is well represented by a non-linear function but has a linear relationship with TT, this variable can also be described by a non-linear function. After reviewing "degree-days" concepts, Bonhomme (2000) explained the limitations of traditional degree-day calculation methods and suggested a beta function to calculate TT, with upper (TD_{max}, °C), lower (TD_{min}, °C), and optimal (TD_{op}, °C) temperatures for development. First proposed by Yin et al. (1995), the beta function is assumed as the best option to calculate thermal times because it is based on relatively flexible mathematical laws and has few parameters, all physically and biologically meaningful, according to:

$$TT=$$
$$TT_{max}\left(\frac{TD_{max}-T}{TD_{max}-TD_{op}}\right)\left(\frac{T-TD_{min}}{TD_{op}-TD_{min}}\right)^{\frac{TD_{op}-TD_{min}}{TD_{max}-TD_{op}}} \tag{18}$$

Assuming user-defined values of TT_{max}, TD_{max}, TD_{op}, and TD_{min} equal to 25, 44, 35, and 0, respectively, the function that particularly describes thermal time as a function of air temperature will only depend on daily air temperature (T, °C), for example:

$$TT=0.0793(44-T)T^{3.89} \tag{19}$$

A helpful approach to establishing a relationship between water and yield productivity was proposed by Doorenbos and Kassam (1979):

$$\frac{Wa}{W}=\prod_{i=1}^{n}\left[1-ky\left(1-\frac{WUa}{WU}\right)\right]_i \tag{20}$$

The ky values for most crops are derived under the assumption that the relationship between relative yield (Wa/W) and relative water use (WUa/WU) is linear and is valid for water deficits up to about 50%, or (1 – WUa/WU) = 0.5. The values of ky are based on an analysis of experimental field data covering a wide range of growing conditions. The experimental results used represent high-producing crop varieties, well adapted to the growing environment and grown under a high level of crop management. The ratio WUa/WU may either occur continuously throughout the entire growth period of the crop, or it may occur during any one of the individual growth periods (i.e., i = 1, 2, …, n). Analysis of the available experimental field data in terms of the more precisely defined stress-day and drought indices proved difficult. On the other hand, if a simulation process is modeled on a one-day scale, the original production model of Doorenbos and Kassam (1979) collapses because the day-to-day multiplication of a decimal factor would quickly deplete the production. To solve this problem, our model assumes a continuous and equally distributed day-to-day water deficit to calculate the productivity depletion, as will be shown later, so that a value of 1.25 might be adopted for maize hybrids in this situation. Thus:

$$1-\frac{Wpad}{Wpa}=1.25\left(1-\frac{ETa}{ET}\right) \tag{21}$$

where $Wpad$ is the depleted, above-ground dry matter productivity rate (kg ha^{-1} d^{-1}), ETa is the actual evapotranspiration (mm d^{-1}), and ET is the potential evapotranspiration (mm d^{-1}).

The relationship ETa/ET could be interpreted as an index of water stress under which a plant population develops (Laar et al., 1992, p. 40), which is equal to one if there is no deficit and zero if the deficit is severe. In reality, this relationship indicates the water supply level (Sw) because its

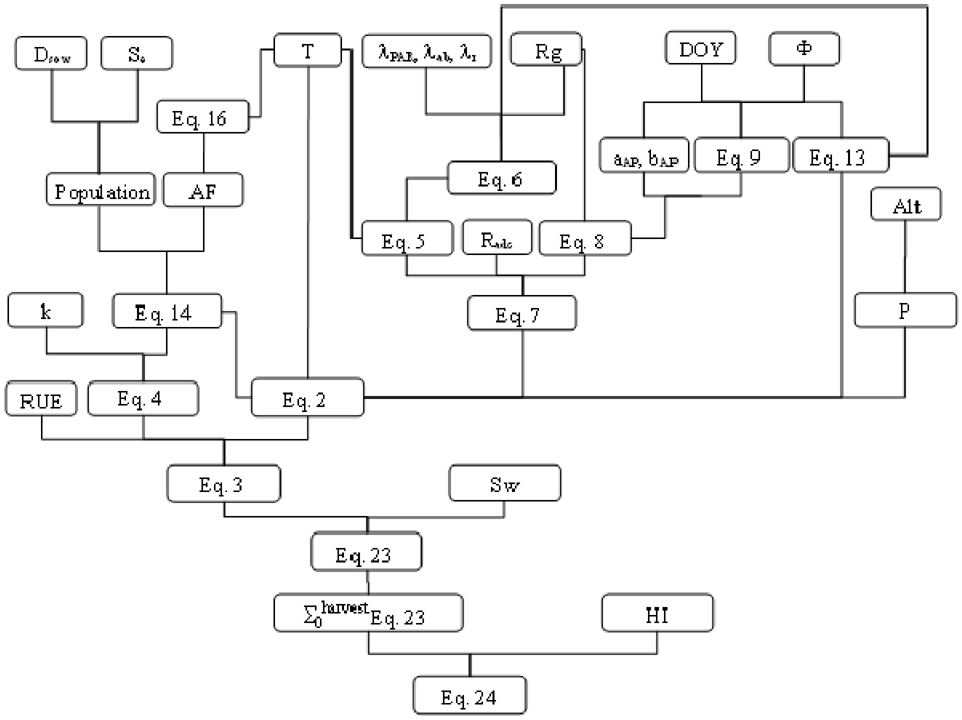

Figure 1. Flowchart summarizing the model used for simulating maize yield.

opposite (i.e., 1 – *ETa*/*ET*) would be the level of stress. If *Wpad* is explicit and *Sw* is inserted into equation 21:

$$Wpad = Wpa\left[1 - 1.25\left(1 - Sw\right)\right] \tag{22}$$

The term inside the brackets is, therefore, a depletion factor (*Fd*; kg potential above-ground dry matter productivity rate kg^{-1} depleted above-ground dry matter productivity rate). By inserting equation 3 into equation 22, it is possible to find a simple model to predict depleted above-ground dry matter productivity rates on a daily basis as:

$$Wpad = 5.48\frac{GAR \cdot LAI \cdot H \cdot P}{T + 273}\lambda_I RUE\left[1 - 1.25\left(1 - Sw\right)\right] \tag{23}$$

In agreement with the approach of Verdoodt et al. (2004), the summation of daily dry matter (eq. 23) until reaching the economically useful stage of a crop, which in this model is the maturity point, results in the final cumulated above-ground dry matter productivity that should account for harvest index (*HI*; kg grain kg^{-1} depleted above-ground dry matter) in the calculation for the grain productivity simulation (*Wg*; kg grain ha^{-1}):

$$Wg = HI\sum_{j=0}^{n} Wpad_j \tag{24}$$

According to equation 24, there is no plant weight loss. This weight stability might be justified because the daily calculations of *Wpa* already account for the efficiency of radiation use (*RUE*), and the senescing losses are accounted for in the *LAI* equation. Equation 24 is the final deterministic model used to predict maize grain productivity (on a dry basis) under specific limited and non-limited water supply, climate, population, site, and time conditions. The stochasticity can be implemented simultaneously through climate variables (i.e., temperature and solar radiation) and the harvest index.

