Extending a first-principles primary production model to predict wheat yields


It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1016/j.agrformet.2020.107932

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR
Central Archive at the University of Reading
Reading’s research outputs online
Extending a first-principles primary production model to predict wheat yields

Shengchao Qiao\textsuperscript{1,2}, Han Wang\textsuperscript{1,2,*}, I. Colin Prentice\textsuperscript{1,3,4}, Sandy P. Harrison\textsuperscript{1,5}

\textsuperscript{1}Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China;
\textsuperscript{2}Joint Center for Global Change Studies (JCGCS), Beijing 100875, China;
\textsuperscript{3}AXA Chair of Biosphere and Climate Impacts, Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, SL5 7PY, UK;
\textsuperscript{4}Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109, Australia;
\textsuperscript{5}School of Archaeology, Geography and Environmental Sciences (SAGES), University of Reading, Reading, RG6 6AH, UK.

*Corresponding author: Han Wang (wanghan_sci@yahoo.com)

Keywords
wheat; photosynthesis; crop yield; crop model; CO\textsubscript{2} fertilization; harvest index

Highlights
1. A novel, simply formulated crop model with quantified uncertainties successfully predicts wheat yields at research sites in China.
2. The model captures the time course of GPP and variations in biomass and yield across sites and between years.
3. Sensitivity analyses and future projections indicate a positive response of wheat yield to rising CO\textsubscript{2}, partially counteracted by a negative response to warming.
Data statement

The climate and flux data from WeiShan can be obtained by contacting Huimin Lei (leihm@tsinghua.edu.cn). The flux data for YuCheng, the climate, LAI, crop data used in this manuscript are publicly available from the National Ecosystem Research Network of China CNERN (http://www.cnern.org.cn/). All climate data driving the PC model runs for future and the model outputs of LPJmL, EPEIC, GEPIC are publicly available from Inter-Sectoral Impact Model Intercomparison Project-2b (ISIMIP2b: https://www.isimip.org/protocol/#isimip2b/). The PC model code will be available from Mendeley data.

Abstract

Climate exerts a major influence on crop development and yield. Despite extensive modelling efforts, there is still considerable uncertainty about the consequences of a changing climate for the yields of major crops. Existing crop models are complex and rely on many assumptions and parameters, motivating a quest for more parsimonious models with stronger theoretical and empirical foundations. This paper presents a prototype of such a model for wheat, informed by measurements of gross primary production (GPP), biomass and yield at research sites across the wheat-growing regions of China. First, GPP was predicted using a recently developed first-principles model driven only by climate, carbon dioxide (CO$_2$) concentration, and light absorbed by leaves. Modelled GPP was shown to agree well with eddy-covariance measurements. Second, the data were used to show that above-ground biomass (AB) is proportional to time-integrated GPP, and that grain yield shows a saturating relationship with AB. Simple empirical equations based on these findings were combined with modelled GPP to predict yield, including propagated errors due to parameter uncertainty in both the GPP model and the
empirical equations. The resulting ‘hybrid’ model, applied in a variety of climates, successfully predicted measured interannual variations in AB and yield. Third, the model was extended to include a phenology scheme, a mass-balance equation relating mean leaf area index to accumulated GPP over growth phase, and an independently observed response of leaf mass-per-area to CO₂. Sensitivity analyses and scenario runs with this extended model showed a positive but saturating (at ~600 ppm) response of yield to rising CO₂, consistent with experimental evidence. This positive effect was partially counteracted by a net negative response of yield to increasing temperature, caused by increasing photorespiration and an accelerated growth cycle.

1. Introduction

An adequate food supply is an essential basis for economic development and social stability in the context of increasing population and continuing anthropogenic climate change (Porter et al., 2014). As one of the world’s four major crops (with maize, rice and soybean), wheat provides about a quarter of the world’s cereal production (FAOSTAT, 2018) which, in turn, provides two-thirds of human caloric intake (Zhao et al., 2017). Wheat was introduced from the Near East and been cultivated in China since the late 6th to early 5th millennium BP (Betts et al., 2014). China is now both the largest producer and the largest consumer of wheat (Wang et al., 2009). Current wheat production in China exceeds 134 Mt of grain per year. This is more than 17% of the total global wheat production, and about 22% of the total cereal production of China (FAOSTAT, 2018). Thus, even a small fluctuation in China’s wheat production could potentially impact not only the Chinese economy but also global food security.

The growth and harvestable yield of wheat are determined by environmental factors (Asseng et al., 2004) but also strongly influenced by management (He et al., 2015). Light, CO₂, temperature, water
and nutrient availability define the basic conditions for the growth and development of the crop. Light and CO₂ directly affect photosynthesis (Gerbaud and Andre, 1980); temperature further influences growth and development processes including germination, anthesis and harvest (Asseng et al., 2011; He et al., 2015; Liu et al., 2018; Porter and Gawith, 1999; Tao et al., 2012); water and nutrient availability principally influence foliage cover (Nielsen and Halvorson, 1991; Pan et al., 2019) and therefore the absorption of light for photosynthesis. However, the basic conditions of wheat growth, especially temperature and CO₂ concentration, are changing. Temperatures in China have risen by 1.2˚C over the past few decades (Cao et al., 2017; Piao et al., 2010) and continued warming is expected in the coming decades (Kirtman et al., 2014). Winter is warming faster than summer (Piao et al., 2010; Wu et al., 2017) and this situation is potentially unfavorable to the production of wheat (Brooking, 1996). On the other hand, atmospheric CO₂ already exceeds 400 ppm, more than 40% above its pre-industrial level and is expected to continue rising (Collins et al., 2014). For C₃ crops (including wheat) the effect of rising CO₂ level on photosynthesis is positive (Ainsworth and Rogers, 2007; Boylan, 2016; Sage et al., 1989), leading to higher photosynthetic productivity and potentially also grain yield (Lawlor and Mitchell, 1991). Improved management practices (e.g. fertilization, irrigation) and crop breeding have also contributed to increased wheat yield (Qin et al., 2015; Yu et al., 2019) under current climate conditions, and this trend is expected to continue.

