

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

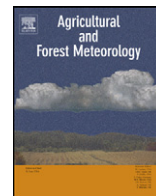
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformetCorrelation between temperature and phenology prediction error in rice (*Oryza sativa* L.)P.A.J. van Oort^{a,*}, Tianyi Zhang^b, M.E. de Vries^c, A.B. Heinemann^d, H. Meinke^{a,1}^a Wageningen University, Department of Plant Sciences, Centre for Crop Systems Analysis, PO Box 430, 6700AK Wageningen, The Netherlands^b LAPC, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China^c Wageningen University, Department of Plant Sciences, Plant Production Systems, PO Box 430, 6700AK Wageningen, The Netherlands^d Embrapa Arroz e Feijão, C.P. 179, 75375-000 Santo Antônio de Goiás, GO, Brazil

ARTICLE INFO

Article history:

Received 23 March 2011

Received in revised form 30 May 2011

Accepted 19 June 2011

Keywords:

Rice

Temperature

Phenology

Calibration

Sensitivity analysis

ABSTRACT

For rice (*Oryza sativa* L.), simulation models like ORYZA2000 and CERES-Rice have been used to explore adaptation options to climate change and weather-related stresses (drought, heat). Output of these models is very sensitive to accurate modelling of crop development, i.e. phenology. What has to date received little attention in phenology calibration is the temperature range within which phenological models are accurate. Particularly the possible correlation between temperature and phenology prediction error has received little attention, although there are indications that such correlation exists, in particular in the study by Zhang et al. (2008). The implication of such correlation is that a phenology model that is accurate within the calibration temperature range can be less accurate at higher temperatures where it can systematically overestimate or underestimate the duration of the phase from emergence to flowering. We have developed a new rice phenology calibration program that is consistent with ORYZA2000 concepts and coding. The existing calibration program DRATES of ORYZA2000 requires an assumption of default cardinal temperatures (8, 30 and 42 °C) and then calculates cultivar specific temperature sums and development rates. Our new program estimates all phenological parameters simultaneously, including the cardinal temperatures. Applied to nine large datasets from around the world we show that the use of default cardinal temperatures can lead to correlation between temperature and phenology prediction error and temperature and RMSE values in the order of 4–18 days for the period from emergence to flowering. Our new program avoids such correlation and reduces phenology prediction errors to 3–7 days (RMSE). Our results show that the often made assumption of a rapid decrease in development rate above the optimal temperature can lead to poorer predictions and systematic errors. We therefore caution against using default phenological parameters for studies where temperatures may fall outside the range for which the phenological models have been calibrated. In particular, this applies to climate change studies, where this could lead to highly erroneous conclusions. More phenological research with average growing season temperatures above the optimum, in the range of 32–40 °C, is needed to establish which phenological model best describes phenology in this temperature range.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The duration of the period from emergence to flowering and flowering to maturity plays an important role in yield formation of crops, including rice. A longer cycle potentially allows for higher yields, but can also increase exposure to stress and hence increase the risk of lower yields in certain environments (Dingkuhn, 1995; Matthews et al., 1997; Saseendran et al., 1998). Climate change

bears risks as well as opportunities for rice producers. For instance, temperature rises may enable regions that are currently too cold to shift from a single to a double rice crop (Matthews et al., 1997). Abiotic stresses such as drought, heat and cold can be avoided by choosing appropriate sowing dates and cultivars with the desirable duration of different development phases (Ekanayake et al., 1989; Matsui et al., 1997; Fukai et al., 1999; Saini and Westgate, 1999; Jagadish et al., 2007). Adaptation options to biotic and abiotic stresses, climate change and climate variability can be explored with crop simulation models. ORYZA2000 (Bouman et al., 2001; Bouman and Van Laar, 2006) and CERES-Rice (Alocija and Ritchie, 1988; Jones et al., 2003) are the most commonly used rice models and are quite similar in terms of processes included. Here we will focus on the ORYZA2000 model.

* Corresponding author. Tel.: +31 317 481357; fax: +31 317 485572.

E-mail address: Pepijn.vanoort@wur.nl (P.A.J. van Oort).¹ Current address: University of Tasmania, Tasmanian Institute for Agricultural Research (TIAR), Hobart TAS 7001, Australia.

In ORYZA2000 many parameters are development stage dependent: assimilate partitioning to organs, senescence of leaves, specific leaf area, the light extinction coefficient and more. Other processes do not depend directly on development stage, but are only active during specific development stages. For instance, spikelet sterility can only occur in the period from panicle initiation to flowering. Reduction of leaf expansion rate in case of drought stress (resulting in extra root growth) occurs only during the vegetative growth phase. Translocation of N from root, stem and leaves to storage organ occurs only in the period from flowering to maturity. Similar dependencies of crop physiological processes on development stage are found in many other models. It is therefore surprising that little attention has so far been given to better phenology calibration, given the importance of phenological processes for the overall performance of the models (Aggarwal and Mall, 2002; Confalonieri et al., 2010).

In previous, phenology-specific research scientists have developed and used their own calibration programs that simultaneously calibrate all parameters of their phenological models (Vergara and Chang, 1985; Summerfield et al., 1991, 1992; Collinson et al., 1992; Gao et al., 1992; Dingkuhn et al., 1995; Dingkuhn and Miezani, 1995; Yin, 1996; Yin et al., 1997b,c; Fukai, 1999). The phenological parameters estimated with these calibration programs cannot be directly fed into models like ORYZA2000 and CERES-Rice, due to conceptual and coding differences between the programs and existing phenology subroutines in ORYZA2000/CERES. To resolve this, tailor made calibration programs have been developed: DRATES and PARAM for ORYZA2000 and GENCALC (Hunt et al., 1993) for DSSAT models such as CERES-Rice. Apart from the conceptual and coding differences with existing work specifically on phenology, the number of parameters estimated differs. The tailor made programs mostly assume default values for most of the phenological parameters. In DRATES/ORYZA the default base, optimum and maximum temperature are 8, 30 and 42 °C. In CERES-Rice, most researchers use per default a base temperature of 9 °C and optimum temperature of 33 °C (e.g., Saseendran et al., 1998; Mall and Aggarwal, 2002; Yao et al., 2007). Although the use of these default values makes the task of calibration much simpler (only one parameter value needs to be estimated, namely the temperature sum), we will show that this often results in highly erroneous results.

The most extensive information on the accuracy of the ORYZA phenology routines is found in Matthews et al. (1995) and Mall and Aggarwal (2002). Both concluded that with the existing model phenology could be accurately simulated. Horie et al. (1995b) suggested that ORYZA's phenological subroutine did not work well, but provided no further details. It remains unclear if the problem lies in the model structure or in the parameterisation. Recent papers in which the ORYZA2000 model is used indicate no reason for concern, but they mostly fail to document the accuracy of phenology estimates and often lack a detailed description of data and methods used to calibrate phenology. Some provide no information at all on the phenology calibration (e.g., Sheehy et al., 2006; Krishnan et al., 2007), others explicitly avoid the topic by separately calibrating phenology for each site (e.g., Bouman and Van Laar, 2006; Boling et al., 2010). This is a feasible work-around and can be justified when the focus of a study is on something else than understanding phenology. But we are faced with a large knowledge gap about the accuracy of the phenological subroutine and parameters when applied in other sites (Sinclair and Seligman, 2000).

Some researchers have tried to separate random from systematic error, whereby systematic errors are errors correlated with an underlying variable. Correlations between phenology error and observed duration have been investigated more thoroughly in wheat (Xue et al., 2004) but not in rice. Plots of observed versus simulated duration of emergence to flowering in Mall and Aggarwal (2002) and Yao et al. (2007) suggest that such correlation also

occurs in rice. We found only one study that looked into correlation of phenological error with temperature. Zhang et al. (2008) showed that phenological errors increased when temperature increased. It is possible that a phenological model yields accurate unbiased predictions within the range of experiments and biased predictions outside this range (higher and lower temperatures). Surprisingly though, this correlation between temperature and phenological error has never been systematically investigated.

The studies cited above raise four questions:

1. Is correlation between temperature and phenological error as reported by Zhang et al. (2008) an exception or is this phenomenon also evident in other datasets?
2. How can we quantify correlation between temperature and phenological error?
3. Is correlation between temperature and phenological error caused by using the wrong phenological (temperature or day length) response model?
4. Is correlation between temperature and phenological error caused by using the wrong phenological parameter set?

The objectives of this paper are to: (1) clearly explain how rice phenology can be calibrated, (2) present a new program for rice phenology calibration that can be directly linked to the ORYZA2000 model and (3) use this improved program to answer the four questions raised above.

2. Materials and methods

First we discuss how (rice) phenology is normally modelled (Section 2.1), then we present the calibration program (Section 2.2) and finally we describe the datasets used for answering the four questions raised above (Section 2.3).

2.1. Modelling rice phenology

Development of rice depends most strongly on temperature and in some cultivars also on day length. Water and nutrient stress can affect development. They can delay or accelerate development relative to the non-limiting case. Their effect cannot be accurately simulated without accurate simulation of the non-limiting case. Here we will limit ourselves to the non-limiting case.