The flowchart in figure 1 summarizes the main concept of the model. Some equations are implied in the flow and are not presented. For example, it is observed that the variable *T* is used in equation 16, but it is important to mention that *T* is first used in equation 19, which leads to equation 17 and then to equation 16. The input variables D_{sow} and S_e (both related to population), *DOY* (related to the date of sowing), Φ and *Alt* (both related to place), and *Sw* (related to water management) are all user-defined variables, as are *RUE*, *k*, λ_{PAR}, λ_{ab}, λ_r, and R_{Adc} (all related to light-plant relations). The user-defined variable *HI* is considered in a stochastic manner through a triangular distribution, whereas *T* and *Rg* can be likewise considered through a triangular (if there is no data series) or bivariate normal distribution.

An initial parameterization was considered as: *RUE* = 3.52, *k* = 0.4257, *HI* (mode = 0.4, maximum = 0.43, minimum = 0.38), and *LA*(*Dr*) according to equation 15. We used equation 5 for gross assimilation rate and equation 19 for thermal time. We also used user-defined values for latitude (-22.425), Angström-Prescott coefficients (a_{AP} = 0.25; b_{AP} = 0.5), plant population (66,666 plants ha^{-1}), row spacing (0.9 m), and date of sowing (24 November), in addition to the experimental weather data. Above-ground dry matter productivity rates (*Wpa*, kg ha^{-1} d^{-1}), cumulated above-ground dry matter (*W*, kg ha^{-1}), and maize grain productivity (*Wg*, kg ha^{-1}) were calculated using equations 23 and 24 through a deterministic simulation. Additionally, R_{Adc} = 0.26 was iteratively adjusted.

A sensitivity analysis was run in a simple manner by analyzing how much the grain yield (in percentage) would be increased or reduced for each added or subtracted unit of MJ m^{-2} d^{-1} of solar radiation, considering a correlation between solar radiation and temperature, in relation to the yield observed from climate data of the field experiment carried out by Detomini (2008). Prior to these analyses, we also performed a sensitivity analysis for the empirical and fitted equation 5, which represents the variation of potential gross assimilation rate as a function of solar radiation for many levels of air temperature (see fig. 2).

Stochastic Procedure Using Either Bivariate Normal Distribution or Triangular Distribution

Considering a variable such as maize grain yield, the model is deterministic because it will reproduce the exact same outputs for a given set of input variables. For situations where externalities or uncertainties can be neglected because they have little effect on the outputs, deterministic models are reasonably acceptable in some situations (e.g., irrigation design, mixture of chemicals under controlled conditions, or fertilizer recommendation for "homogeneous" soils). On the other hand, variability might be determinant on the final output (e.g., on the same day of the year in a given place, the average air temperatures might be around 10°C, 20°C, or even 30°C, highly affecting the final plant growth rate). A stochastic procedure would consider this variability by releasing too many outputs instead of only one, with each one related to its corresponding probability of occurrence.

Air temperature, for example, can reveal infinite possibilities of occurrence, even though some frequency around a given value will most likely occur. This frequency can be low or high for extreme or expected values. A group of infinite possibilities associated with their corresponding values can be eventually expressed by a probability density function, and its integration gives the probability function if all function properties are confirmed. A Monte Carlo method consists of inverting the resulting integration and explicitly declaring the variable, when possible, remaining only using known values of the probabilities for the simulation initiation. In fact, pseudo-random numbers are the starting point of a Monte Carlo simulation. Matsumoto and Nishimura (1988) developed the Mersenne-Twister algorithm that is used in most recognized statistical packages to generate the starting points. The simulation error (ε) is minimized inversely according to the iteration number (*Ni*), yet it also relies on the deviation (σ) of a dataset:

$$\varepsilon = \frac{3\sigma}{\sqrt{Ni}} \tag{25}$$

Within the various existing probability distributions, either discrete or continuous, the normal distribution is the most important from the agronomic knowledge viewpoint, not only because many processes are well explained by it but also because it is a sort of "outer limit" of most of the distributions. Ideally, a distribution is adopted for a dataset if it describes the distribution properly, which can be verified by tests like Kolmogorov-Smirnov. Because a variable distribution is an average, the central limit theorem states that if a sample is too large ($n \rightarrow \infty$), variance is then minimized [$(\sigma^2/n) \rightarrow 0$] by trending to zero, leading a variable to approximately follow a normal distribution [$\sim N(\mu, \sigma^2/n)$] (i.e., strongly bell-shaped) even for non-normal populations. Additionally, the normal distribution fits many sample probability distributions very well. A simple starting point for simulating two variables that presumably follow normal distributions, based on a high *Ni*, is to use the Box-Muller transformation to obtain auxiliary variables 1 and 2 (N_1 and N_2), which depend on a previous generation of pseudo-random numbers U_1 and U_2 that are independent from each other and are uniformly distributed according to (Box and Muller, 1958, p. 610):

$$N_1 = \left[-2\ln\left(U_1\right)\right]^{0.5}\cos\left(2\pi U_2\right) \tag{26}$$

$$N_2 = \left[-2\ln\left(U_1\right)\right]^{0.5}\sin\left(2\pi U_2\right) \tag{27}$$

For example, if one intends to generate solar radiation and air temperature for each day of the year, the first step would be to generate two pseudo-random values (U_1 and U_2, both between 0 and 1) and to replicate this step *Ni* times (i.e., Ni = 10,000), i.e., there will be 10,000 values for both U_1 (${}^1U_1$, ${}^2U_1$, …, ${}^{10000}U_1$) and U_2 (${}^1U_2$, ${}^2U_2$, …, ${}^{10000}U_2$) for each day. As a result, variables ${}^1N_1$, ${}^2N_1$, …, ${}^{10000}N_1$ and also ${}^1N_2$, ${}^2N_2$, …, ${}^{10000}N_2$ will exist. As solar radiation drives practically all processes on Earth, the value of this variable is hierarchically generated first. By adapting the procedure provided by Hogg and Craig (1978), the *i*th value of solar radiation (${}^iRg$) for whatever day for the *i*th computer-generated auxiliary-1 variable (${}^iN_1$) is:

$${}^iRg = \mu_{Rg} + \sigma_{Rg}\,{}^iN_1 \tag{28}$$

where μ_{Rg} and σ_{Rg} are calculated (from the dataset) from the average and standard deviation of solar radiation for one day. Note that equation 28 is the inversion of the standardized variable *Z*, where $Z \sim N(\mu_{Rg}, \sigma_{Rg})$.