The combined effects of changes in climate, CO₂ and management are highly uncertain (Challinor and Wheeler, 2008) and numerical models are needed to project future trends in yield in different regions, and thereby to facilitate adaptation in the food production system. Such models should integrate knowledge from experiments and observations with theoretical understanding. Crop models have been under development for at least 40 years, and there are now many models with the technical
capacity to simulate the growth and development of wheat (Blanc, 2017; Huang et al., 2016; Palosuo et al., 2011). However, current crop models require detailed input information that is challenging to collect over large scales and potentially subject to change in a dynamic environment. Moreover, inter-model comparisons have revealed large differences between model predictions of both current yields and future trends (Liu et al., 2019; Nelson et al., 2014; Ostberg et al., 2018). This situation parallels that for natural vegetation models (Prentice et al., 2015), and suggests that current crop models contain untested and potentially incorrect assumptions. Recently, however, progress has been made in developing simpler vegetation models, based on theoretical principles but drawing extensively on empirical data to test each model component (Franklin et al., 2017; Wang et al., 2017). Here we adopt this novel approach to develop a prototype model to predict wheat growth and yield.

The starting point for this model (‘PC’, for P crop: see Fig. 1) was the universal C_3 primary production model ‘P’ (Stocker et al., 2019; Wang et al., 2017). The P model is a theoretically derived and extensively tested light use efficiency (LUE) model that predicts gross primary production (GPP) as a function of climate, absorbed light and CO_2 on time steps of a week to a month. Further model development and testing of the PC model, presented here, used measurements of GPP, biomass and yield at research sites across the wheat-growing regions of China. In the first step of the analysis, the original P model was applied to predict GPP at two flux-tower sites situated in wheat crops in order to test its performance. In the second step, simple empirical equations were fitted to experimental data at several field research sites in order to relate accumulated GPP to aboveground biomass (AB), and AB to grain yield. These equations were combined with the P model to predict yields for different sites and years; these predictions were compared to observed yields. Uncertainties in predicted yields due to key parameters of the P model, and to the fitted coefficients of the empirical equations, were quantified. In
the third step, the model was extended to represent the responses of yield to environmental change by
the inclusion of (a) a general scheme to predict phenology, (b) a mass-balance equation quantifying
leaf area index (LAI) consistent with a given GPP, and (c) an observed relationship between leaf mass-
per-area (LMA) and the CO$_2$ concentration experienced during crop growth. These extensions allowed
modelled light absorption to be influenced by changes in growing-season length, and changes in
modelled GPP to feed back to LAI. The extended model was applied at six field sites to project future
wheat yields under different combinations of increasing CO$_2$ and temperature, in a sensitivity analysis
for combinations of CO$_2$ and temperature increase, and in alternative scenario runs for future CO$_2$ and
temperature change.

2. Material and methods

2.1 The P model

The P model is based on the standard biochemical model of C$_3$ photosynthesis (Farquhar et al.,
1980), with additional formulations that allow photosynthetic capacities and stomatal behaviour to
acclimate to environmental conditions on weekly to monthly time scales (Wang et al., 2017).
Instantaneous photosynthetic rates according to the standard model are the lesser of the electron
transport-limited rate ($A_J$) and the carboxylation-limited rate ($A_C$). $A_C$ is proportional to the Rubisco
capacity ($V_{cmax}$). In the P model, $V_{cmax}$ is assumed to acclimate to growth conditions such that the two
rates are co-limiting under average daytime conditions (Maire et al., 2012; Smith et al., 2019). $A_J$ is
proportional to the absorbed photosynthetic photon flux density (PPFD) at low PPFD, increasing with
PPFD towards a light-saturated rate that is proportional to the electron-transport capacity ($J_{max}$). In the
P model, acclimation of $J_{max}$ parallels that of $V_{cmax}$ and their ratio is set to maximize the benefit ($A_J$)
minus the cost of maintaining \( J_{\text{max}} \). Both \( A_{C} \) and \( A_{J} \) are functions of the leaf-internal CO\(_2\) partial pressure \((c_{i})\), whose ratio \((\chi)\) to the ambient \( \text{CO}_2 \) partial pressure \((c_{a})\) is determined by stomatal responses to the relative rates of carbon gain and water loss. In the P model, \( \chi \) is determined by the least-cost criterion (Prentice et al., 2014), which minimizes the combined costs of maintaining the Rubisco and water transport capacities. The three constraints (on \( V_{\text{cmax}}, J_{\text{max}} \) and \( \chi \)) lead to an expression for weekly to monthly GPP under field conditions that has the mathematical form of a LUE model. That is, accumulated GPP is proportional to absorbed PPFD:

\[
\text{GPP} = \Phi_{0}I_{\text{abs}}m\sqrt{1 - (c^{*}/m)^{2/3}} \quad (1a)
\]

where

\[
m = (c_{i} - \Gamma^{*}) / (c_{i} + 2\Gamma^{*}) \quad (1b)
\]

\[
c_{i} / c_{a} = \Gamma^{*} / c_{a} + (1 - \Gamma^{*} / c_{a}) \xi / (\xi + \sqrt{D}) \quad (1c)
\]

\[
\xi = \sqrt{[\beta (K + \Gamma^{*}) / 1.6\eta^{*}]} \quad (1d)
\]

Here, \( \Phi_{0} \) is the intrinsic quantum yield (mol \( \text{CO}_2 \) mol\(^{-1}\)); \( I_{\text{abs}} \) is the PPFD intercepted and absorbed by the canopy (mol m\(^{-2}\) s\(^{-1}\)); \( c_{a} \) is the ambient atmospheric \( \text{CO}_2 \) partial pressure (Pa); \( \Gamma^{*} \) is the photorespiratory compensation point (Pa); \( \eta^{*} \) is the viscosity of water, relative to its value at 25 \(^\circ\)C (dimensionless); \( D \) is the vapour pressure deficit (Pa); \( K \) is the effective Michaelis-Menten coefficient of Rubisco (Pa); and \( c^{*} = 0.41 \) and \( \beta = 146 \) are dimensionless constants, where \( c^{*} \) has been estimated from observed \( J_{\text{max}}; V_{\text{max}} \) ratios and \( \beta \) from observed stable carbon isotope ratios (Wang et al., 2017).