2.1.1. Temperature response functions

To simulate phenology it is assumed that there is a cultivar specific temperature sum, or thermal time, needed to complete a certain development stage. This temperature sum can be made day length dependent, which we have addressed in the subsequent section (Section 2.1.2). The temperature response function determines how temperature sum is calculated. The simplest (Blackman) equation is:

$$TI_t = \max(0, \min(T_t - TBD, TOD - TBD)) \quad (1a)$$

where TI_t (°C) is the increment in thermal time over time unit t , T_t is the temperature during t , TBD is the base temperature and TOD the optimal temperature. In this model there is no development below TBD, TI_t increases linearly with T_t up to the optimal temperature TOD and then remains at its maximum TOD – TBD. Temperature sum for development phase p ($TSUM_p$) is then the sum of daily TI_t values during this phase (in degree days, °Cd).

Development rates are calculated as the inverse of $TSUM_p$. Development stages are ranked on a numerical scale. In ORYZA2000 the scale runs from 0 at emergence, to 0.4 (start of photoperiod sensitive phase), 0.65 (panicle initiation), 1.0 (flowering) and 2.0 (maturity). If the temperature sum needed to complete the phase

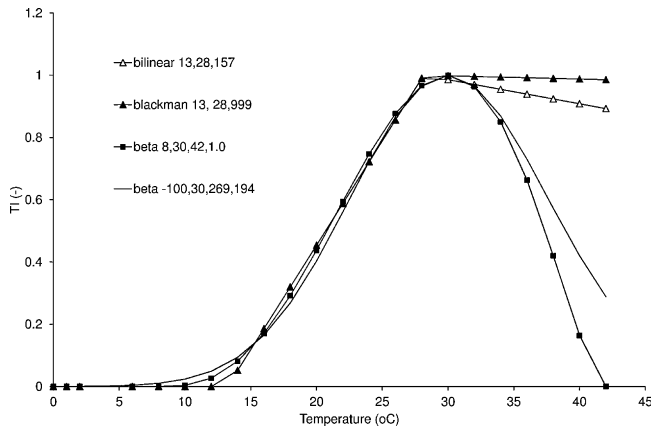


Fig. 1. Temperature response functions (function name: TBD, TOD, TMD, TSEN). As long as data are on the left side of the response function, say below 32 °C, the same shape can be obtained with any of the response functions, with very different values for TBD (beta with TBD = 8 °C or –100 °C, bilinear or Blackman with TBD = 13 °C) and with very different values for TMD (beta with TMD = 42 °C or 269 °C, bilinear with TMD = 157 °C or Blackman with TMD = 999 °C).

from panicle initiation to flowering is $TSUM_{PIFL}$, then the development rate for this phase is calculated as $(1.0 - 0.65)/TSUM_{PIFL}$.

The fact that the maximum thermal time accumulation in Eq. (1a) is $TOD - TBD$ makes comparisons among response functions of different parameter sets (TBD, TOD) complicated and also complicates comparisons of temperature sums calculated with different underlying TBD and TOD values. Eq. (1a) can be normalised to Eq. (1b) such that the minimum thermal time increment is always 0 at TBD and maximum thermal time increment is always 1 at TOD:

$$TI_t = \max\left(0, \min\left(1, \frac{T_t - TBD}{TOD - TBD}\right)\right) \quad (1b)$$

TI_t values can be calculated at any time step t , but are often conveniently aggregated to daily values in the range of 0–1. Let us assume $TBD = 8$, $TOD = 30$, $T_1 = 10$ and $T_2 = 40$ °C. This results in $TI_1 = 2/22 = 0.09$ and $TI_2 = 1$. Let us further assume that a development phase is completed within these 2 days. Now the duration of this phase under optimal temperatures is $OPTDUR = 0.09 + 1 = 1.09$ days and development rates are calculated similar as above but with $OPTDUR$ instead of $TSUM$.

The Blackman functions assume development rates remain at their maximum above TOD . In other response functions, including the one implemented in ORYZA2000, development slows linearly once an optimum value has been exceeded, coming to a complete standstill at maximum temperature TMD (Figs. 1 and 3c). The bilinear function is defined as:

$$TI_t = \max(0, T_t - TBD) \quad \text{for } T_t < TOD \quad (2a)$$

$$TI_t = \max\left(0, (TMD - T_t) \frac{TOD - TBD}{TMD - TOD}\right) \quad \text{for } T_t \geq TOD \quad (2b)$$

This bilinear model can also be normalised:

$$TI_t = \max\left(0, \frac{T_t - TBD}{TOD - TBD}\right) \quad \text{for } T_t < TOD \quad (2c)$$

$$TI_t = \max\left(0, \frac{TMD - T_t}{TMD - TOD}\right) \quad \text{for } T_t \geq TOD \quad (2d)$$

The Blackman response function can be seen as a special case of the bilinear model in which TMD is set to infinity. Type 1 (Eqs. (1a) and (1b)) and type 2 (Eqs. (2a) and (2b)) response functions are most commonly used: CERES-Rice uses 1a (with some modification for soil temperature during early growth); SIMRIW (Horie et al., 1995a) is a variant of type 1 with a logistic instead of linear response; ORYZA2000 uses 2a/b (Bouman and Van Laar, 2006); the Rice Clock Model (Gao et al., 1992) and the beta model (Yin, 1996;

Yin et al., 1995, 1997a,b,c) use bell shaped versions of the type 2 bilinear model. Among many temperature response models, the beta function has proven to be slightly yet consistently more accurate than other response models (Yin et al., 1997b), also in other crops (e.g., Ceglar et al., 2011). The normalised beta model is:

$$TI_t = \left\{ \left(\frac{T_t - TBD}{TOD - TBD} \right) \left(\frac{TMD - T_t}{TMD - TOD} \right)^{((TMD - TOD)/(TOD - TBD))^{TSEN}} \right\} \quad (3)$$

With a parameter $TSEN$ that determines the curvature of the bell-shape, see Fig. 1.

2.1.2. Rice, day length and transplanting

The descriptions in previous section are generally applicable for any crop. Now let us look in more detail into rice phenology. In ORYZA2000 the following four phases are distinguished:

1. BVP is *basic vegetative phase*, temperature dependent, from emergence (E) to end of BVP (=SPSP, start of PSP), development stage (DVS) 0–0.4, with development rate DVRJ.
2. PSP is *photoperiod sensitive phase*, temperature and day length dependent, from SPSP (DVS 0.4) to Panicle Initiation (PI, DVS 0.65), with development rate DVRI.
3. PPP is *post PSP phase*, temperature dependent, from panicle initiation (PI, DVS 0.65) to flowering (FL, DVS 1.0), with development rate DVRP.
4. GFP is *grain filling phase*, temperature dependent, from flowering (FL, DVS 1.0) to maturity (M, DVS 2.0), with development rate DVRR.

For any day in a simulation, the development stage on a continuous scale from 0 to 2 is simulated, based on temperature, day length and the abovementioned development rates. In case of cultivars that are not photoperiod sensitive, BVP and PSP can be treated as one phase, although there is indication that cardinal temperatures ($TOD, TSEN$) during BVP and PPP are different from those during PSP (Yin et al., 1997c), also for cultivars such as ShanYou 63 that are only mildly photoperiod sensitive. In case of photoperiod sensitivity, the start of PSP (parameter SPSP) needs to be derived indirectly because there are no means of direct measurement in the field (Vergara and Chang, 1985; Yin, 1996). The simplest (linear) day length model is:

$$EFP = 1 \quad \text{for } DL < MOPP \text{ and in phases other than PSP} \quad (4a)$$

$$EFP = \max(0, PPSE(DL - MOPP)) \quad \text{for } DL > MOPP \text{ during PSP} \quad (4b)$$

where EFP is the effect of day length (range 0–1), $PPSE$ the photoperiod sensitivity parameter and $MOPP$ the maximum optimum photoperiod. Development is delayed if during PSP day length DL is larger than $MOPP$. During the PSP, thermal increment (TI_t as defined above) is multiplied with EFP . An additive model or a model taking the minimum of EFP and (normalised) daily TI_t is also possible, but uncommon. Whether or not the additive or the multiplicative model ($EFP \times TI_t$) results in better simulations remains to be tested. EFP is calculated on a daily time step and can be multiplied with TI_t regardless of the step size t . The interaction between temperature effect and photoperiod effect can hamper calibration of temperature and photoperiod parameters. Let us define the daily photo-thermal time increment on a given day t as $PTI_t = EFP \times TI_t$. This approach is prone to compensating errors: if EFP is underestimated due to a low estimate of $PPSE$, this can be compensated by a higher TI_t value. This can result in multiple parameter sets of $TBD, TOD, TMD, MOPP, PPSE$ that all simulate the same PTI values. Because of this ambiguity, it is impossible to tell, which of these parameters reflects the true (physiological) parameter values of a cultivar. Interaction can be excluded in greenhouse/climate chamber experiments (Yin, 1996) with same temperature at different day lengths, or same day length at different temperatures.

In field experiments such interactions cannot be avoided. Further, photoperiod and temperature are continuously changing, further complicating the analyses of field experiments. Hence, much more experimental data are needed to unravel the effects of temperature and day length in the field (Dingkuhn and Miezán, 1995; Dingkuhn et al., 1995; Horie et al., 1995a).