The *i*th value of temperature (${}^iT$) can be subsequently simulated for the same day by considering the *i*th computed auxiliary-2 generated variable (${}^iN_2$), the calculated average and standard deviation of air temperature (μ_T and σ_T, respectively) for the same day, the simulated value from equation 28, and Pearson's product-moment coefficient correlation (ρ) existing between the variables *Rg* and *T*:

$${}^iT = \left[\mu_T + \left(\rho\frac{\sigma_T}{\sigma_{Rg}}\right)\left({}^iRg - \mu_{Rg}\right)\right] + \left(1-\rho^2\right)^{0.5}\sigma_T\,{}^iN_2 \tag{29}$$

Although it is not the main focus of this work, bivariate normality can be validated through asymmetry and kurtosis coefficients, whereas Bartlett's test is useful for checking variance homogeneity of simulated values in comparison to the observed values. This validation would allow us to identify the type of simulation that best agrees with the observations. Pearson's correlation coefficient may vary from -1 to 1 and equals zero if variables are independent, whereas covariance evaluates how the dependent variables "walk" together independent of sample size (Wonnacott and Wonnacott, 1985, p. 132). If ρ = 0 in equation 29, the T calculation is arranged in an analogous manner to equation 28. The correlation measures the strength and linear direction between two quantitative variables. For our samples (Moore, 1995, p. 111):

$$\rho = \frac{\sum_{i=1}^{n}\left[\left(Rg_i - \mu_{Rg}\right)\left(T_i - \mu_T\right)\right]}{\sqrt{\sum_{i=1}^{n}\left(Rg_i - \mu_{Rg}\right)^2}\sqrt{\sum_{i=1}^{N}\left(T_i - \mu_T\right)^2}} \tag{30}$$

If there are no climate data, it is necessary to search for information from a specialist to derive the subjective parameters of the triangular distribution, which is often used in agribusiness when one wants to subjectively describe a population of a continuous variable. The three key parameters of the triangular distribution are mode or "most likely" (Mo), maximum ($V_{\max}$), and minimum ($V_{\min}$). To initiate the triangular generation through the Monte Carlo method, a critical pseudo-random number (Uc) should first be calculated according to:

$$Uc = \frac{Mo - V_{\min}}{V_{\max} - V_{\min}} \tag{31}$$

Because of the function discontinuity, three possibilities of function inversion become:

$$X = Mo \text{ if } U = Uc \tag{32}$$

$$X = V_{\min} + \left[U\left(V_{\max} - V_{\min}\right)\left(Mo - V_{\min}\right)\right]^{0.5} \quad \text{if } U < Uc \tag{33}$$

$$X = V_{\max} - \left[\left(1-U\right)\left(V_{\max} - V_{\min}\right)\left(V_{\min} - Mo\right)\right]^{0.5} \quad \text{if } U > Uc \tag{34}$$

Evidence from field experiments shows that as biomass increases, harvest index usually decreases with a non-mechanistic explanation. The uncertainty of the harvest index and environment relationships also encourages the use of a stochastic procedure during the simulation processes, regardless of assuming empirical relationships or even a single value. Thus, the model assumes a triangular distribution for simulating the harvest index by considering user-defined values of the mode, maximum, and minimum. In summary, 10,000 pseudo-random values are generated to produce the same amount of radiation values and their subsequently correlated air temperatures, both of which follow normal distributions. This procedure is done for each day (after emergence) of a crop cycle. Ten thousand values of the remnant dependent variables are subsequently calculated, resulting in 10,000 values of grain yield that have to be classified per frequency class in a previously calculated class number. For a given iteration, each calculated value of *Wpad* (eq. 23) in a given day is summed with the calculated *Wpad* of the previous day. There is certainly an implied error by doing this calculation because two opposite extreme values of radiation may be simulated for two consecutive days, sometimes without any physical sense. Nevertheless, such error is inversely minimized according to the number of iterations (eq. 25). Ten thousand *Ni* values is a large enough sample size.

The implementation of the stochastic procedures in conventional worksheets would be possible but tiring and quite confusing in terms of presentation because several outputs would be generated and, consequently, a large number of lines or rows would be required. Therefore, it became convenient to develop a tool in the Visual Studio 2005 (C#) platform to run the model to allow for the best visualization of the results. Some input parameters are mandatory, such as sowing date, latitude and hemisphere, crop spacing, sowing density, mass of a thousand seeds, Angström-Prescott coefficients (the program suggests additional values), and the distribution probability. One should opt for triangular distribution if no climate data are available or for bivariate normal distribution if climate data are available. In the former case, a meteorology text file needs to be selected. Setting the plant parameters is optional. The main output is a histogram of yield probabilities.

To stochastically evaluate the model, a deterministic simulation was first run to check the growth of the above-ground dry matter throughout the maize cycle. For specific purposes, we consider the term "calibration" when a known observation of a dependent variable (i.e., maize grain yield) is used to predict an input variable (i.e., R_{Adc}). Thus, the maize deterministic model was calibrated according to an iterative procedure to find R_{Adc} from Wg = 10,472 kg ha^{-1} (averaged grain yield) observed in the field and assuming constant the other input variables. The input variables *RUE*, *k*, and *LA*(*Dr*) were chosen according to Detomini et al. (2008); the variables λ_{PAR}, λ_{ab}, λ_r, a_{AP}, and b_{AP} were adopted according to the literature; R_{Adc} was iteratively adjusted as a function of grain productivity (Wg = 10,472 kg ha^{-1}) found by Detomini (2008); and *DOY*, Φ, D_{sow}, and S_e were defined to meet field experiment conditions. Then the stochastic procedure was run for a different place, e.g., Rockhampton, Australia. The climate variables *Rg* and *T* were simulated according to a bivariate normal distribution fed by a climate dataset for this location, and *HI* was set according to a triangular distribution fed by practical suggestion based on field experiment results.

By selecting the weather dataset for Rockhampton (23° 22′ 30″), localized near the Tropic of Capricorn in Australia, for comparison with the Piracicaba weather dataset, stochastic simulations were run by setting the same date and variable values assumed for the deterministic simulation, i.e., *RUE* = 3.52, *k* = 0.4257, *HI* (mode = 0.40, maximum = 0.43, minimum = 0.38), *LA*(*Dr*) (see eq. 15), equation 5 for gross assimilation rate and equation 19 for thermal time, Rockhampton latitude and Angström-Prescott coefficients

(a_{AP} = 0.25; b_{AP} = 0.5), plant population (66,666; row spacing = 0.9 m), and date of sowing (24 November). Four situations were considered for the stochastic simulation of maize grain: (1) bivariate normal distribution for solar radiation and air temperature under non-water deficit conditions, (2) triangular distribution for solar radiation and air temperature under non-water deficit conditions, (3) reduction of plant population by taking D_{sow} = 5 plants m^{-1} under non-water deficit conditions, and (4) water supply of 75% (Sw = 0.75). The corresponding results are shown in figure 6. In case 2, the mode as well as maximum and minimum values for both variables were based on the Rockhampton radiation and temperature dataset (from 24 November up to 120 days after) as follows: mode of 18 MJ m^{-2} s^{-1} and 21°C, maximum of 32 MJ m^{-2} s^{-1} and 32°C, and minimum of 6 MJ m^{-2} s^{-1} and 15°C.

RESULTS AND DISCUSSION

DETERMINISTIC SIMULATION AND SENSITIVITY ANALYSIS

A comparison of the observed (field experiment) vs. simulated (eq. 24) data reveals that the model has a slightly overestimated grain productivity of 1.4%, giving an output, for example, of 10,620 kg ha^{-1}. If general data are interpolated from Heemst (1986) for obtaining R_{Adc}, we would find values of over 0.3 for this variable. Because specific values of R_{Adc} are seldom explored in the literature for specific genotypes, this problem would exemplify the importance of models, namely, that they allow for the identification of specific magnitudes of a given variable of interest without needing to experimentally derive them through alternative experimental conditions, which might require additional expenses.