The P model thus calculates GPP as the product of \( I_{\text{abs}} \), which is the product of incident PPFD and \( f_{\text{APAR}} \) (the fraction of incident PPFD absorbed by foliage) and LUE. LUE is the product of \( \Phi_{0}, m \) and...
the square-root term in equation (1a). The parameters $I^*$, $\eta^*$ and $K$ depend on temperature (Bernacchi et al., 2001; Wang et al., 2017) and $I^*$ and $K$ depend on atmospheric pressure (Farquhar et al., 1980). The inputs to the model are air temperature ($T$), relative humidity (RH), incident PPFD, $f$APAR, elevation (to calculate atmospheric pressure) and $c_a$ (the product of the current year’s mole fraction of CO$_2$ with atmospheric pressure). When driven by satellite-derived $f$APAR data, the model has been shown to reproduce monthly GPP well at eddy-covariance flux tower sites in natural vegetation worldwide (Stocker et al., 2019; Wang et al., 2017) and geographic patterns, seasonal cycles and interannual variability of GPP at flux tower sites located in different biomes, including croplands (Balzarolo et al., 2018).

![Diagram of PC model](image)

**Figure 1: The structure of the PC model.** $\Sigma$GPP: accumulated gross primary production over growing season (g C m$^{-2}$); $N$: total application of nitrogen (kg N hm$^{-2}$). LAI: leaf area index (dimensionless). $\phi$: light use efficiency (%). $\sum I$: the sum of incident light over the growth phase (mol
photo m\(^{-2}\)). \(k\): a dimensionless constant, \((k = 0.5)\). \(\eta\): the fraction of \(\Sigma\text{GPP}\) allocated to leaves (dimensionless). LMA: leaf mass per area (g m\(^{-2}\)). Climate here comprises temperature, relative humidity and incident photosynthetic photon flux density. Boxes with grey dash line indicate the already published model or known information, whereas the boxes with blue dash line indicate the new model and its extension developed here.

We ran the P model on a weekly time step. The model has already been shown to work well on a ten-daily time step (Balzarolo et al., 2018). We applied the ‘BRC’ model set-up (Stocker et al., 2019), which differs from the original published version (Wang et al., 2017) by incorporating an observed temperature-dependence of \(\Phi_0\) (Bernacchi et al., 2003):

\[
\Phi_0 = (0.352 + 0.021T - 0.00034T^2) / 8
\]  

(2)

2.2 Sites and field data

Data from 12 agricultural sites in the main wheat-growing area of China (see Fig. 2) were used for the second step of model development and testing (see Table 1 and Table 2). More than 90% of wheat production occurs in the areas where these sites are located (Wang et al., 2009), so they are representative of the environmental conditions for wheat production in China.
Figure 2: Locations of sites providing experimental data.
<table>
<thead>
<tr>
<th>Site</th>
<th>Code</th>
<th>Longitude (E, °)</th>
<th>Latitude (N, °)</th>
<th>Elevation (m)</th>
<th>Mean annual air temperature (℃)</th>
<th>Mean annual precipitation (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WeiShan</td>
<td>WS</td>
<td>116.83</td>
<td>36.23</td>
<td>20</td>
<td>13.3</td>
<td>532</td>
</tr>
<tr>
<td>YuCheng</td>
<td>YC</td>
<td>116.57</td>
<td>36.82</td>
<td>22</td>
<td>13.2</td>
<td>530</td>
</tr>
<tr>
<td>ChangShu</td>
<td>CS</td>
<td>120.7</td>
<td>31.55</td>
<td>3.1</td>
<td>16.6</td>
<td>1321.2</td>
</tr>
<tr>
<td>ChangWu</td>
<td>CW</td>
<td>107.68</td>
<td>35.23</td>
<td>1220</td>
<td>9.1</td>
<td>580</td>
</tr>
<tr>
<td>LuanCheng</td>
<td>LC</td>
<td>114.68</td>
<td>37.88</td>
<td>50.1</td>
<td>12.2</td>
<td>536.8</td>
</tr>
<tr>
<td>FengQiu</td>
<td>FQ</td>
<td>114.4</td>
<td>35</td>
<td>67</td>
<td>13.9</td>
<td>605</td>
</tr>
<tr>
<td>YanTing</td>
<td>YG</td>
<td>105.45</td>
<td>31.27</td>
<td>420</td>
<td>17.3</td>
<td>836</td>
</tr>
<tr>
<td>HaiLun</td>
<td>HL</td>
<td>125.92</td>
<td>47.45</td>
<td>240</td>
<td>4</td>
<td>550</td>
</tr>
<tr>
<td>LaSa</td>
<td>LA</td>
<td>91.33</td>
<td>29.67</td>
<td>3688</td>
<td>6</td>
<td>425</td>
</tr>
<tr>
<td>LinZe</td>
<td>LZ</td>
<td>100</td>
<td>39.35</td>
<td>1384</td>
<td>1.5</td>
<td>550</td>
</tr>
<tr>
<td>NaiMan</td>
<td>NM</td>
<td>120.7</td>
<td>42.92</td>
<td>358</td>
<td>5</td>
<td>425</td>
</tr>
<tr>
<td>ShaPotou</td>
<td>SP</td>
<td>104.95</td>
<td>37.45</td>
<td>1250</td>
<td>9.6</td>
<td>1250</td>
</tr>
</tbody>
</table>

Table 1: Background information about the sites.
Table 2: Data details and use. √: data are available; a: sites were used to test the P model; b: sites were used to derive the GPP to AB relationship; c: sites were used to derive the AB to yield relationship; d: sites were used to test the final model. * Two years (2004-2005) during the data span are available for the flux data at YuCheng.
Two flux tower sites (WeiShan, YuCheng; see Table 2 and Fig. 2) were used to test the GPP predictions. One full year of observations from WeiShan (2006) and two years of observations from YuCheng (2004, 2005) were available. The climate data (PPFD, T and RH) and canopy coverage (here estimated from LAI by Beer’s law), required as input to the P model, were obtained from on-site measurements for WeiShan provided by the original authors (Lei and Yang, 2010) and downloaded from the National Ecosystem Research Network of China (CNERN: http://www.cnern.org.cn/) for YuCheng. CO₂ concentrations used are the annual global average obtained from the United States National Oceanic & Atmospheric Administration (NOAA: https://www.esrl.noaa.gov/gmd/ccgg/trends/).