Instead of the linear photoperiod response functions (Eqs. (4a) and (4b)) researchers have also used logistic functions (Gao et al., 1992; Horie et al., 1995a; Yin, 1996; Yin et al., 1997a,b,c). Proper calibration of the logistic response requires one treatment with day length close to MOPP and at least two treatments with $DL > MOPP$. In the greenhouse calibration of a logistic response function requires at least 3 photoperiods: two close to MOPP and one much larger than MOPP. For example Yin et al. (1997a,b,c) used treatments of 10, 12.5 and 14 h at a constant temperature.

Rice can be direct seeded or transplanted. Transplanting is normally at around 15–35 days after emergence. In the ORYZA2000 model, it is assumed that upon transplanting development is halted for some time after transplanting. The duration of this period of no development is calculated as:

$$TSHCKD = SHCKD \times TSTR \quad (5)$$

where SHCKD is the transplanting shock parameter, TSTR the temperature sum (sum of T_{lt} from emergence to transplanting) at the day of transplanting and TSHCKD the time in degree days (or unitless in case of a normalised model) during which development is halted after transplanting. The default value for SHCKD in ORYZA2000 is 0.4. Consider an experiment in which rice is direct seeded and the temperature sum from emergence to flowering is calculated as $TSUMFL_0$. Consider further a second experiment in which rice is transplanted at day a , for this second experiment the temperature sum at flowering is calculated as $TSUMFL_a = TSUMFL_0 + TSHCKD_a$. Then SHCKD is the only unknown in the equation $TSUMFL_a = TSUMFL_0 + SHCKD \times TSTR_a$, thus $SHCKD = (TSUMFL_a - TSUMFL_0) / TSTR_a$. In case of transplanting at two different dates a and b , SHCKD can be calculated as:

$$SHCKD = \frac{TSUMFL_b - TSUMFL_a}{TSTR_b - TSTR_a} \quad (6)$$

As in the case of photoperiod sensitivity, interaction with other parameters hampers estimation of SHCKD. A too low value for SHCKD can be compensated by a too high value for TSTR, yielding the same TSHCKD in Eq. (5). A too high value for TSTR can be obtained by assuming a too high value TBD or with an appropriately chosen set of TBD, TOD and TMD. Thus an error in parameter SHCKD can cancel out an error in TBD TOD and/or TMD. There can exist multiple parameter sets TBD, TOD, TMD, SHCKD that give identical simulations of phenology and it may be impossible to tell which of these parameters reflects the true (physiological) parameter values of a cultivar. In addition, it has been suggested that the true value of SHCKD might vary considerably, depending on handling of the young plants during transplanting (T. Lafarge, IRRI, personal communication. To our best knowledge, there are no published studies explicitly targeted at estimating SHCKD.

In summary, we can define the number and type of parameters of a rice phenological model as follows:

- For a photo period insensitive cultivar: TBD, TOD, $TSUM_{BVP+PSP}$, $TSUM_{PPP}$, $TSUM_{GFP}$.
- For a photo period sensitive cultivar: TBD, TOD, SPSP, MOPP, EFP, $TSUM_{BVP}$, $TSUM_{PSP}$, $TSUM_{PPP}$, $TSUM_{GFP}$.
- In case of transplanting: as above +SHCKD.

The set of parameters becomes larger when TBD and TOD are different for day and night (Yin et al., 1997b,c), when for different

phases different TBD and TOD parameters are assumed (Yin et al., 1997a; Dingkuhn, 1995) and when using the bilinear (+TMD) or beta (+TSEN) response model. Below, we will describe conceptually how these parameters can be estimated, in Section 2.2 we describe our phenology calibration program.

2.1.3. Calibration, random and systematic error

The first step to calibrate the phenology of a given cultivar is to define in drawings or photographs how the different development stages can be recognised (for example see the protocol for rice of the international rice research institute IRRI,² Lancashire et al., 1991 or the BBCH protocol by Meier, 2001). Next, the cultivar is grown at different temperatures and/or day length, and the dates of emergence, panicle initiation, flowering and maturity are noted. Ideally experiments will include a wide range of environmental conditions, which can be obtained through sowing at different sites or sowing at the same site on monthly intervals. Once temperature, photoperiod and transplanting shock parameters are fixed, temperature sum (TSUM, °Cd) or shortest possible duration (OPTDUR, days) can be calculated. The resulting TSUM or OPTDUR will be different depending on assumptions made on the other parameter values. Ideally all parameters are estimated simultaneously, but in practice very often one or more of the parameters is assumed constant, an assumption that simplifies the estimation of the other parameters. Estimation of the TSUM or OPTDUR parameter is particularly easy when all but one parameters are assumed constant. This is indeed the common approach for model calibration. Commonly assumed parameter values are: TBD = 9 °C and TOD = 33 °C (in CERES-Rice) and 8, 30, 42 °C (in ORYZA2000). In the Beta model, Yin et al. (1996) per default assumed TBD = 8 °C and TMD = 42 °C, after which TOD, TSEN and OPTDUR were simultaneously estimated. Yin (1996) found TOD values ranging from 25 to 30 °C, TSEN equal to 1 during BVP and PPP, TSEN ranging from 0.6 to 5.6 during PSP and OPTDUR in the range 35–73 days. Gao et al. (1992) suggests 10–28–40 °C (TBD–TOD–TMD) for Japonica and 12–30–40 °C for Indica and, plus power function parameters that produce a bell shaped response curve. Dingkuhn and Miezán report base temperatures in the range of 10–15 °C, and TOD in the range of 23–31 °C, based on water temperatures.

The average TSUM (or OPTDUR) from a set of temperature/day length treatments $i = 1, n$ is used to simulate the duration of phases in days. Model accuracy can be calculated as the coefficient of variation in TSUM or OPTDUR. After simulating the duration of a phase the accuracy can be calculated as the error (E_{ip}) for each treatment i and from that root mean square error ($RMSE_p$):

$$E_{ip} = OBSDUR_{ip} - SIMDUR_{ip} \quad (7)$$

$$RMSE_p = \sqrt{\frac{1}{n-1} \sum_i E_{ip}^2} \quad (8)$$

where $OBSDUR_{ip}$ is the observed duration in days of phase p for treatment i , $SIMDUR_{ip}$ is the simulated duration and n the number of treatments. $RMSE_p$ in Eq. (8) combines systematic and random error in one metric. A novelty in this study is that we unambiguously separate random from systematic variations:

$$TSUM_{ip} = a_p + b_p \times TM_{ip} \quad (9)$$

$$E_{ip} = a_p + b_p \times TM_{ip} \quad (10)$$

where TM_{ip} is the average temperature during phase p in treatment i , parameter a_p is the random error and $b_p \times TM_{ip}$ the systematic error. Parameters a_p and b_p are calculated using linear

² www.knowledgebank.irri.org/extension/index.php/growthstages.

regression. Slope parameter b_p measures how much the phenology error in days increases with a 1 °C increase in TM_{ip} . The calculation is repeated for every parameter set, thus we can search for the parameter set with the lowest a_p , b_p and $RMSE_p$ (and ideally with b_p equal zero). Separation of random from the systematic error has been done previously in wheat (Xue et al., 2004) but not in rice. Xue et al. (2004) tested for systematic error as in Eq. (11):

$$MSE_s = \frac{1}{n} \sqrt{\sum_i ((a_p + b_p \text{OBSDUR}_{ip}) - \text{OBSDUR}_{ip})^2} \quad (11)$$

where parameters a_p and b_p are derived from linear regression as in Eq. (12):

$$\text{SIMDUR}_{ip} = a_p + b_p \text{OBSDUR}_{ip} \quad (12)$$

Thus Xue et al. (2004) tested if phenological error of phase p is correlated with observed duration of p . Plots of simulated versus observed duration in Mall and Aggarwal (2002) and in Yao et al. (2007) suggest that such correlation can also occur in rice. Xue's approach does not identify the source of systematic error. Where the phenological model is driven by temperature, it seems logical to directly go to the plausible cause of bias, as we do in Eq. (10).

We have shown in Section 2.1.2 that the temperature and photoperiod response models are non-linear and that an error in the estimated value of one parameter can be compensated by an error in another parameter. Due to these non-linearities and interactions there can be local optima. If an optimisation program is used to estimate parameters (by minimising RMSE) we must ensure that it converges to a global (and not a local) optimum. We also need to establish whether multiple and very different parameter sets exist that produce identical simulations of duration of development stages. If in all simulations a single set of default parameters yields near optimal RMSE values then that can be a valid reason to retain the default set. Reduction of the number of parameters to be estimated at the expense of a small loss of accuracy may be very well acceptable. The phenology model that we have developed (Section 2.2) allows to investigate not only the optimal solutions but also the near optimal solutions, so that users can assess if there is clear convergence to a unique parameter set. Failure to converge generally indicates that more experimental data across a wider set of environmental conditions are needed.