It can be seen from a sensitivity analysis performed for equation 5 that the value of GAR_p, which depends on solar radiation, decreases from a maximum of 33°C even if R_{ab} is rising, according to figure 2, which could also be represented by a surface. The graph is presented in a convenient, illustrative manner. For example, GAR_p is less responsive to R_{ab} when under 25°C compared to 39°C, which is a similar behavior to that found by Lizaso et al. (2005). In spite of being built empirically, the curves satisfactorily represent the process of interest when compared to the curves (similar to a rectangular hyperbolic shape, except for low temperature values) presented by Pachepsky et al. (1996). For physiological reasons, potential gross assimilation rates substantially decrease for overcast days such that it is recommended to make the necessary corrections prior to estimating the actual GAR, following the approaches of Heemst (1986) and Verdoodt et al. (2004). This decrease is the justification for the existing equation 7 to deplete GAR_p. We believe one good route to specify GAR for each maize hybrid would be to adjust R_{Adc} and biological meaningful parameters of the beta function (eq. 18).

Identifying the sensitivity of the maize model yield to solar radiation (the main driving variable) is essential because temperature variation is also calculated using correlation (i.e., eqs. 29 and 30). For example, by adopting climate data from the field experiment of Detomini (2008), figure 3 reveals that each extra unit of MJ m^{-2} d^{-1} added to the daily solar radiation flow values throughout the maize cycle would increase the yield by approximately 4%, which is a variation smaller than that of each unit of MJ m^{-2} d^{-1} subtracted from each daily value. By correlation, the temperature varies linearly by more or less than 0.6°C for each added or subtracted MJ m^{-2} d^{-1}, respectively. If the correlation between solar radiation flow and air temperature is neglected (i.e., varying the former but keeping the latter constant), the magnitudes of the simulated yields resulting from added units of radiation become even greater, and yields resulting from subtracted units become even smaller, in comparison to the yields simulated from the aforementioned field experiment climate data.

It is worth noting that increasing the amount of radiation by 6 MJ m^{-2} d^{-1} (and ~3.6°C in temperature by correlation) would shorten the time to tasseling by 12 days if the assumptions of thermal time concepts are valid. A decrease in

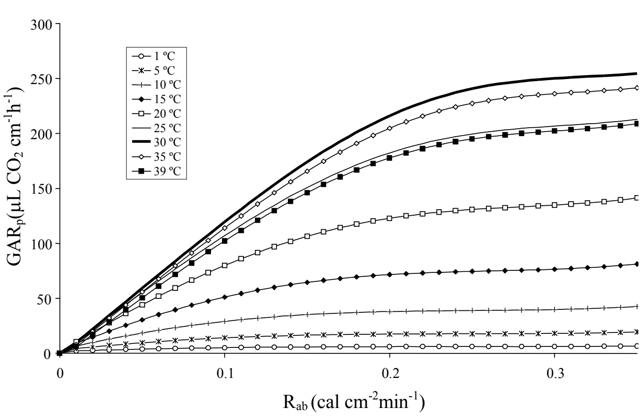

Figure 2. Potential gross assimilation rate as a function of absorbable photosynthetically active radiation flows for different air temperatures.

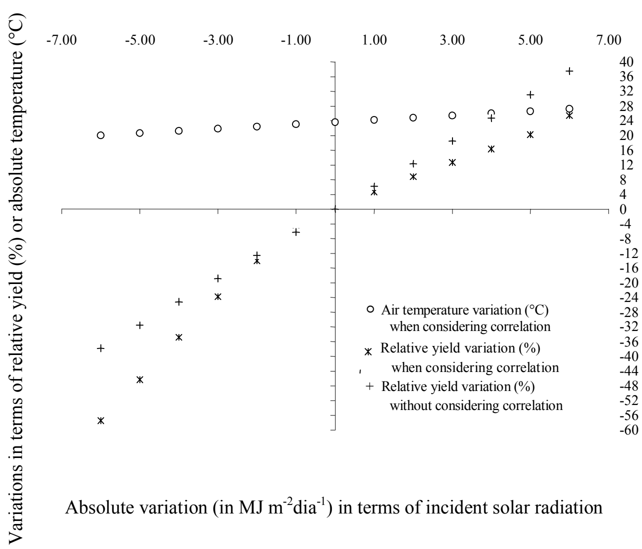

Figure 3. Sensitivity analysis for the absolute variation in temperature and relative yield from each unit of MJ m^{-2} d^{-1} varied in terms of incident solar radiation.

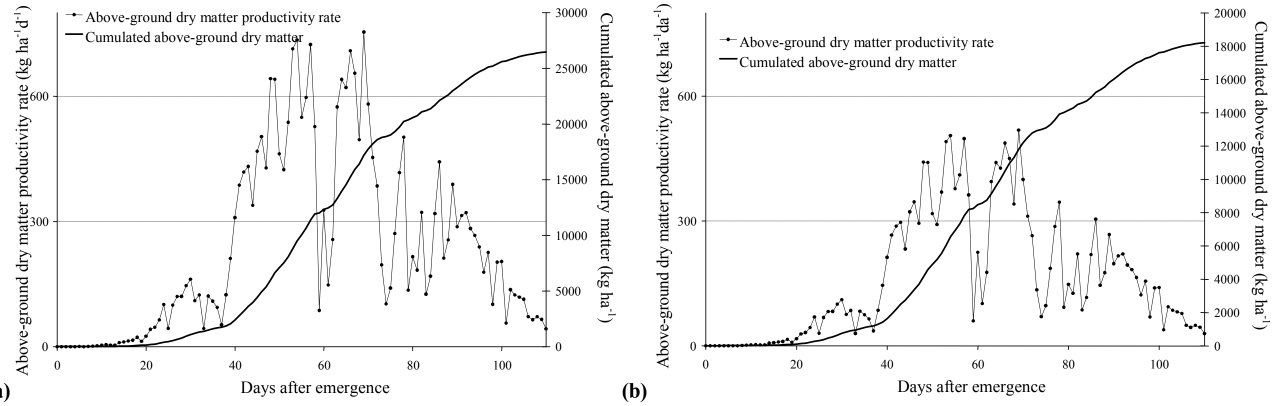

Figure 4. Variation of simulated above-ground dry matter productivity rate and cumulated productivity as functions of days after emergence: (a) *Sw* = 1 and (b) *Sw* = 0.75.

radiation by the same value would stretch this time to 23 days. In fact, this phenological stage was observed in the field at 55 days after emergence. The proposed model does not release the outcome of tasseling. However, these variations in time are realistic because shortening the maize cycle could eventually lead to a premature definition of the number of leaves, whereas stretching it could provoke excessive growth of vegetative components, especially with regard to stalk height, mass, and carbohydrate storage. In the case of increasing radiation, for example, the anticipated growth determination could reduce the productivity in a manner contradictory to that identified by sensitivity analysis unless the hybrid was sown in a non-recommended place and season. In a second case, for constant given water and nitrogen conditions, the sink-source relationships could be altered to reduce the yield because the competition of assimilates among the different plant tissues would increase.