There are no data on AB or grain yield from WeiShan, although this information is available for YuCheng. We therefore used the two years of data from YuCheng to derive the relationship between GPP and AB. We obtained experimental data from ten additional agricultural sites providing information on AB and grain yield from CNERN. CNERN also provided data on climate (including PPFD, T and RH), LAI, dates of the growing period, wheat varieties planted and management practices (including irrigation and the supply of total nitrogen, phosphate and potassium) for all of these sites. However, the records cover different periods (see Table 2): some sites only have data for one year (LZ, SP, NM); one site has data for two years (HL); the remaining sites have records spanning multiple years (CS, CW, FQ, LC, YG, LS). We used all the available data from these ten sites and YuCheng together (584 data points) to estimate the allocation relationship between AB and grain yield. We used data from six sites (CS, CW, FQ, LC, YG, YC) with records for more than two years to test the final model. We could not use the Lasa site for testing because there are no LAI data from this site.

Climate data were pre-processed on a weekly time step, with PPFD summed and T and RH averaged. Then vapour pressure deficit (D, kpa) was calculated according to the following equation (Campbell and Norman, 2012):

\[ D = 610.8 \exp\left[\frac{17.27T}{T+237.3}\right] \frac{(100 - \text{RH})}{100} \]  

(3)
The fraction of absorbed photosynthetically active radiation was estimated by Beer’s law (Monsi, 1953) from LAI:

\[ f_{\text{APAR}} = 1 - \exp(-k \cdot \text{LAI}) \]  

(4)

where \( k \) is a dimensionless constant, assigned a generic value of 0.5. LAI was measured several times over the growing season, but the times of measurement varied from year to year and site to site. The LAI data used as input to test the P model (WS 2006, YC 2004-2005) are based on eight to ten observations at each site over growing season. We interpolated the data to weekly mean values using a polynomial regression of LAI against time. Measurements of LAI at the sites used to test the crop model (PC model) were made more sporadically (less than five observations per growing season) and therefore inadequate for polynomial regression. We derived LAI values for PC model test from the MODIS LAI product (MCD15A3H: 4-day time-step and 500m resolution, https://modis.gsfc.nasa.gov/). Since MODIS severely underestimates the observed LAI at the six test sites, we calculated the relationship between observed and MODIS LAI by linear regression and used the slope of this regression to rescale the MODIS LAI data and derive weekly mean LAI.

2.3 Derivation of allocation relationships

We hypothesized that a fixed proportion (\( \varepsilon \)) of accumulated GPP during the growing season would be allocated to above-ground biomass (AB). We calculated GPP accumulation (\( \Sigma \text{GPP} \)) from the beginning of the growing season to the day when AB was measured, then derived an empirical relationship between AB and \( \Sigma \text{GPP} \) by linear regression. We used the slope of this linear regression as an estimate of \( \varepsilon \).

We hypothesized that grain yield should increase, monotonically but not necessarily linearly, with AB, and that this relationship might be influenced by wheat varieties and management practices. Non-linear regression was used to derive an empirical relationship between grain yield and AB, taking
account of the effect of nitrogen supply and wheat variety on the relationship, meanwhile testing the
effects of irrigation and the application of phosphate and potash. Non-linear regression was performed
using a mixed-effects model in the nlm package of R. The form of the fitted equation is as follows:
\[
Yield = (a \cdot N + b) \left[1 - \exp(c \cdot AB)\right] + d
\]  
where \(N\) is the total application of nitrogen (kg N hm\(^{-2}\)), and \(a, b, c\) and \(d\) are parameters to be fitted.
Wheat variety was considered as a random effect added to parameter \(b\), thus allowing maximum yields
to differ by variety. The potential effects of other factors (irrigation, precipitation, mean temperature
during the growth season) were tested by examining the residuals from this regression. To check the
goodness of fit of the non-linear regression, linear regression was also performed both using all the
data together, and for each variety separately. The root mean squared error (RMSE) and Akaike
Information Criterion (AIC) were calculated as an indicator of the goodness of fit of each model.

2.4 Model evaluation

We tested the performance of the PC model by comparing interannual variations in predicted and
observed AB and grain yield over multiple years at the six test sites (CS, CW, FQ, LC, YC, YG), using
meteorological observations from each site to drive the model. The simulated accumulated GPP during
the growing season was allocated to AB using the fixed ratio \((\varepsilon)\) obtained by linear regression, and AB
at harvest was allocated to grain yield using the non-linear relationship described above (Equation 5).
The growing season was defined as the period when mean daily temperature was above 0°C.
Interannual variation in yield provides an independent test of the model as no information on
interannual variability was used in the derivation of empirical relationships used in the model.

There are two sources of uncertainty in the model predictions: the input data (climate and LAI)
and the model parameters. We assumed the input data were reliable and focused on parameter
uncertainty. We considered each of the sources of uncertainty in the individual equations in the P model
independently, and combined these uncertainties using the standard error propagation formula (Prentice and Thomas, 2018):

\[ u^2(y) = \sum_i (\frac{\partial f}{\partial x_i})^2 u^2(x_i) \]  

(6)

where \( u(y) \) is the standard uncertainty (of GPP or AB or yield), \( \frac{\partial f}{\partial x_i} \) is the sensitivity to variable \( x_i \) (obtained by differentiating the individual equations), and \( u(x_i) \) is the standard uncertainty of \( x_i \).