2.1.4. What temperature, what day length

Temperature and day length playing an essential role in phenological development. Hence, it is important to consider how these two variables are defined and measured. Normally we use air temperature at 2 meters height in an open field, measured at a nearby meteorological station. In cool environments standing water temperature can be consistently higher than air temperature, with beneficial effects on yield (Confalonieri et al., 2005; Shimono et al., 2007b). In hot environments water temperature can have a (positive) cooling effect for crop production (Dingkuhn, 1995; Lobell and Bonfils, 2008). Studies have shown that accuracy of phenology modelling can be improved by inclusion of water temperatures in the calculations (Collinson et al., 1995; Dingkuhn et al., 1995; Shimono et al., 2007a). This leads to calibrated TBD, TOD and TMD values that differ from those calibrated from air temperature. If water temperature is systematically lower than air temperature then optimisation of parameters will show that $TBD_{\text{water}} < TBD_{\text{air}}$. Parameter estimates are therefore contingent on the type of temperature used as input. Water temperature is rarely measured. Models for simulating water temperature exist (Dingkuhn et al., 1995; Confalonieri et al., 2005; Kuwagata et al., 2008) but have to our knowledge not yet been validated outside the environment in which they were calibrated. They may require additional data

(such as temperature of inflowing irrigation water) which are generally also not known. A water temperature model also challenges the assumption normally made that growth and development are independent. Air temperature³ is not affected by growth of the plant and so we can assume that growth does not affect development. Water temperature depends on leaf area index (LAI) thus on growth. To calibrate water temperature driven phenology, additional measurements of water temperature and LAI are needed. A model that poorly simulates LAI will also poorly simulate water temperature and the resulting water temperature driven phenology. The theoretical accuracy gained by using water temperature may well be lost again due to uncertainties in the estimation of water temperature. Therefore, we will in this study restrict discussion to air temperature alone.

Another issue is whether daily average temperature should be used or hourly temperature. Daily mean temperature (TM) and hourly temperature (TH) are both calculated from daily minimum and maximum temperature (TMIN and TMAX), given that hourly measurements are rarely available. In ORYZA2000 the following equations are used:

$$TM = 0.5 \times (TMAX + TMIN) \quad (13)$$

$$TH = TM + 0.5 \times (TMAX - TMIN) \times \cos(0.2618 \times (h - 14)) \quad (14)$$

where h is hour and it is assumed that peak temperature TMAX is reached at 14.00 h. Both equations implicitly assume a day length of 12 h, which is valid for only 2 days of the year and with errors disproportionally increasing with the spatial distance from the equator and temporal distance for the equinoxes. At higher latitudes these equations will underestimate daily mean temperature during summer (when daytime is longer than 12 h) and overestimate daily mean temperature during winter (when night time is longer than 12 h). Slightly more complex models can take this into account (Parton and Logan, 1981; Goudriaan and van Laar, 1994; Ephraïm et al., 1996) and are implemented in some models (APSIM; Keating et al., 2003) but require additional parameters that need to be calibrated based on measured hourly temperatures. With temperature response functions as in Eqs. (1)–(3), the daily thermal time increment (TI_t) can be different depending on whether it is calculated with TM or TH. Few researchers have quantified the consequences of using daily average rather than hourly temperatures. Purcell (2003) has shown that differences are very small, except when temperature is for a long time close to TOD. Rice is often grown at temperatures close to TOD so we decided to work with hourly temperatures as already implemented in ORYZA2000. The use of hourly temperatures and determination of the duration of daytime and night time also becomes relevant when, as Yin (1996), Yin et al. (1997a,b,c) shows, day/night temperature parameter values differ markedly.

For day length, an additional concern is the exact definition of threshold light intensities. When the sun is just below the horizon there is still some light, declining to zero as the sun drops further below the horizon. Sun angles of 0° to –6° are found in the literature, resulting in 0 to 1.5 h of twilight, depending also on day of the year and latitude. Strictly speaking in calibrating phenology from field experiments we must also estimate the sun angle at which twilight still affects day length. Yin et al. (1997b) have calibrated phenology under different sun angles and found only small differences in terms of accuracy. This suggests that for prediction purposes, consistency on methods used is more important than the exact value of sun angle. Consider a set of phenological parameters calibrated with

³ Here we mean air temperature at 2 m height in the open field; of course the air temperature inside the canopy does depend on the LAI.

a sun angle of -2° . Now feed the MOPP, PPSE and other calibrated parameters into a model that uses a sun angle of -6° , i.e. with systematically longer days. In that case, the effect of photoperiod (EFP) would be systematically overestimated. This example shows the importance of having transparency and consistency in modelling approaches and phenological calibration programs.

Day length can easily be calculated from latitude and day of year (with an assumption on twilight sun angle), the equations are documented, for instance, in Parton and Logan (1981) and Goudriaan and van Laar (1994) and can be found in subroutine SASTRO of the ORYZA2000 model (which can be downloaded from the IRRI website). In the current implementation of ORYZA2000 a simpler approach is followed: effective day length is calculated as the day length at a sun angle of 0° plus 0.9 h. This corresponds roughly with a sun angle of -6° .

2.2. Rice phenology calibration program

Design criteria for the new calibration program were:

1. Compatibility with existing ORYZA2000 concepts and code;
2. Inclusion of temperature sensitivity, photoperiod sensitivity and transplanting shock as stated above;
3. Inclusion of the Blackman (Eq. (1a) and (1b)), bilinear (Eqs. (2a), (2b), (2c) and (2d)), and beta (Eq. (3)) response functions, both for temperature and day length. For the beta response function we use the model presented in Yin et al. (1997b), with different TOD and TSEN values during day and night, with one set of TOD, TSEN values for the BVP and PPP phase and a different set for the PSP phase;
4. Ability to treat the BVP and PSP as one phase in case of non-photoperiod sensitive cultivars and as separate phases in case of photoperiod sensitive cultivars;
5. In the photoperiod sensitive mode also statistically infer the thermal time SPSP at which BVP ends and PSP starts;
6. Use normalised TI_t to allow for easy comparison among temperature response functions and within the same response function across different parameter sets. Note that the existing ORYZA2000 model uses a temperature sum concept. After a small change in the subroutine SUBDD of ORYZA2000 our development rates can be fed directly into ORYZA2000. Development stage dependent parameters can be estimated with the existing DRATES program (also with modified SUBDD) after having calibrated phenological parameters;
7. Calculation of both $RMSE_p$ and slope parameter b_p as in Eqs. (8) and (10);
8. Ability for users to assess possible convergence to a unique optimal parameter set. This is achieved by looping through all possible (user defined) parameter sets and calculating summary statistics for each of these sets;
9. Ability to mimic water temperature in the simplest possible way (for demonstration purposes only), namely by assuming that it is consistently $TM_CORR^\circ C$ degrees higher or lower than air temperature. We do not know water or canopy temperature and have no intent of estimating it. The sole purpose of this TM_CORR approach is to illustrate the effect of consistently lower (or higher) water temperature on parameter estimates.

The program needs as input a text file with dates of emergence, transplanting, panicle initiation, flowering and maturity. This phenology input file can contain multiple treatments, each in a separate row. The program needs as input annual weather files (structured in the same way as in ORYZA2000 weather files) with latitude (for calculating day length), station number, day numbers and for each day minimum and maximum temperature. The user

selects the response function to be calibrated and specifies minimum value, maximum value and step size for each parameter. The Blackman response function (Eqs. (1a) and (1b)) is practically obtained by selecting the bilinear model and setting in the TMD to a very high value, in our analyses $999^\circ C$. For direct seeded rice the parameter SHCKD (Eq. (5)) is set to zero. In simulations in which no photo-period sensitivity is assumed the value of PPSE (Eq. (4b)) is set equal to zero and just one level for MOPP is chosen. In the current implementation, the same TBD, TOD and TMD values (and TSEN if the beta function is used) are assumed for each developmental phase. This is the simplest possible approach, whether this simplification is justified we will see in the results.

The program assumes that the duration of each phase p under optimal temperature ($OPTDUR_p$) is a constant parameter. It generates sets of parameters (TBD, TOD, etc.) and for each set it searches for the value of $OPTDUR_p$ such that the average error of duration of the phases from emergence to PI, emergence to flowering and flowering to maturity is less than 1 day (i.e. $\frac{1}{n} \sum_i (SIMDUR_{ip} - OBSDUR_{ip}) < 1$, where $i=1$, n is the number of treatments). This set-up corresponds to the standard approach to averaging of $TSUM_{ip}$ or averaging of development rates. This set-up does not exclude parameter sets with slope parameter b_p (Eq. (10)) differing strongly from 0. Thus, we can assess possible bias caused by the assumption of default ORYZA2000 parameters. Looping through all possible parameters is computationally demanding (can take a several hours) depending on the number of parameter sets evaluated and the response function chosen. It does however give the user maximum freedom in later selecting his or her ideal parameter set and it allows the user to check if there is clear convergence to a unique optimal parameter set and if there is interdependency between model parameters.

The program generates two output files. One listing for each parameter set and for each treatment the simulated and observed duration of the development phases. The second (summary) output file lists for each parameter set the $RMSE_p$ (Eq. (8)) and parameter b_p (Eq. (10)). In the results section we will report these as "Error \times TMEFL" for slope b_p during the phase from emergence to flowering, $RMSE_EFL$ for the root mean square error of this phase. Likewise, "Error \times TMFLM" and $RMSE_FLM$ are for the phase from flowering to maturity. The summary file can be sorted by $RMSE_EFL$ to obtain the parameter set with the highest accuracy for the phase from emergence to flowering or sorted by $RMSE_FLM$ to obtain the parameter set with the highest accuracy for the phase from flowering to maturity. By plotting $RMSE_EFL$ (y-axis) against parameter TBD (x-axis) we can see if there is clear convergence towards an optimal value of TBD. The same plots can also be made for the other parameters.