The harvest index is another attribute that could be strongly modified with the variation of both climate variable magnitudes and maize biological cycle length, resulting in either a reduction or increase in grain yield. However, this sensitivity is difficult to analyze, as the harvest index generally has an inverse relationship with plant growth, which does not necessarily mean that the productivity would be smaller, as observed by Detomini (2008). Additionally, it is difficult to study the effects of climate inputs on *HI* because the variation is not even throughout the cycle, besides the fact that the main periods (vegetative, flowering, and grain filling) contribute through different routes to the *HI* and are different from each other in terms of weather sensitivity.

Quite high values of above-ground dry matter productivity rates (*Wpa*) were calculated, which surpassed 600 kg ha^{-1} d^{-1} during the most exigent stage of the cycle (flowering); yet low values of *Wpa* were simulated for the initial stages (i.e., before 40 DAE), as expected due to low values of *LAI* and light interception during theses stages. Loomis and Connor (1992, p. 41) reported rates at a magnitude of 520 kg above-ground dry matter ha^{-1} d^{-1}, presumably for less efficient hybrids in *RUE*. According to figure 4, above-ground dry matter productivity rates tend to naturally decrease as a consequence of senescence processes, although there is still dry matter accumulation because of a partitioning process and carbohydrate transference. Dry matter accumulation nearly followed a sigmoidal shape, similar to the explanatory model developed by Detomini et al. (2008), who adjusted field experiment observation data to follow a sigmoidal curve. Data from Andrade (1995) corroborate the magnitudes of yield and the curve shape of the accumulation of above-ground dry matter, albeit analyzed by a different hybrid and different experimental harvest index. In fact, the model in the present study does not simulate processes such as senescence or carbon transfers; these processes are approached here only to justify some behaviors of the deterministic simulation.

High values of above-ground dry matter productivity rates during some days of flowering can be supported by coincidences of high values of air temperature with greater leaf area index values during the flowering period. Even so, full-cover canopy is subjected to abrupt reductions in terms of productivity rates if climate conditions are not favorable for plant development, as shown near 60 days after emergence (figs. 4a and 4b). Extremely high values of productivity rates are not frequent in agricultural practice because irrigation and other non-controlled variables are rarely controllable for the purpose of achieving such desirable combinations. The experimental data were simulated through APSIM (APSRU, 2007), and there were no water or nitrogen stresses throughout the entire maize cycle, according to APSIM released indexes. Thus, we point out that the productivity rates verified in the field experiment are related to potential conditions of plant growth and development, although different environmental situations could lead to different potential productivities for maize, in agreement with Dobermann et al. (2003).

On the other hand, if maize was grown under a 25% water deficit uniformly distributed throughout a cycle (i.e., $Sw = 0.75$ and thus $Fd = 0.69$), there would be a 30.26% yield reduction, as grain productivity would be 7,303 kg ha^{-1} (fig. 4b) instead of 10,620 kg ha^{-1}. This "what if" question is not realistic for non-irrigated cases because soil water continuously fluctuates throughout the entire growing season. However, as long as there is a necessity for controlling soil water contents due to reducing water allocations in irrigated agriculture, the assumption of a 25% water deficit could provide different results in terms of an economic viability extrapolation. For example, what is the magnitude of yield and therefore the income that could be achieved by reducing the water levels by 25% as a result of a reduced water allocation imposed by a government? Would it be viable to irrigate the crop under this situation? Technically, soil water devices allow for the checking and controlling of soil water levels with good accuracy, so that setting the level of soil water is somewhat reasonably manageable today (i.e., irrigation precision).

One source of error in figure 4b could be the assumption of a constant k (eq. 17). As k varies with *LAI*, especially during the early stages, Meinke (1996) showed for wheat that in situations where a high leaf area can be achieved, the slight underestimation of early dry matter production by conventional simulations that use a constant k has no significant consequences for dry matter production at anthesis, unless the constant k is substantially underestimated. However, in situations where maximum *LAI* values are low, as is frequently the case in dryland production areas, cumulated dry matter (i.e., until anthesis) can be significantly underestimated. The value of light coefficient extinction should be considered in every simulation model (Monteith, 1965; Jones and Kiniry, 1986; Meinke et al., 1997). For example, the software APSIM 5.2 (APSRU, 2007) assumes a constant $k = 0.45$ for maize. Another difference is that APSIM (APSRU, 2007) also considers non-continuous functions for describing *TT* (i.e., *TT* described linearly and for more than one condition), different from the function of equation 18, which is continuous.

The curve in figure 4a was simulated for non-limiting conditions of water and nitrogen. It can be compared to an explanatory curve derived from experimental data by Detomini (2008). The explanatory (statistical) model was fitted to a sigmoidal curve with three parameters ($p < 0.05$, F-test = 5083.07, and $r^2 = 0.9866$). By comparing the deterministic model curve (for non-limiting conditions) and explanatory model curve, it can be seen in figure 5 that the former (shown in fig. 4a) tends to underestimate the cumulated above-ground dry matter during the initial stages of the maize cycle and overestimate it after these stages.

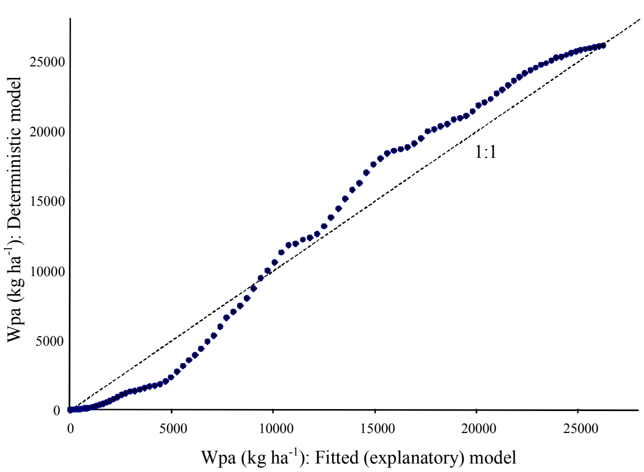

Figure 5. Comparison between deterministic model curve and explanatory model curve (statistically fitted from field experimental data).

According to the comparison (fig. 5) between the deterministic model curve and the explanatory model curve fitted from field experimental data, it is presumed that the model satisfactorily simulates maize grain productivity (*Wg*, kg ha^{-1}), although the above-ground dry matter productivity rates (*Wpa*, kg ha^{-1} d^{-1}) and the cumulated above-ground dry matter (*W*, kg ha^{-1}) are either underestimated or overestimated in some specific stages throughout the time course of the maize cycle. The time courses (daily basis) shown in figures 4a and 4b express the natural behavior of maize growth as a whole and confer a correct biophysical meaning to the model, in agreement with results obtained by Verdoodt et al. (2004). The values implied on the curves are also realistic. Therefore, the evaluation of the model through deterministic simulation and comparison of the models with field experiment observations enable the implementation of stochastic procedures on the deterministic model.