### 2.5 Model extension

#### 2.5.1 Phenology scheme

The phenology scheme for wheat was adopted from the LPJmL4 model (Bondeau et al., 2007; Schaphoff et al., 2018). Sowing and maturity dates were obtained from the datasets provided in the global gridded crop model inter-comparison project (Elliott et al., 2015). A phenological scalar (\( f_{PHU} \)) ranging from 0 at sowing to 1 at harvest was computed:

\[ f_{PHU} = \sum_i (T_m - T_b) / PHU \]  

(7)

where \( n \) is the number of days from sowing, \( T_m \) is the daily mean air temperature (°C) and \( T_b \) is the base temperature (here 0°C) and used to determine LAI development, using a sigmoid curve during the growth phase and a quadratic curve during the senescent phase. (In the absence of water and nutrient stresses, LAI is assumed to follow this optimal curve. During the growth phase:

\[ f_{LAI_{max}} = \frac{f_{PHU}}{[f_{PHU} + \exp(l_1 - l_2 \cdot f_{PHU})]} \]  

(8)

where LAI is a time-dependent fraction \( f_{LAI_{max}} \) of maximum LAI (\( LAI_{max} \)), and \( l_1 \) and \( l_2 \) are the first and second inflection points. During the senescence phase:

\[ f_{LAI_{max}} = \left[\frac{(1 - f_{PHU})^2}{(1 - f_{PHU \cdot sen})^2}\right] (1 - f_{LAI_{max} - harvest}) + f_{LAI_{max} - harvest} \]  

(9)
where $f_{PHU,\text{sen}}$ is the fraction of PHU when senescence begins, and $f_{LAI,\text{max-harvest}}$ is the fraction of maximum LAI at harvest (here fixed to zero). The $f_{PHU}$ values corresponding to the $l_1$ and $l_2$ inflection points were derived from Figure 4 in Bondeau et al. (2007) and the parameter values of $l_1$ and $l_2$ were calculated for these $f_{PHU}$ values using the method of Neitsch et al. (2011). We used values for the $l_1$ and $l_2$ inflection points of 0.89 and 10, respectively.

2.5.2 Dynamic leaf area index

Prognostic calculation of LAI was enabled by fitting a linear relationship between leaf biomass and AB based on data from the field sites, then solving for LAI in the mass-balance equation:

$$\frac{\text{LMA}}{\eta} \cdot \text{LAI} = \varphi \Sigma I [1 - \exp(-k \cdot \text{LAI})]$$

(10)

where $\eta$ is the fraction of $\Sigma \text{GPP}$ allocated to leaves, LMA is the leaf mass-per-area ($\text{g m}^{-2}$), $\varphi$ is the modelled LUE (the ratio of modelled GPP, following equation (1), to absorbed PPFD), and $\Sigma I$ is the accumulated incident PPFD (mol photon m$^{-2}$). $k$ is a dimensionless constant ($k = 0.5$). Equation (10) indicates that the LAI demand (left hand side) must equal to its supply (right hand side). The LAI demand represents the allocation of accumulated GPP to canopy to support a certain level of LAI. The LAI supply represents the carbon accumulation supported by a certain level of LAI. $\eta$ was estimated from the observed data on leaf biomass and $\Sigma \text{GPP}$ from the YuCheng site in 2004 and 2005, LMA was set at 35.7 g m$^{-2}$ corresponding to the mean observed value at YuCheng and allowed to increase linearly with CO$_2$ using the observed slope (0.05 g m$^{-2}$ per ppm) from Thilakarathne et al. (2013).

2.6 Model application

2.6.1 Sensitivity analysis

Using 2005 as a baseline (baseline temperature is the weekly mean temperature over the growing season and baseline CO$_2$ is 380 ppm), we ran simulations with the extended model, including
prognostic phenology and dynamic LAI and LMA, with temperature increasing by 0.05° increments up to 5° above the baseline temperature and CO₂ concentration increasing by increments of 5 ppm up to 500 ppm above the baseline CO₂ concentration. These changes were superimposed on the weekly mean temperatures and on the annual CO₂ concentration. All other inputs (radiation, relative humidity, management practices and wheat variety) were fixed at their 2005 values.

2.6.2 Future scenarios

We used the model to examine the consequences of potential future climate changes on wheat yields, following the protocol used by the Inter-Sectoral Impact Model Intercomparison Project-2b (ISIMIP2b: https://www.isimip.org/protocol/#isimip2b/). Climate data, including daily mean temperature, photosynthetically active radiation (assumed to be half of downward shortwave radiation) and relative humidity from the MIROC5 simulations, and CO₂ concentrations, for two scenarios (RCP2.6 and RCP6.0) were used to run the PC model at six test sites (CS, CW, FQ, LC, YC, YG) from 2006 to 2099. Management practices and wheat varieties were unchanged from 2005.

The LPJmL (Bondeau et al., 2007; Muller and Robertson, 2014; Schaphoff et al., 2018), GEPIC (Liu et al., 2007) and PEPIC (Liu et al., 2016) crop models have been used to simulate future wheat yields in ISIMIP2b. We compared our future projections of yield with results from these three models. In ISIMIP2b, these models ran simulations with full irrigation and no irrigation. In order to eliminate the effect of variable water supply, we compared our results with those from the full-irrigation run. We extracted simulated wheat yields at our six test sites from the results of each model for the RCP2.6 and RCP6.0 scenarios. Further information about these three models, including input data, leaf area, phenological development, yield formulation and stresses considered, is given in Table S1.
3. Results

3.1 Modelled versus observed GPP

Predicted weekly GPP values were consistent with the observations from the flux towers, both in their magnitudes (Fig. 3a) and their patterns over the growing season (Fig. 3b). Observed and predicted GPP were highly correlated ($R^2 = 0.81$, RMSE = 10.1 g C m$^{-2}$ week$^{-1}$) and the slope of the relationship was close to 1:1 (slope = 1.07 ± 0.08) with a non-significant offset (intercept = 0.79 ± 4.67 g C m$^{-2}$ week$^{-1}$).

![Plot showing modelled versus observed GPP](image-url)

**Figure 3:** Comparison of predicted and observed gross primary production at two sites. (a) Scatterplot. The thick black line is the linear regression. The grey number is not significant; (b) GPP during the growing season (weekly sums). WS-2006, YC-2005 and YC-2004, represents WeiShan in 2006, YuCheng in 2005 and YuCheng in 2004, respectively.
3.2 The relationship between AB and GPP

A strong linear relationship (Fig. 4) was shown between AB and accumulated GPP \((r = 0.92)\) with an estimated 72\% (slope, \(\varepsilon = 0.72\)) of accumulated GPP allocated to AB. The intercept was statistically significant, but small enough to be neglected.

\[ y = 0.72x \]
\[ r = 0.92 \]
\[ n = 98 \]

**Figure 4:** The proportion of accumulated gross primary production (GPP) allocated to aboveground biomass. The data are observations during the wheat growth season at YuCheng from 2004 to 2005. All values were accumulated from green-up to measurement time.