2.3. Data

The program was applied to large datasets from a wide range of environments. Table 1 lists all data used. Table 2 lists the range of observed durations of development phases of the nine cultivars. In all experiments, the IRRI protocol for identification of phenological events was used. The Chinese and Senegalese datasets contained records of emergence day (EMD), date of transplanting, date of panicle initiation (PI), date of flowering (FL) and date of maturity (M). For the Brazilian datasets aerobic rice was used that was not transplanted, emergence day and flowering day were recorded on all sites, panicle initiation and maturity were only recorded in two treatments. For each dataset weather data were obtained from a nearby official weather station.

Table 1
Experiments.

Cultivar	Location	Time	Average temperature (°C) during period from emergence to flowering	Reference
Qiuguang (QG)	China, Tonghua station (41.67N, 125.75E)	1985–2005, sown 1 × per year	17.8–20.8; temperature rise of 1.5 °C during the 21 years	Zhang et al. (2008)
Shanyou-63 (SY)	China, Zhongxiang station (112.3N, 31.1E)	'89–'00, sown 1 × per year	24.6–27.0	Zhang et al. (2010)
	China, Gushi station (115.4N, 32.1E)	'98, '99, sown 1 × per year	25.0–25.3	
	China, Liu'an station (116.3N, 31.5E)	'95–'97, sown 1 × per year	25.4–26.2	
	China, Hengyang station (112.4N, 26.5E)	'96–'02, sown 1 × per year	27.4–30.2	
	SY combined dataset	21 treatments	24.6–30.2	
BRSPrimavera (Bra.P)	Brazil, Sto Antônio de Goiás (–16.4N, –49.2E)	'04–'09, multiple sowings per year	24.1–24.8	Heinemann et al. (2009, 2011), Lorençoni et al. (2010)
	Brazil, Sinop (–12.0N, –55.6E)	'08–'09, sown 1 × per year	24.9–25.5	
	Brazil, Vilenha (–12.73N, –60.15E)	'08–'09, sown 1 × per year	25.1	
	Brazil, Gurupi (–12.47N, –49.18E)	'08–'09, sown 1 × per year	27.0–27.6	
	Brazil, Teresina (–5.08N, –42.8E)	'07–'10, sown 1 × per year	26.1–28.5	
	Primavera combined dataset	24 treatments	24.1–28.5	
BRSMG Curinga (Bra.C)	Same locations as Primavera	21 treatments	24.1–28.5	Heinemann et al. (2009, 2011)
IR64	Senegal, Fanaye (16.53N, –15.18E)	'05–'06, 15 sowing dates at 30 days interval; 1 treatment never reached maturity	24.2–31.3	de Vries et al. (2011)
	Senegal, Ndiaye (16.18N, –16.25E)	'05–'06, 15 sowing at 30 days interval	24.8–29.3	
IR32370	Same locations as IR64	Same treatments as IR64	24.0–31.3	de Vries et al. (2011)
ITA344	Same locations as IR64	Same treatments as IR64	24.5–31.4	de Vries et al. (2011)
SAHEL	Same locations as IR64	Same treatments as IR64	23.5–31.4	de Vries et al. (2011)
WAS161	Same locations as IR64	Same treatments as IR64	24.2–31.3	de Vries et al. (2011)
Total: 9 cvs	12 locations	232 treatments	17.8–31.4 °C	

3. Results and discussion

The four questions raised in the introduction can be answered as follows:

1. Correlation between temperature and phenological prediction error as reported by Zhang et al. (2008) for cultivar Qiuguang (QG) is not unusual. It was also found in the Chinese SY cultivar and the five Senegal cultivars. However the direction was

different, positive correlation in QG, negative in SY and negative in the five Senegalese cultivars. As a consequence of assuming default ORYZA2000 parameters, duration of emergence to flowering was underestimated in QG at high temperatures (Fig. 2). For cultivars SY and the five Senegalese cultivars we found that at high temperatures the default parameter set would overestimate the duration of the phase of emergence to flowering. This is illustrated in Fig. 3a–c for IR64 (the same pattern was found in the other cultivars). At low temperatures around 24 °C the error

Table 2
Observed duration of developmental phases.

	Emergence to panicle initiation (days)	Panicle initiation to flowering (days)	Emergence to flowering (days)	flowering to maturity (days)
Qiuguang (QG)	69–87	18–33	94–115	35–44
Shanyou-63 (SY)	59–85	18–33	83–118	24–49
Primavera (Bra.P)			63–94	
Curinga (Bra.C)			55–85	
IR64	59–110	23–35	85–140	19–37
IR32370	54–107	23–40	78–135	23–48
ITA344	67–118	23–56	93–156	15–34
SAHEL	53–118	21–29	80–135	24–38
WAS161	59–111	21–31	84–142	16–37

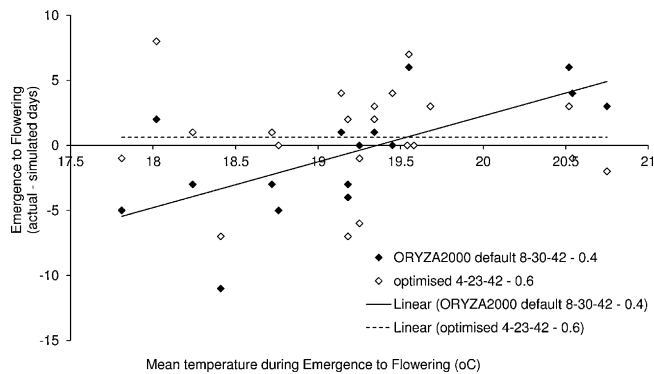


Fig. 2. Error in duration of emergence to flowering for cultivar Qiuguang, simulated with default ORYZA2000 parameters and with parameters optimised such that there is no longer a trend with temperature. The sequence 8–30–42–0.4 means TBD = 8 °C, TOD = 30 °C, TMD = 42 °C and SHCKD = 0.4.

is caused by the fact that default TBD = 8 °C is too low, at high temperatures (>30 °C) the error is caused by the fact that default TMD = 42 °C is too low.

- Correlation between temperature and phenological error can be quantified using Eqs. (8) and (10). We demonstrate this in Figs. 2 and 3a, b. The slope bp of Eq. (10) is shown in column “Error × TMEFL” in Table 3. “Error × TMEFL” indicates how much days the phenological error (emergence to flowering) increases with a unity increase in TMEFL (mean temperature from emergence to flowering). A value of zero for “Error × TMEFL” means no correlation between error and temperature. A negative error means that at high temperatures, simulated duration of this phase is longer than observed duration (Fig. 3a). The cause of this correlation can be identified from comparing the default temperature response function with the optimised temperature response function (Fig. 3c). Apparently for IR64, the default parameter set overestimates development rate at temperatures below 30 °C and underestimates development rate at temperatures above 30 °C.
- Within our range of temperatures, correlation between temperature and phenological error is not caused by using the wrong phenological response model. With temperatures largely in the left side part of the temperature response function, say 32 °C and lower, almost the same shape response function can be obtained with any response model, simply by properly tuning the parameters (see Fig. 1). Just slightly above the optimum the beta and Blackman function will predict similar development (Fig. 1). Our results, with average temperatures from emergence to flowering in the range of 17.8–31.4 °C clearly indicated that the assumption of a rapid decline in development rate above the optimum (bilinear function) gives poor predictions. With the current range of temperatures no meaningful difference in accuracy of the beta and Blackman function could be found (Table 3). It is however clear from Fig. 1 that with average temperatures from emergence to flowering of around 34 °C and higher, the two functions will give very different results. With lack of data on the shape of response functions at the very high temperatures, we caution against using untested models for climate change scenarios and call for more field experiments at high temperatures.
- Correlation between temperature and phenological error is in many cases caused by using the wrong phenological parameter set. In the Brazilian datasets the ORYZA2000 default values happened to be valid, but in the Chinese and Senegalese datasets using the default values lead to bias. It is clear from our results is that the sharp decrease in development rate from TOD = 30 °C to TMD = 42 °C in ORYZA2000 leads to overestimation of the duration of emergence to flowering. A simpler model with minimal