It is convenient to mention that the model assumes that there is no nutrient stress during the crop cycle. Soil fertility is considered to be sufficiently upgraded to such a level that it is not necessary to be concerned about it when micronutrients are correctly supplied. This is very realistic because farmers engaged in professional agriculture generally search for necessary corrections to all soil nutrient shortages prior to sowing the crop. In doing so, they are only concerned about maintaining adequate nitrogen levels to the set plant population management variable. On the other hand, the model simulates the nitrogen requirement for all simulated yields and outputs only the most likely maximum and minimum requirements. This nitrogen model relies on many input parameters and was not considered in this article. Nitrogen requirement is actually one of the outputs of the model.

Stochastic Simulations

Several values of daily average solar radiation and daily average air temperature are possible, but there is a corresponding probability for each one of the daily values according to the distribution to which they are likely to be fit. As a consequence, maize dry matter productivity rates and grain productivity can also vary. A stochastic procedure would approach these variations toward the identification of not only a single value but also the various classes of grain productivities associated with their corresponding probabilities. The generation of thousands of pseudo-random numbers is the beginning step and, particularly for the case of the model presented here, works to generate input values of solar radiation and air temperature according to either a bivariate normal distribution or a triangular distribution (if no climate series data are available) and also to generate the harvest index according to a triangular distribution. Finally, thousands of above-ground dry matter productivity rates and dry matter accumulations would have to be generated, regardless of single curves, as presented in figures 4a and 4b. For convenience, histograms of maize grain productivity are presented instead.

Ideally, the necessity of testing the multivariate normality hypothesis becomes explicit when a researcher intends to evaluate whether the presupposed conditions were met in terms of the inference validation that will be done. The existence of a "most suitable" multinormality test has been always undermined (Cantelmo and Ferreira, 2007), such as the test based on asymmetry and kurtosis deviation (Mardia, 1975) or the test based on Shapiro-Wilk generalization (Royston, 1983). Nevertheless, Mecklin and Mundfrom (2004) reinforced that a single method is not sufficient to approach the multinormality issue and that it is appropriate and even useful to run multiple tests together from an applied statistical perspective. For these reasons, the model proposed herein is pragmatically not subjected to a previous test before releasing a value of radiation or temperature to start the stochastic simulations.

The existence of a multivariate normal distribution implies the existence of evenly distributed marginal distributions, although the combined distribution of two variables that follow two separate, normal, univariate distributions does not necessarily follow one normal, multivariate distribution. According to Hair et al. (2005), in the case in which a univariate normality presupposition is not violated, this distribution will not necessarily result in an acceptance of multivariate normality, although it will help to obtain multivariate normality. Therefore, if one variable follows a multivariate normal distribution, it is also univariate such that the simulation of climate data through a bivariate normal distribution will favor the non-rejection of multivariate normality.

The most likely grain productivity for case 1 was 10,604 kg ha^{-1} (11.53% probability) according to figure 6a, which is very similar to the productivity simulated using deterministic conditions for Piracicaba. Probabilities around 0.02% were verified for both extreme (maximum and minimum) grain productivities (magnitudes of 12,673 and 8,634 kg ha^{-1}, respectively). One peculiarity of a normal distribution is that the grain productivity average is also the most likely value to occur, as the average is equal to the mode. If climate data were not available, one could use a triangular distribution. There is an unknown correlation in this situation, in principle, given that there is no dataset. Because of this, it makes no sense to run an adherence test such as the Kolmogorov-Smirnov (non-parametric) or chi-squared (parametric) test, in spite of the reciprocal situation being true, i.e., from a dataset, it would be possible to fit the data to a triangular distribution and run the tests.

When simulating maize grain productivity from a triangular distribution, the highest probability (11.5%) corresponds to the achievement of around 11,000 kg ha^{-1}. By selecting case 2, the grain mode becomes 400 kg ha^{-1} greater than that of case 1, and the "shape" of the histogram in figure 6b is not as symmetric as the other mentioned simulation situations. The small increase in productivity can be justified because, for most of the days of a crop cycle, both solar radiation and air temperature triangular distributions eventually allow the computation of productivity rates greater than those computed when a bivariate normal distribution is used. This computation is acceptable because there is an implied error in using a triangular distribution, as the mode, maximum, and minimum values are kept constant throughout the maize cycle, which is, indeed, not true.

By keeping the water supply conditions constant but reducing the sowing density to 5 plants m^{-1} (i.e., reducing the plant population by ~11,111 plants ha^{-1} as a consequence), the resulting histogram shows a productivity break around 2191 kg ha^{-1}, which is roughly 20% smaller than the potential productivity achieved in case 1, as can be seen in figure 6c. Provided that the plant population is kept as previously mentioned and the water level supply is reduced to 75% (Sw = 0.75 and Fd = 0.69), the model simulates a mode of 7,576 kg ha^{-1} (11.97% probability) for grain-depleted productivity (fig. 6d). As a conclusion, the model reveals that the best decision in a case of water scarcity would be to reduce the population instead of irrigating with a 75% water deficit, at least for the reductions in both population and water supply tested here. It is certainly in this way that models are useful, i.e., they allow for the determination of how much a change in a variable affects the output to aid in assertive decision making.

All the histograms shown in figure 6 are based on simulated production data. Measured crop yield or production data do not behave typically like a Gaussian distribution (unlike the histograms in fig. 6). Nevertheless, measured or production data are often not sufficient in number of observations to build a large sample (i.e., a sample with 10,000 observations). As mentioned in the Materials and Methods section, the central limit theorem states that if a sample is too large, then variance is minimized by trending to zero, leading a variable to approximately follow a normal distribution (i.e., strongly bell-shaped) even for non-normal populations.

Although harvest index values have been simulated by a triangular distribution, it might be seen that all histograms of yield frequency prevail with a normal distribution shape. This is in part expected because driving variables of solar radiation and air temperature are both simulated to follow normal distributions. On the other hand, the smaller the amplitude of HI (maximum simulated value of HI minus minimum simulated value of HI), the smaller the resulting range of Wg values. Extremely high values (i.e., a right "tail" of the histogram) of maize grain productivity are likely to occur due to, for example, the least likely and least frequent combinations among optimal air temperature, more intense radiant regimes, and high harvest index values. On the contrary, a left tail is obtained. The grain productivity mode represents the central portion of the histogram and is derived from the most likely values of solar radiation and air temperature frequently combined throughout the maize cycle. Additionally, there is a possibility of simulating an extremely high value of solar radiation for a given day and simulating an extreme low value for the day after, which can be physically meaningless. This misrepresentation would be repaired by equation 25, considering that 10,000 pseudo-number values are enough.