3.3 The relationship between yield and AB

Yield was shown to be a saturating function of AB (Fig. 5a). Nitrogen addition affected the overall level of allocation (Fig. 5b), with higher nitrogen supply causing high allocation to AB. The relationship was affected by wheat variety, and a saturating relationship can also be shown within each of the varieties that covers a large range of AB (range > 1800 g m\(^{-2}\)), with a substantially smaller RMSE and AIC compared to linear fits (Fig. S1). Moreover, the slopes of linear regressions fitted to each variety separately decline with the mean value of AB for the variety (Fig. 7 and Fig. S2). In other words, at the high end of AB values, the increment in yield diminishes with the increment in AB. These
results indicate that the non-linear, saturating relationship of yield to AB applies generally, both within and across varieties.

Figure 5: Results of the mixed-effects model. (a) Yield versus aboveground biomass (AB). Pink line is linear regression and black line is non-linear regression. (b) Yield trend with AB, including the effect of nitrogen addition. The solid line is the fitted curve of yield with AB at high nitrogen level (pure nitrogen added \(= 300 \text{ kg hm}^{-2}\)) and the dotted line is at low nitrogen level (pure nitrogen added \(= 100 \text{ kg hm}^{-2}\)). The open and closed circles represent the observations with total application of pure nitrogen at levels above and below 200 kg hm\(^{-2}\), respectively. (c) Scatterplot including AB and nitrogen as predictors. (d) Scatterplot including AB, nitrogen and variety as predictors. Grey numbers are non-significant.
Figure 6: The relationship between residuals and wheat varieties. (a) Without the random effect of variety; (b) with this effect. Each box represents a wheat variety. The red line is zero and the black dots are outliers.

Figure 7: The fitted slope and mean value of aboveground biomass based on the linear regression of yield against aboveground biomass within each variety. (See Fig. S2 for the separate linear regressions.)

The comparison between simulated and observed yields improved when variety was taken into account (Fig. 5c compared to Fig. 5d). Residuals of the non-linear regression were reduced (Fig. 6), and the correlation between predicted and observed yield improved ($R^2$ increased from 0.68 to 0.83). Irrigation, mean temperature over the growing season, and the supply of phosphate and potassium had
no significant effects on the relationship between AB and yield (P > 0.05) on yield (Fig. S3), indicating that the first-order effects of these factors are already subsumed in AB.

3.4 Model evaluation

3.4.1 Prediction of AB and yield variations

The model captured the pattern of interannual variation in AB (see Fig. 8) and, although there were some anomalous years, the predicted AB was generally within the range of the observations. The correlation between predicted and observed AB was moderate ($r = 0.40$). The simulated and observed yields matched reasonably well ($r = 0.61$) and interannual variations in grain yield were captured (Fig. 9 and Table S2), with observations almost always within the uncertainty range of the predictions.

**Figure 8:** Comparisons of observed and modelled aboveground biomass. (a) Interannual variation at four sites: ChangShu, ChangWu, YuCheng, YangTing. (b) Scatterplot of predicted and observed AB at all sites.
Figure 9: Comparisons of observed and modelled yield. (a) Interannual variations at four sites: ChangShu, ChangWu, YuCheng, YangTing. (b) Scatterplot of predicted and observed yield at all sites.

3.4.2 Uncertainty analysis

Uncertainty in model predictions could originate either in the input data (climate, LAI) or in the model. We assumed that the input data were reliable, and used YuCheng 2005 as a case study to analyse the uncertainties due to the following model parameters:

- The two most uncertain quantities ($\beta, c^*$) in the P model (Prentice and Thomas, 2018). $\beta$ is the ratio of the unit costs for the maintenance of carboxylation and transpiration capacities, evaluated at 25°C. It determines the value of the ratio of leaf-internal to ambient CO$_2$ (an index of stomatal behaviour) under standard conditions. $c^*$ is the unit cost of maintaining electron-transport capacity and determines the extent to which optimum carboxylation capacity is lowered because of the cost of maintaining an equivalent capacity for electron transport (Equation 1).

- The proportionality constant ($\varepsilon$) between biomass and accumulated GPP.

- The four main parameters ($a, b, c, d$) from the formula relating yield to AB (Equation 5).
The calculated uncertainties with respect to different model parameters for predicted GPP and yield are shown in Fig. 10. When the parameters were varied by ± 10%, the total uncertainty of predicted GPP was ~9%. The largest source of this uncertainty (~6%) was associated with the parameter $c^*$, which is an important control on the magnitude of simulated GPP. This parameter also contributes substantially (~4%) to the uncertainty in simulated yield. The other parameter contributing substantially (~7%) to this uncertainty is the main slope parameter ($b$), which is the principal control over the yield attained for a given biomass.

![Figure 10: The impact of parameter uncertainty on the prediction of GPP accumulation and yield. YuCheng 2005 was selected as a case study. Triangles represent GPP and dots represent yield. The x-axis represents the progressive inclusion of ± 10% uncertainty in successive parameters, indicated by their symbols.](image)

### 3.5 Model extension and application

#### 3.5.1 Testing the phenology scheme

The phenology scheme was shown to reproduce seasonal patterns of LAI today (Fig. 11) reasonably well ($R^2 = 0.84$).
Figure 11: Observed versus predicted leaf area index (LAI) with the LPJmL4 phenology scheme. The inset shows the observed (circles, green is mean value) and predicted (line) seasonal time course of LAI at the YuCheng site in 2005. The grey number is not significant.