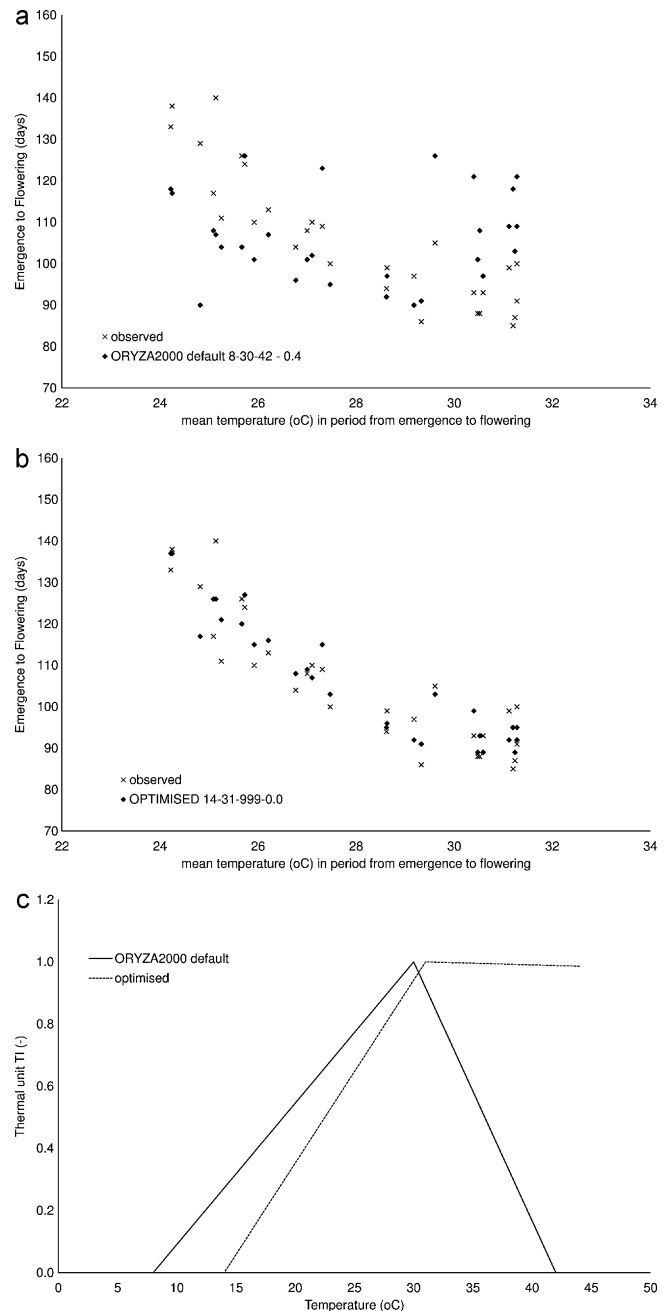


Fig. 3. (a) Observed and simulated duration of emergence to flowering for cultivar IR64, simulated with default ORYZA2000 parameters. (b) Observed and simulated duration of emergence to flowering for cultivar IR64, simulated with optimised parameters. (c) Default and optimised temperature functions for IR64.

high temperature constraints results in a better fit (e.g. by setting TMD to an extremely high value such as 999 °C, i.e. a Blackman response function). For the Senegalese cultivars we found that, relative to the Blackman curve, small further increase in accuracy could be obtained with a bilinear model and TMD values somewhere in between 100 and 1000 °C (data not shown). In the QG dataset the mean temperature in the period from emergence to flowering was 17.8–20.8 °C (Table 1) thus the shape of the response function above TOD (22–25 °C, Table 3) could not be determined. In the QG dataset, the systematic error was caused by default values of TBD and TOD being too high (Table 3), resulting in development at lower temperatures being underestimated. However, the range of temperatures in the QG dataset was limited (Table 1, Fig. 2) questioning if this temperature range

Table 3

Optimised parameter sets and default ORYZA2000 parameter set. To fit the table to this page photoperiod parameters are left out; in most cultivars adding PSP did not improve results. The column “objective” indicates what was minimized: RMSE.EFL is the RMSE for the phase of emergence to flowering, RMSE.FLM the RMSE for the phase of flowering to maturity. Error \times TMEFL is the slope parameter b_p in Eq. (10), which is ideally zero.

Cultivar	Photoperiod sensitivity	Objective	Model	TBD	TOD	TMD	SHCKD	OPTDUR		E. to flowering		Fl. to maturity	
								EFL	FLM	Error \times TMEFL	RMSE	Error \times TMFLM	RMSE
QG	noPSP	Default	Bilinear	8	30	42	0.4	49	17	3.5	4.0	2.8	5.7
	noPSP	Error \times TMEFL	Bilinear	4	23	42	0.6	63	31	0.0	4.0	0.5	4.1
	noPSP	RMSE.EFL	Blackman	0	25	999	0.6	67	32	1.4	3.1	1.6	4.5
	PSP	RMSE.EFL	Blackman	2	22	999	0.4	41	34	0.7	2.6	1.4	3.4
	noPSP	RMSE.EFL	beta					62	24	1.1	3.6	0.3	4.6
	noPSP	RMSE.FLM	bilinear	2	19	42	0.8	61	32	−2.1	5.1	−0.8	2.0
SY	noPSP	Default	bilinear	8	30	42	0.4	67	23	−4.0	7.9	0.6	4.3
	noPSP	RMSE.EFL	bilinear	12	32	42	−0.2	73	20	−0.3	5.0	1.1	6.3
	PSP	RMSE.EFL	Blackman	12	28	999	0.8	44	24	−0.4	3.7	−0.3	4.9
	noPSP	RMSE.EFL	beta					76	21	−0.2	4.5	0.2	7.4
	noPSP	RMSE.FLM	bilinear	9	27	42	0.4	67	23	−5.1	9.0	−0.1	3.6
Bra.P	noPSP	Default	bilinear	8	30	42	0	55	31	−1.4	5.2	1.3	1.8
	noPSP	RMSE.EFL	Blackman	9	32	999	0	51	28	0.1	4.5	2.0	2.7
	PSP	RMSE.EFL	Blackman	6	28	999	0	59	34	−0.5	4.8	1.4	1.8
	noPSP	RMSE.EFL	beta					54	33	0.0	5.0	1.5	2.0
Bra.C	noPSP	Default	bilinear	8	30	42	0	62	30	−1.1	5.3	1.3	2.0
	noPSP	RMSE.EFL	bilinear	5	33	42	0	59	28	−0.4	4.9	1.7	2.4
	PSP	RMSE.EFL	Blackman	0	28	999	0	40	41	−0.1	4.4	0.9	1.4
	noPSP	RMSE.EFL	beta					56	24	0.1	4.8	1.3	2.0
IR64	noPSP	Default	bilinear	8	30	42	0.4	67	20	−5.9	17.4	−0.4	4.3
	noPSP	RMSE.EFL	Blackman	14	31	999	0	78	22	−0.1	6.0	1.3	4.8
	PSP	RMSE.EFL	Blackman	16	28	999	0	82	23	−0.1	6.1	0.9	4.7
	noPSP	RMSE.EFL	beta					56	14	0.0	6.3	1.1	5.7
	noPSP	RMSE.FLM	bilinear	4	35	42	0.2	71	21	−4.5	13.9	0.1	3.4
IR323	noPSP	Default	bilinear	8	30	42	0.4	65	22	−5.7	18.4	0.3	5.1
	noPSP	RMSE.EFL	Blackman	14	32	999	0	73	23	0.1	8.0	1.4	6.9
	noPSP	RMSE.EFL	beta					63	20	−0.1	7.4	1.4	6.9
	noPSP	RMSE.FLM	bilinear	8	28	42	0	70	22	−6.5	19.8	0.2	4.9
ITA344	noPSP	Default	bilinear	8	30	42	0.4	77	19	−6.3	17.6	0.4	5.3
	noPSP	RMSE.EFL	Blackman	18	26	999	0	95	22	0.4	6.2	2.1	7.0
	noPSP	RMSE.EFL	beta					83	20	0.6	5.6	1.3	7.2
	noPSP	RMSE.FLM	Bilinear	8	29	42	−0.2	85	19	−6.7	18.1	0.3	5.1
Sahel108	noPSP	Default	Bilinear	8	30	42	0.4	63	22	−5.4	17.9	−0.2	4.2
	noPSP	RMSE.EFL	Blackman	15	34	999	0.4	60	21	0.6	7.3	2.0	6.2
	noPSP	RMSE.EFL	Beta					62	17	0.0	9.5	0.7	10.5
	noPSP	RMSE.FLM	Bilinear	1	36	42	−0.2	74	23	−4.0	14.3	0.2	3.3
WAS161	noPSP	Default	Bilinear	8	30	42	0.4	68	20	−6.1	18.0	0.0	4.1
	noPSP	RMSE.EFL	Blackman	16	30	999	0	77	22	0.3	6.2	1.0	5.5
	noPSP	RMSE.EFL	Beta					64	18	0.1	5.5	0.9	5.3
	noPSP	RMSE.FLM	Bilinear	9	34	42	0.6	62	19	−4.3	14.2	0.2	3.8

is sufficiently large for calibrating phenological parameters. In the SY and Senegalese datasets, the low TBD default was too low, as a result of which simulated duration from emergence to flowering was too short at temperatures around 24 °C (Fig. 3a). In the SY and Senegalese datasets the default TMD was too low as a result of which simulated duration from emergence to flowering was too long at temperatures above 30 °C (Fig. 3a).

In addition to answering the above four questions, we found it interesting to note that:

- Existing phenology calibration programs (Summerfield et al., 1991, 1992; Collinson et al., 1992; Gao et al., 1992; Dingkuhn et al., 1995; Dingkuhn and Miezani, 1995; Yin, 1996; Yin et al., 1997b,c; Fukai, 1999) only minimised RMSE. We found that minimising RMSE (Eq. (8)) also leads to correlation (temp \times error) converging to zero (b_p in Eq. (10) converges to zero). In Table 3 this can be seen from comparing the “Error \times TMEFL” and RMSE values of the default and the optimised phenology. The fortunate

consequence is that these previous studies unintentionally also eliminated correlation of phenology error with temperature.