Deterministic models may exhibit lots of errors through model design constraints, parameter uncertainty, errors in forcing data, and calibration requirements. Moreover, in models that require inputs such as air temperature or solar radiation, which can be obtained from weather stations, there can be errors associated with these measurements because of instrumentation limitations. This is because, for

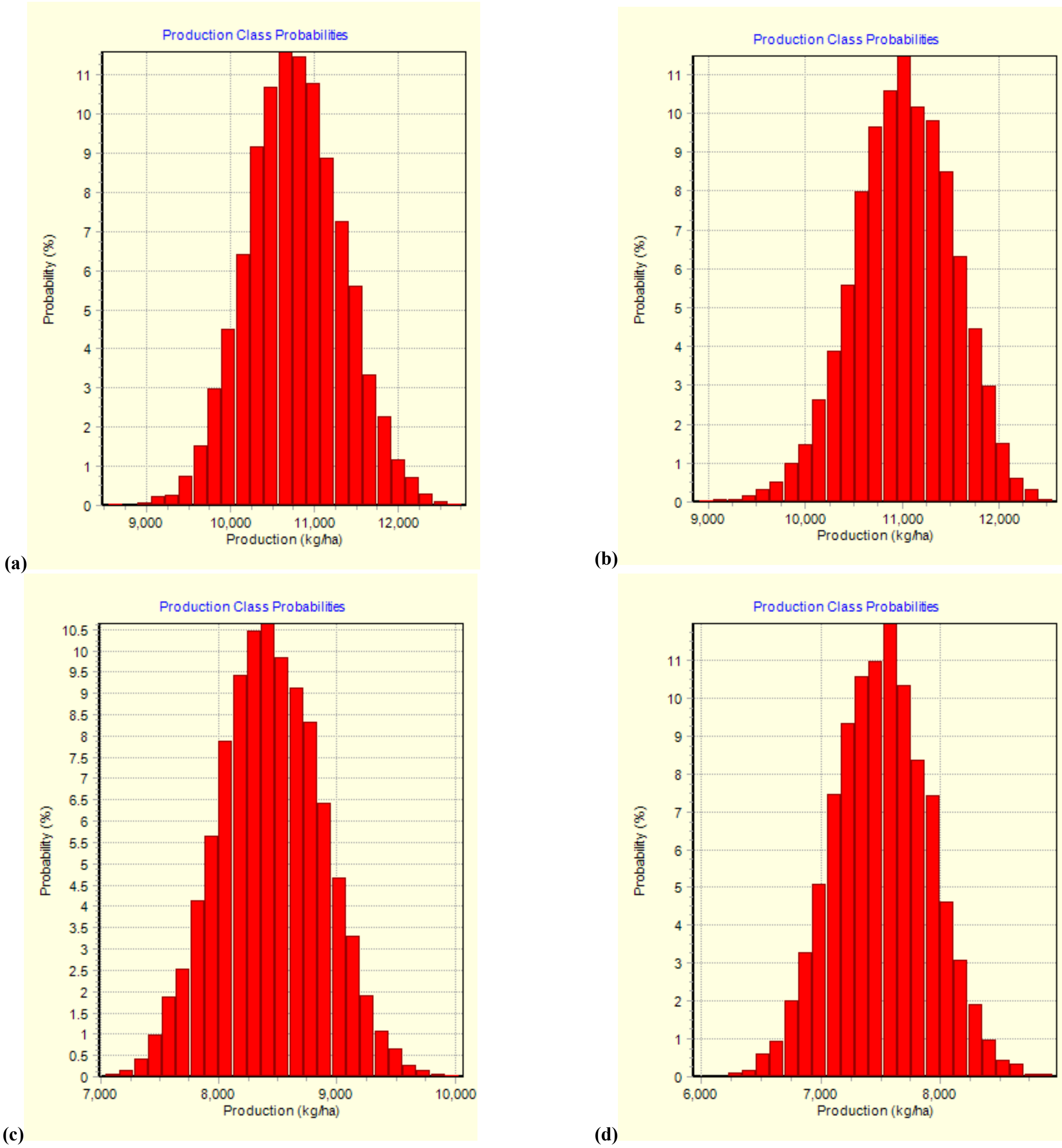

Figure 6. Maize grain productivity histograms considering that plant populations are sufficiently supplied with nitrogen: (a) bivariate normal distribution for solar radiation and air temperature under no-water-deficit conditions (case 1), (b) triangular distribution for solar radiation and air temperature under no-water-deficit conditions (case 2), (c) reduction of the plant population by taking D_{sow} = 5 plants m^{-1} under no-water-deficit conditions (case 3), and (d) consequence of using a water supply that is 75% of the optimal condition (Sw = 0.75; or 25% water deficit) (case 4).

example, the weather station is some distance away from the specific crop location or because different methods are used to log data and compute daily averages. One of the purposes of stochastic approaches is thus to statistically account for errors such as these.

General Considerations

One particularity of the herein presented model that uses a stochastic procedure is that it exhibits some similarities with Dutch generic growth models, which were properly approached by Ittersum et al. (2003), as carbon dioxide assimilation was taken into account and the prediction of phenological events was not considered. The model also brings some particularities with the family of specific models such as CERES-Maize and APSIM, as it omits respiration processes by adopting the radiation use efficiency to compute above-ground dry matter productivity rates. Moreover, the model relies on air temperature as the main variable to predict leaf area. The family of generic models assumes only a few parameters, most of them with physical or biological meaning, and it is not specific for any genotypes to capture photoperiod sensitivity and genetic potential in terms of kernel number and ear rows, as is done empirically by the specific models. Interestingly, the model presented in this work considers both empirical specificity and mechanistic features.

It is finally convenient to highlight that the depletion factor is only a simplification to express the reduction of productivity as a function of water deficit. In fact, if water availability is limited, then leaf area expansion is negatively affected and results in a decrease of light interception, which in turn decreases above-ground dry matter productivity rates. The best way to compute *Wpad* prior to *Wpa* would be to collate models such as those based on leaf area expansion as a function of water status (Reymmond et al., 2003) with those based on the variation of the light extinction coefficient as a function of leaf area.

In summary, the proposed deterministic model simulates realistic and robust values for maize grain productivity and for above-ground dry matter productivity rates and accumulation. Besides being considered satisfactory as the basis of the suggested stochastic model, this model is superior because it presents a range of outcomes that account for the probabilities of variable values and some implied errors associated with these. Through either a bivariate normal distribution or a triangular distribution for generating solar radiation and temperature values and through a triangular distribution for generating harvest index values, the stochastic model predicted similar values (i.e., mode) of grain productivity to those values observed in a field experiment. It is expected that some model improvements will be made as other genotypes are used to contribute to its evaluation in different sites, especially for some specific parameters.

CONCLUSIONS

The construction of a model with stochastic components was presented, and it was validated for two distant sites with similar climate conditions. The stochastic simulations output pointedly similar mode values to those obtained from a deterministic simulation, which itself was similar to the data observed from a field experiment with no water stress. The model can be satisfactorily used to simulate maize yield, although it should be improved if one is interested in simulating some specific above-ground dry matter data throughout the crop cycle or even in predicting some phenological stages of maize growth.

ACKNOWLEDGEMENTS

We acknowledge the following Brazilian Institutions: CNPq (National Council of Scientific and Technological Development) for providing a scholarship to the first author, and CNPQ/FAPESP (São Paulo State Scientific Foundation)/INCTEI (National Institute of Science and Technology in Irrigation Engineering) for financial support.

REFERENCES

Andrade, F. 1995. Analysis of growth and yield of maize, sunflower, and soybean grown at Balcarce, Argentina. *Field Crops Res.* 41(1): 1-12.

APSRU. 2007. Maize science documentation. Toowoomba, Queensland, Australia: Agricultural Systems Research Unit. Available at: www.apsim.info/Wiki/Maize.ashx. Accessed 14 May 2012.