3.5.2 The relationship between leaf biomass and accumulated GPP

A strong linear relationship ($r = 0.94$) was found between leaf biomass and accumulated GPP, allowing us to estimate $\eta = 0.32$ (Fig. 12) and thereby solve equation (10) for mean LAI over growth phase.
Figure 12: The proportion of gross primary production (GPP) allocated to leaves. The data are observations during the wheat growth phase in 2004 and 2005 at YuCheng. All values were accumulated from green-up to measurement time.

Projections of changing LAI as a function of CO$_2$ concentration are shown in Figure 14b and Figure S4. Modelled LAI increases in response to increasing CO$_2$, but when the effect of increasing LAI on LMA is considered, the increase is much smaller and reaches a maximum at around 600 ppm. This behaviour is consistent with the maximum yield enhancement indicated by raised CO$_2$ experiments, as summarized in the meta-analysis by Broberg et al. (2019).

3.5.3 Sensitivity analyses

Modelled grain yields increase with rising CO$_2$ concentration, and decrease with increasing temperature, when other variables are kept fixed (Fig. 13). Higher temperature shortens the growing season (Fig. 14d) leading to a reduction in total absorbed light. In addition, the response of modelled LUE to rising temperature follows a unimodal curve (see Fig. 14c) such that increasing temperature above the optimum reduces GPP (Long, 1991). Lower GPP means lower yield. On the other hand, rising CO$_2$ monotonically increases LUE, GPP and yield; and rising GPP leads to rising LAI, amplifying this effect. But the net effect of CO$_2$ is limited by an increase in LMA.

For these scenarios and sites, the modelled positive effect of rising CO$_2$ concentration on yield was greater than the negative effect of increasing temperature. However, the modelled reductions in yield caused by rising temperature differed among the sites (Fig. 13). Modelled wheat yields in warmer regions today, like YG and CS, are more sensitive to warming than cooler regions such as LC.
Figure 13: The response of predicted yield to rising CO₂ and increasing temperature at six sites. Dots mean increasing on temperature and mean rising on CO₂ concentration in the last decade (2090-2099) comparing with the first decade (2006-2015) under two future scenarios (RCP2.6, the black dot, and RCP6.0, the red dot). The temperature is the mean value over growing season.
Figure 14: The response of light use efficiency (LUE), leaf area index (LAI) and growing season length (GS) to CO₂ concentration and temperature (T). Using the climate and CO₂ measurements from YuCheng 2005 as a baseline condition. (a) The response of LUE change to CO₂ increment. (b) The response of LAI change to CO₂ increment. Solid line includes the effect of CO₂ on LMA; dash line excludes this effect. (c) The response of LUE to temperature. (d) The response of changes in growing season length to temperature increment.

3.5.4 Comparison with future scenario runs by other crop models

We compared future scenarios with the PC model to ISIMIP2b model runs performed with the same scenarios using complex crop models. PC and LPJmL simulated contemporary yields reasonably well across all the sites, but the PEPIC and GEPIC models showed unrealistically low yields except at site CS. PC showed an increase in wheat yield ~6.6% in the RCP2.6 and ~15.1% in the RCP6.0 simulations by the end of the 21st century (Fig. 15). Although the different crop models predict different absolute magnitudes of wheat yields, the trend and the interannual variations are similar among all models. Moreover, the magnitude of increase shown by PC is similar to that shown with LPJmL. All models showed increases in wheat yield at individual sites over the 21st century, with the exception of the PEPIC model at the YG site.
Figure 15: Comparison of different crop models: future scenarios at six sites. The climate data to drive the crop models were derived from the MIROC5 climate model. Lines represent modelled interannual yield trends; Points represent measured yields.

4. Discussion

The P model has been extensively tested against GPP derived from flux measurements in natural vegetation (Stocker et al., 2019; Wang et al., 2017) and has also been shown to perform well in simulating the GPP of croplands (Balzarolo et al., 2018). The present study has confirmed that the P model can predict the GPP of irrigated and fertilized wheat crops in China (Fig. 3); that above-ground crop biomass can be modelled as a fraction of accumulated GPP (Fig. 4); and that yield can be modelled as a saturating function of AB (Fig. 5). Further extensions and tests of the model in a wider range of environmental and economic settings will be needed to allow application to model wheat crops under water and/or nutrient stresses (which are expected to result in different relationships among GPP, AB and yield), or in a wider range of climates.

The ratio of above-ground biomass production to GPP is typically 0.41 (natural) to 0.53 (managed) for forests, ~0.2 for natural grasslands and 0.6 to 0.7 for managed grasslands (Campioli et al., 2015). Values of this ratio, up to ~0.8, have been found for intensively managed crops (Campioli et al., 2015; Chen et al., 2018; Huang et al., 2018). The value of 0.72 estimated in our study is in the upper part of the expected range. This is not unreasonable. As an annual crop, wheat does not require strong roots for support. In addition, the study sites are irrigated (to eliminate water stress) and fertilized (to reduce or eliminate nutrient stress), so the below-ground carbon allocation needed to acquire nutrients and water is minimal. Modern varieties of wheat are highly efficient in converting GPP to biomass because selective breeding has aimed to increase the GPP allocation to biomass, and ultimately to grain.

Most crop models assume that grain yield is a fixed proportion of above-ground or total biomass, the so-called harvest index (HI) (Donald, 1962; Hay, 1995). However, the grain yield data analysed
here follow a saturation function with AB (Fig. 5), so that HI declines with increasing AB. The level of AB at which saturation occurs is largely determined by the wheat variety (Fig. 6 and Fig. S1). The maximum yield given by \((a \cdot N + b)\) in equation (5) is influenced by the wheat variety and the amount of nitrogen added. The actual yield is also determined by the amount of biomass accumulated and, therefore, by the GPP during the growing season – which depends on CO\(_2\), climate, canopy development and incident PPFD. The negative intercept \((d)\) quantifies the requirement for a certain minimum biomass accumulation before any carbon is allocated to seeds. When a linear regression was fitted instead of a saturating function, the estimate of \(d\) became positive, which is biologically impossible as it suggests a positive yield when AB is zero (Fig. S2).