- Normally accuracy of a phenological model (+parameters) is reported in one metric (the RMSE) and shown in graphs with simulated duration on the y-axis and observed duration on the x-axis (e.g., Matthews et al., 1995; Mall and Aggarwal, 2002; Yao et al., 2007). Such a way of reporting conveys no information on the temperature range in which phenology was calibrated and conveys no information on the relation between error and temperature. We recommend plotting the observed and simulated duration of phase p on the y-axis versus average temperature during phase p on the x-axis, as in Figs. 2 and 3a,b. Such figures provide information on the temperature range used for calibration and the accuracies at different temperatures.
- Using this method to plot observed duration from flowering to maturity (Fig. 4) against mean temperature from flowering to maturity showed that this phase is much less responsive to temperature than the pre-flowering phase. We only found a significant relation between duration and temperature of this phase for cultivar SY. For the other cultivars almost the same accuracy

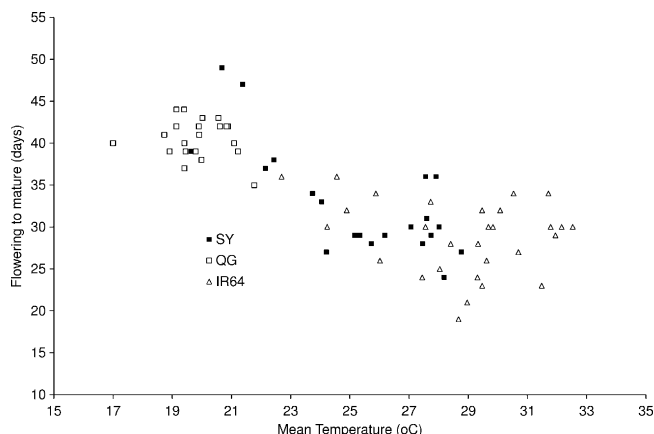


Fig. 4. Observed duration from flowering to maturity versus mean temperature during this period.

could be obtained simply by assuming the phase had a fixed number of days (for clarity we show in Fig. 4 only IR64, the same pattern was found for the other four Senegalese cultivars).

4. Our results consistently suggested that TBD during the post-flowering phase is markedly lower than in the pre-flowering phase (Table 3). This has also been shown for wheat (Wang and Engel, 1998). A lower TBD in our models means that below TOD (Fig. 1) the slope of the bilinear or Blackman model is less steep, which means that T_{fl} increases relatively little in case of a 1°C increase in temperature. This is in line with the above finding that the duration of the post-flowering phase is less responsive to temperature. The RIDEV model (Dingkuhn, 1995) uses a different approach for simulating pre- and post-flowering phenology of rice but from the paper it is not clear whether development rate is equally or less responsive to temperature compared with the pre-flowering phase. Models like ORYZA2000 and CERES-Rice assume the same temperature parameters for the pre- and the post-flowering phase. Our results and previous work by Dingkuhn (1995) suggest that it may be better to use different temperature parameters for the pre- and post-flowering phases. Our results and those of Dingkuhn (1995) further suggest that the duration of the post-flowering phase is less variable among cultivars. Additional analyses might show if there exists a set of cultivar independent temperature response parameters for the phase from flowering to maturity.
5. In some cultivars, our optimisations resulted in very high values of TBD in the pre-flowering phase: ITA344: 18°C , WAS161: 16°C and Sahel 108: 15°C (Table 3). We cannot tell from our calibration data what the shape of the response function looks like around $15\text{--}18^{\circ}\text{C}$ because temperature was never that low for a prolonged time in the Senegal experiments. What we do know is that within the range of temperatures in our experiments ($24.2\text{--}31.4^{\circ}\text{C}$, Table 1), development could accurately be modelled with these TBD values.
6. The effect of transplanting shock remains difficult to quantify. In most cultivars there was no clear convergence to a transplanting shock parameter greater than zero. In all cultivars there was a very flat response surface, i.e. almost the same accuracy could be obtained with values for SHCKD in the range of -0.2 to 0.6 . For IR64, IR323, ITA344 and WAS161 we found SHCKD = 0, for the Brazilian cultivars there was no transplanting, for SY we found SHCKD values from -0.2 to 0.8 depending on whether or not photoperiod sensitivity was assumed (an example of possible cancelling out of effects of different parameters). For Sahel 108 SHCKD converged to 0.4 , in QG SHCKD converged to 0.6 . The results imply that in terms of predictive power, there is little or

no value in including a transplanting shock parameter in a rice phenology model.

4. Conclusions

The use of existing default values for phenological parameters can lead to correlation between phenological error and temperature. Phenological parameters that give seemingly accurate predictions under current temperatures can be unexpectedly less accurate under higher or lower temperatures. This is relevant when using crop growth models for exploring potential for adaptation or intensification of cropping systems (e.g., double cropping) and in scenario development for climate change impacts. For pragmatic reasons model users often use the same default values for base, optimum and maximum temperature parameters and only estimate the temperature sum as a cultivar specific parameter. Major crop simulation models such as ORYZA2000 and CERES-Rice lack a tailor made phenology calibration program that allows for simultaneously calibrating all phenological parameters. We have developed such a new phenology calibration program consistent with ORYZA2000 code and concepts. Applied to nine large datasets (>20 treatments per dataset) from across the globe we have shown that the use of ORYZA2000's default base, optimum and maximum temperature parameter can lead to systematic phenology error (estimated number of days from emergence to flowering) at high and low temperatures. It became clear that to calibrate phenology it is necessary to simultaneously estimate all phenological parameters and not just the temperature sum. Assumptions on the shape of the response function above the optimum temperature led to very different simulated durations of development phases. We therefore caution against using untested models for climate change scenarios, when not calibrated in the range of temperatures considered in the climate change scenarios. This could lead to highly erroneous conclusions.

References

- Aggarwal, P.K., Mall, R.K., 2002. Climate change and rice yields in diverse agro-environments of India. II. Effect of uncertainties in scenarios and crop models on impact assessment. *Climatic Change* 52 (3), 331–343.
- Alocija, E.C., Ritchie, J.T., 1988. Upland rice simulation and its use in multicriteria optimization. In: International Benchmark Sites Network for Agrotechnology Transfer. Research Report Series. University of Hawaii, Hawaii, 96 pp.
- Boling, A.A., Tuong, T.P., van Keulen, H., Bouman, B.A.M., Suganda, H., Spiertz, J.H.J., 2010. Yield gap of rainfed rice in farmers' fields in Central Java, Indonesia. *Agricultural Systems* 103 (5), 307–315.
- Bouman, B.A.M., Van Laar, H.H., 2006. Description and evaluation of the rice growth model ORYZA2000 under nitrogen-limited conditions. *Agricultural Systems* 87 (3), 249–273.
- Bouman, B.A.M., Kropff, M.J., Tuong, T.P., Wopereis, M.C.S., ten Berge, H.F.M., M.C.S., H.H., 2001. ORYZA2000: Modeling Lowland Rice. International Rice Research Institute/Wageningen University and Research Centre, Los Baños, Philippines/Wageningen, The Netherlands.
- Ceglar, A., Črepinšek, Z., Kajfež-Bogataj, L., Pogačar, T., 2011. The simulation of phenological development in dynamic crop model: the Bayesian comparison of different methods. *Agricultural and Forest Meteorology* 151 (1), 101–115.
- Collinson, S.T., Ellis, R.H., Summerfield, R.J., Roberts, E.H., 1992. Durations of the photoperiod-sensitive and photoperiod-insensitive phases of development to flowering in four cultivars of rice (*Oryza sativa* L.). *Annals of Botany* 70 (4), 339–346.
- Collinson, S.T., Ellis, R.H., Summerfield, R.J., Roberts, E.H., 1995. Relative importance of air and floodwater temperatures on the development of rice (*Oryza sativa*). *Experimental Agriculture* 31 (2), 151–160.
- Confalonieri, R., Bellocchi, G., Tarantola, S., Acutis, M., Donatelli, M., Genovesi, G., 2010. Sensitivity analysis of the rice model WARM in Europe: exploring the effects of different locations, climates and methods of analysis on model sensitivity to crop parameters. *Environmental Modelling and Software* 25 (4), 479–488.
- Confalonieri, R., Mariani, L., Bocchi, S., 2005. Analysis and modelling of water and near water temperatures in flooded rice (*Oryza sativa* L.). *Ecological Modelling* 183 (2–3), 269–280.
- de Vries, M.E., Leffelaar, P.A., Sakané, N., Bado, B.V., Giller, K.E., 2011. Adaptability of irrigated rice to temperature change in Sahelian environments. *Experimental Agriculture* 47 (1), 69–87.