Bonhomme, R. 2000. Basis and limits to using "degree-days" units. *European J. Agron.* 13(1): 1-10.

Box, G. E. P., and M. E. Muller. 1958. A note on the generation of random normal deviates. *Ann. Math. Stat.* 29(2): 610-611.

Cantelmo, J. F., and D. F. Ferreira. 2007. Desempenho de testes de normalidade multivariados avaliado por simulação Monte Carlo. *Ciência e Agrotec.* 31(6): 630-636.

Detomini, E. R. 2008. Ecophysiological attributes of hybrid Dkb-390 and stochastic model for predicting grain productivity. PhD diss. Piracicaba, Brazil: University of São Paulo, Luiz De Queiroz School of Agriculture. Available at: www.teses.usp.br/teses/disponiveis/11/11143/tde-15072008-140048/publico/eurodetomini.pdf. Accessed 10 October 2010.

Detomini, E. R., L. F. Massignan, M. S. Bernardes, and D. Dourado Neto. 2008. Light extinction coefficient for hybrid Dkb-390 under two levels of nitrogen supply. *Revista Brasileira Agrometeorol.* 16(2): 155-162.

De Wit, C. T. 1958. Transpiration and crop yields. *Versl. Landbouwk. Onderz* 64(6): 1-88.

Dobermann, A., T. Arkebauer, K. Cassman, R. Drijber, J. Lindquist, J. Specht, D. Walters, H. Yang, D. Miller, D. Binder, G. Teichmeier, R. Fergunson, and C. Wortmann. 2003. Understanding corn yield potential in different environments. Fresno, Cal.: Fluid Fertilizer Foundation.

Doorenbos, J., and A. H. Kassam. 1979. Yield response to water. Irrigation and Drainage Paper No. 33. Rome, Italy: United Nations FAO.

Dwyer, L. M., and D. W. Stewart. 1986. Leaf area development in field-grown maize. *Agron. J.* 78(2): 334-343.

Hair, J. F., R. E. Anderson, R. L. Tatham, and W. C. Black. 2005. *Análise Multivariada de Dados***.** 5th ed. Porto Alegre, Brazil: Bookman.

Heemst, H. D. J. 1986. Physiological principles. In *Modeling of Agricultural Production: Weather, Soils, and Crops*, 13-26. Wageningen, The Netherlands: Pudoc.

Hogg, R. V., and A. T. Craig. 1978. *Introduction to Mathematical Statistics*. 4th ed. New York, N.Y.: Macmillan.

Ittersum, M. K., P. A. Leffelaar, H. Keulen, M. J. Kropff, L. Bastiaans, and J. Goudriaan. 2003. On approaches and applications of the Wageningen crop models. *European J. Agron.* 18(3-4): 201-234.

Jones, C. A., and J. R. Kiniry. 1986. *CERES-Maize: A Simulation Model of Maize Growth and Development*. College Station, Tex.: Texas A&M University Press.

Laar, H. H., J. J. Goudriaan, and H. Van Keulen. 1992. Simulation of crop growth for potential and water-limited production situations: As applied to spring wheat. Simulation Report No. 27**.** Wageningen, The Netherlands: Center for Agrobiological Research (CABO-DLO).

Lizaso, J. I., W. D. Batchelor, K. J. Boote, and M. E. Westgate. 2005. Development of a leaf-level canopy assimilation model for CERES-Maize. *Agron. J.* 97(3): 722-733.

Loomis, R. S., and D. J. Connor. 1992. *Crop Ecology: Productivity and Management in Agricultural Systems.* Cambridge, U.K.: Cambridge University Press.

Mardia, K. V. 1975. Assessment of multinormality and the robustness of Hotelling's T^2 test. *J Royal Stat. Soc. C* 24(2): 163-171.

Matsumoto, M., and T. Nishimura. 1998. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comp. Simul.* 8(1): 3-30.

Mecklin, C. J., and D. J. Mundfrom. 2004. An appraisal and bibliography of tests for multivariate normality. *Intl. Stat. Rev.* 72(1): 123-138.

Meinke, H. 1996. Improving wheat simulation capabilities in Australia from a cropping perspective. PhD diss. Wageningen, The Netherlands: Wageningen Agricultural University.
Meinke, H., G. L. Hammer, H. Van Keulen, R. Rabbinge, and B. A. Keating. 1997. Improving wheat simulation capabilities in Australia from a cropping systems perspective: Water and nitrogen effects on spring wheat in a semi-arid environment. *European J. Agron.* 7(1-3): 75-88.
Monsi, M., and T. Saeki. 1953. Über den Lichtfaktor in den Pflanzengesellschaften und seine Bedeutung für die Stoffproduktion. *Japanese J. Botany* 14(1): 22-52.
Monteith, J. L. 1965. Light distribution and photosynthesis in field crops. *Ann. Botany* 29(1): 17-37.
Moore, D. S. 1995. *The Basis Practice of Statistics*. New York, N.Y.: W. H. Freeman.
Oguntunde, P. G., and N. Van De Giesen. 2004. Crop growth and development effects on surface albedo for maize and cowpea fields in Ghana, West Africa. *Intl. J. Biometeorol*. 49(2): 106-112.
Pachepsky, L. B., J. D. Haskett, and B. Acock. 1996. An adequate model of photosynthesis parameterization, validation, and comparison of models. *Agric. Syst.* 50(2): 209-225.
Passioura, J. B. 1996. Simulation models: Science, snake oil, education, or engineering? *Agron. J.* 88(5): 690-698.
Penman, H. L. 1948. Natural evaporation from open water, bare soil, and grass. *Proc. Royal Soc. London A* 193(1032): 120-145.
Reymmond, M., B. Muller, A. Leonardi, A. Charcosset, and F. Tardieu. 2003. Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiol.* 131(2): 664-675.
Royston, J. B. 1983. Some techniques for assessing multivariate based on the Shapiro-Wilk *W. Appl. Stat.* 32(2): 121-133.
Sinclair, T. R., and R. C. Muchow. 1999. Radiation use efficiency. *Advances in Agronomy* 65: 215-265.
Valentinuz, O. M., and M. Tollenaar. 2006. Effect of genotype, nitrogen, plant density, and row spacing on the area-per-leaf profile in maize. *Agron. J.* 98(1): 94-99.
Verdoodt, A., E. van Ranst, and L. Ye. 2004. Daily simulation of potential dry matter production of annual field crops in tropical environments. *Agron. J.* 96(6): 1739-1756.
Wonnacott, R., and T. Wonnacott. 1985. *Introductory Statistics*. 4th ed. New York, N.Y.: John Wiley and Sons.
Yang, J., and M. Alley. 2005. A mechanistic model for describing corn plant leaf area distribution. *Agron. J.* 97(1): 41-48.
Yin, X., M. J. Kropff, G. McLaren, and R. M. Visperas. 1995. A nonlinear model for crop development as a function of temperature. *Agric. Forest Meteorol.* 77(1-2): 1-16.
Zadoks, J. C., and R. Rabbinge. 1985. Modelling to a purpose. *Advances in Plant Pathology* 3: 231-244.