It follows from this simple empirical representation of the experimental data that improving grain yield is not simply a case of adding more fertilizer (which also comes with significant monetary and environmental costs). Moreover, yields will also not automatically increase in proportion to the effect of CO\(_2\) on photosynthesis, because the saturating nature of this relationship implies a “diminishing return” on increases in AB. The differences among varieties are potentially important here. They suggest that a key target for crop improvement should be the ability of the plants to allocate more carbon to grain as AB increases, and thereby to profit from higher CO\(_2\) levels.

Both the current level and the trend in yield over the 21\(^{st}\) century simulated by the PC model are similar to that shown by the LPJmL model (Fig. 15). This similarity is probably due to the fact that the LPJ component of LPJmL, is also, ultimately, based on the standard model of photosynthesis and the acclimation of \(V_{\text{cmax}}\) – the latter process now supported by a wealth of evidence (Smith et al., 2019). However, the PC model is simpler, has fewer parameters and is more transparent, with major advantages both for the credibility of the results and the ease with which uncertainties can be quantified and traced to their source.

Quantification of prediction uncertainties in complex models requires extensive computation to estimate the sensitivity to their many parameters. In contrast, the PC model consists of a single central
equation (1a), which can readily be differentiated with respect to its (far fewer) parameters. This process allows uncertainties to be attached to predictions without excessive computational demands and allows the major sources of uncertainty to be pinpointed. We have shown that the parameter $c^*$, related to the metabolic costs of maintaining electron transport capacity, and parameter $b$, related to the potential maximum yield, accounts for a large fraction of the prediction uncertainty (Fig. 10) – indicating the importance of further work to improve these aspects of the model.

Studies have suggested that rising temperatures could greatly reduce the grain yield of wheat (Asseng et al., 2014; Asseng et al., 2011; Zhao et al., 2017; Zhao et al., 2016), because of the shortened growing season. However, many studies have neglected the effects of rising CO$_2$ on C$_3$ photosynthesis, which has the potential to mitigate the impact of rising temperatures on production by improving LUE, particularly as the temperature optimum shifts to higher temperatures with rising CO$_2$. The effects of elevated CO$_2$ have been shown by Free Air Carbon dioxide Enrichment (FACE) experiments, with positive impacts on wheat yields and net assimilation rates (Broberg et al., 2019).

The primary mechanism by which increasing CO$_2$ increases GPP is through the improvement of LUE. This mechanism is amplified by the positive feedback by which increased GPP allows greater LAI, which in turn implies greater light absorption and GPP. On the other hand, LMA increases with CO$_2$, resulting in a diminished response and, according to our model, a peak of the positive response of LAI to CO$_2$ above ~600 ppm (Fig. 14b and Fig. S4). However, we see no peak in the positive response of yield to CO$_2$. This appears to be because $f_{\text{APAR}}$ is comparatively insensitive to changes in LAI at the high end. However, according to our simulations, a positive but saturating response of yield and LUE to CO$_2$ are found at high CO$_2$ levels (Fig. 13 and Fig. 14a). This appears to be inconsistent with the yield response shown in the meta-analysis of enhanced CO$_2$ experiments by Broberg et al. (2019). However, the response shown in that paper is small, solely derived from chamber rather than FACE experiments, and seems to reflect the reduced sensitivity of higher yield wheat varieties to CO$_2$ changes.
The magnitude of yield enhancement simulated by our model is consistent with the findings of Broberg et al. (2019) for the mid-range of wheat yields. Field warming experiments, summarised in Zhao et al. (2016), show negative responses of yield to warming of between 0.5 to 3.0°C at individual sites in northwestern and northern China, with an average response of −4.4% per °C in northwestern China to −2.8% per °C in northern China. Across our study sites (which are in the same region), we predict a net negative response to increased temperature of between −2.3 and −5.7 per °C. This response is caused by the reduction in the length of the growing season and the negative impact of temperature on LUE. The modelled response of yield to combinations of raised CO₂ and temperature, as shown by sensitivity analysis, reflects a combination of net positive effects of CO₂ and net negative effects of rising temperature. Scenario runs show that, under the scenarios tested, the positive effects of CO₂ on yield however outweigh the negative effects of temperature, consistent with the findings of other crop models for China (Liu et al., 2019). Warmer regions are more sensitive to warming than cooler regions (also consistent with Liu et al. (2019)), indicating that wheat production in warmer regions of China will be more challenged by climate change.

5. Conclusions

The yield of irrigated and fertilized wheat crops at research sites across the wheat-growing region of China was simulated successfully using a parsimonious model based on a combination of first-principles theoretical considerations governing GPP with empirical analyses of the relationships amongst GPP, AB and yield. The model reproduced the general magnitude and patterns of interannual variability in both AB and yield. When driven by future climate and CO₂ scenarios, it produced results similar to those of the most credible of the more complex crop models.

The model also provided insights into how wheat yields may respond to global environmental change. The effect of rising CO₂ on photosynthesis does not imply proportionately increased yield. The model results suggested that the positive response of yield to rising CO₂ may saturate at around
600 ppm. The model also predicted a negative effect of warming on wheat yields. Sensitivity analysis showed this negative effect to be stronger in regions with warmer climates today. Nonetheless, in common with other crop models, the simulations indicated an increase of ~6.6% in wheat yields under the RCP2.6 and ~15.1% under the RCP6.0 scenarios of future CO₂ and climate.

Acknowledgements

The authors thank Huimin Lei and Dawen Yang for providing the data at the WeiShan site, and Yiqi Luo, Kun Yang, Chaoqing Yu and Zaichun Zhu for their comments on the analysis. This research was supported by the National Key R&D Program of China (no. 2018YFA0605400), National Natural Science Foundation of China (no. 31600388). ICP and SPH acknowledge support from the High-End Foreign Expert programme of the China State Administration of Foreign Expert Affairs (no. 161207002). SPH acknowledges funding from the European Research Council (ERC) for “GC2.0: Unlocking the past for a clearer future”. This research contributes to the AXA Chair Programme in Biosphere and Climate Impacts and the Imperial College initiative on Grand Challenges in Ecosystems and the Environment (ICP). ICP also acknowledges funding from the ERC, under the European Union’s Horizon 2020 research and innovation programme (grant agreement No: 787203 REALM).

References