- Dingkuhn, M., 1995. Climatic determinants of irrigated performance in the Sahel – III. Characterizing environments by simulating crop. *Agricultural Systems* 48 (4), 435–456.
- Dingkuhn, M., Miezani, K.M., 1995. Climatic determinants of irrigated rice performance in the Sahel – II. Validation of photothermal constants and characterization of genotypes. *Agricultural Systems* 48 (4), 411–433.
- Dingkuhn, M., Sow, A., Samb, A., Diack, S., Asch, F., 1995. Climatic determinants of irrigated rice performance in the Sahel – I. Photothermal and micro-climatic responses of flowering. *Agricultural Systems* 48 (4), 385–410.
- Ekanayake, I.J., Datta, S.K.D., Steponkus, P.L., 1989. Spikelet sterility and flowering response of rice to water stress at anthesis. *Annals of Botany* 63 (2), 257–264.
- Ephraïm, J.E., Goudriaan, J., Marani, A., 1996. Modelling diurnal patterns of air temperature, radiation wind speed and relative humidity by equations from daily characteristics. *Agricultural Systems* 51 (4), 377–393.
- Fukai, S., 1999. Phenology in rainfed lowland rice. *Field Crops Research* 64 (1–2), 51–60.
- Fukai, S., Pantuwan, G., Jongdee, B., Cooper, M., 1999. Screening for drought resistance in rainfed lowland rice. *Field Crops Research* 64 (1–2), 61–74.
- Gao, L., Jin, Z., Huang, Y., Zhang, L., 1992. Rice clock model: a computer model to simulate rice development. *Agricultural and Forest Meteorology* 60 (1–2), 1–16.
- Goudriaan, J., van Laar, H.H., 1994. Modelling potential crop growth processes. In: *Current Issues in Production Ecology*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 239 p.
- Heinemann, A.B., Stone, L.F., Fageria, N.K., de Castro, L.M., 2009. Evaluation of physiological traits in upland rice for adaptation to no-tillage system. *Brazilian Journal of Plant Physiology* 21, 113–122.
- Heinemann, A.B., Stone, L.F., Nand, K.F., 2011. Transpiration rate response to water deficit during vegetative and reproductive phases of four upland rice cultivars. *Scientia Agrícola* 68, 24–30.
- Horie, T., Nakagawa, M., Centeno, H.G.S., Kropff, M.J., 1995a. The rice crop simulation model SIMRIW and its testing. In: Matthews, R.B., Kropff, M.J., Bachelet, D., Laar, H.H. (Eds.), *Modeling the Impact of Climate Change on Rice Production in Asia*. IRRI/CAB International, pp. 51–66.
- Horie, T., Nakagawa, M., Ohnishi, M., Nakno, J., 1995b. In: Matthews, R.B., Kropff, M.J., Bachelet, D., Laar, H.H. (Eds.), *Modeling the Impact of Climate Change on Rice Production in Asia*. IRRI/CAB International, pp. 143–164.
- Hunt, L.A., Pararajasingham, S., Jones, J.W., Hoogenboom, G., Imamura, D.T., Ogoshi, R.M., 1993. GENCALC: software to facilitate the use of crop models for analyzing field experiments. *Agronomy Journal* 85 (5), 1090–1094.
- Jagadish, S.V.K., Craufurd, P.Q., Wheeler, T.R., 2007. High temperature stress and spikelet fertility in rice (*Oryza sativa* L.). *Journal of Experimental Botany* 58 (7), 1627–1635.
- Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Ritchie, J.T., 2003. The DSSAT cropping system model. *European Journal of Agronomy* 18 (3–4), 235–265.
- Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N.G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J.P., Silburn, M., Wang, E., Brown, S., Bristow, K.L., Asseng, S., Chapman, S., McCown, R.L., Freebairn, D.M., Smith, C.J., 2003. An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy* 18 (3–4), 267–288.
- Krishnan, P., Swain, D.K., Chandra Bhaskar, B., Nayak, S.K., Dash, R.N., 2007. Impact of elevated CO₂ and temperature on rice yield and methods of adaptation as evaluated by crop simulation studies. *Agriculture, Ecosystems and Environment* 122 (2), 233–242.
- Kuwagata, T., Hamasaki, T., Watanabe, T., 2008. Modeling water temperature in a rice paddy for agro-environmental research. *Agricultural and Forest Meteorology* 148 (11), 1754–1766.
- Lancashire, P.D., Bleiholder, H., Langelüdde, P., Stauss, R., van den boom, T., Weber, E., Witzsenberger, A., 1991. A uniform decimal code for growth stages of crops and weeds. *Annals of Applied Biology* 119 (3), 561–601.
- Lobell, D.B., Bonfils, C., 2008. The effect of irrigation on regional temperatures: a spatial and temporal analysis of trends in California, 1934–2002. *Journal of Climate* 21 (10), 2063–2071.
- Lorençoni, R., Dourado Neto, C., Heinemann, A.B., 2010. Calibration and evaluation of the ORYZA-APSIM crop model for upland rice in Brazil. *Revista ciência agrônômica* 41, 605–613 (summary in English and text in Portuguese).
- Mall, R.K., Aggarwal, P.K., 2002. Climate change and rice yields in diverse agro-environments of India. I. Evaluation of impact assessment models. *Climatic Change* 52 (3), 315–330.
- Matsui, T., Namuco, O.S., Ziska, L.H., Horie, T., 1997. Effects of high temperature and CO₂ concentration on spikelet sterility in indica rice. *Field Crops Research* 51 (3), 213–219.
- Matthews, R.B., Kropff, M.J., Bachelet, D., Laar, H.H. (Eds.), 1995. *Modeling the Impact of Climate Change on Rice Production in Asia*. IRRI/CAB International, p. 289.
- Matthews, R.B., Kropff, M.J., Horie, T., Bachelet, D., 1997. Simulating the impact of climate change on rice production in Asia and evaluating options for adaptation. *Agricultural Systems* 54 (3), 399–425.
- Meier, U., 2001. Growth stages of mono- and dicotyledonous plants. BBCH Monograph. www.jki.bund.de/fileadmin/dam.uploads/_veroeff/bbch/BBCH-Skala.english.pdf.
- Paeton, W.J., Logan, J.A., 1981. A model for diurnal variation in soil and air temperature. *Agricultural Meteorology* 23 (C), 205–216.
- Purcell, L.C., 2003. Comparison of thermal units derived from daily and hourly temperatures. *Crop Science* 43 (5), 1874–1879.
- Saini, H.S., Westgate, M.E., 1999. Reproductive development in grain crops during. *Advances in Agronomy* 68 (C), 59–96.
- Saseendran, S.A., Hubbard, K.G., Singh, K.K., Mendiratta, N., Rathore, L.S., Singh, S.V., 1998. Optimum transplanting dates for rice in Kerala, India, determined using both CERES v3.0 and ClimProb. *Agronomy Journal* 90 (2), 185–190.
- Sheehy, J.E., Mitchell, P.L., Ferrer, A.B., 2006. Decline in rice grain yields with temperature: models and correlations can give different estimates. *Field Crops Research* 98 (2–3), 151–156.
- Shimono, H., Hasegawa, T., Iwama, K., 2007a. Modeling the effects of water temperature on rice growth and yield under a cool climate: I, model development. *Agronomy Journal* 99 (5), 1327–1337.
- Shimono, H., Hasegawa, T., Kuwagata, T., Iwama, K., 2007b. Modeling the effects of water temperature on rice growth and yield under a cool climate: II, model application. *Agronomy Journal* 99 (5), 1338–1344.
- Sinclair, T.R., Seligman, N., 2000. Criteria for publishing papers on crop. *Field Crops Research* 68 (3), 165–172.
- Summerfield, R.J., Collinson, S.T., Ellis, R.H., Roberts, E.H., Vries, F.W.T.P., 1992. Photothermal responses of flowering in rice (*Oryza sativa*). *Annals of Botany* 69 (2), 101–112.
- Summerfield, R.J., Roberts, E.H., Ellis, R.H., Lawn, R.J., 1991. Towards the reliable prediction of time to flowering in six annual crops. I. The development of simple models for fluctuating field environments. *Experimental Agriculture* 27 (1), 11–31.
- Vergara, B.S., Chang, T.T., 1985. The flowering response of the rice plant to photoperiod. In: *A Review of Literature*, 4th ed. The International Rice Research Institute, Manila, Philippines, <http://books.irri.org/9711041510.content.pdf>.
- Wang, E., Engel, T., 1998. Simulation of phenological development of wheat crops. *Agricultural Systems* 58 (1), 1–24.
- Xue, Q., Weiss, A., Baenziger, P.S., 2004. Predicting leaf appearance in field-grown winter wheat: evaluating linear and non-linear. *Ecological Modelling* 175 (3), 261–270.
- Yao, F., Xu, Y., Lin, E., Yokozawa, M., Zhang, J., 2007. Assessing the impacts of climate change on rice yields in the main rice areas of China. *Climatic Change* 80 (3–4), 395–409.
- Yin, X., 1996. Quantifying the effects of temperature and photoperiod on phenological development to flowering in rice. PhD thesis. Wageningen University, 173 p.
- Yin, X., Kropff, M.J., Goudriaan, J., 1997a. Changes in temperature sensitivity of development from sowing to flowering in rice. *Crop Science* 37 (6), 1787–1794.
- Yin, X., Kropff, M.J., Horie, T., Nakagawa, H., Centeno, H.G.S., Zhu, D., Goudriaan, J., 1997b. A model for photothermal responses of flowering in rice. I. Model description and parameterization. *Field Crops Research* 51 (3), 189–200.
- Yin, X., Kropff, M.J., McLaren, G., Visperas, R.M., 1995. A nonlinear model for crop development as a function of temperature. *Agricultural and Forest Meteorology* 77 (1–2), 1–16.
- Yin, X., Kropff, M.J., Nakagawa, H., Horie, T., Goudriaan, J., 1997c. A model for photothermal responses of flowering in rice. II. Model evaluation. *Field Crops Research* 51 (3), 201–211.
- Zhang, T., Zhu, J., Wassmann, R., 2010. Responses of rice yields to recent climate change in China: an empirical assessment based on long-term observations at different spatial scales (1981–2005). *Agricultural and Forest Meteorology* 150 (7–8), 1128–1137.
- Zhang, T., Zhu, J., Yang, X., 2008. Non-stationary thermal time accumulation reduces the predictability of climate change effects on agriculture. *Agricultural and Forest Meteorology* 148 (10), 1412–1418.