Reply to the community comments #1

We are grateful to Owen Cooper for reviewing the manuscript and for providing the comments. We improved the manuscript based on these suggestions.

Line 34 When citing IPCC it would be best to cite the most recent update from 2021.

Line 34- 36 We have now cited the updated IPCC report reference to the manuscript as:

'At the same time, climate change is considered a substantial threat to arable productivity through changes in average and extreme temperature and precipitation profiles across the region (IPCC, 2021)'

Line 55 When discussing the differences between exposure metrics (e.g. AOT40) and the stomatal ozone flux method, a useful reference is Ronan et al., 2020.

Line 58-60 We have now added the reference to the manuscript as:

'At present, methods to assess the risk to crop productivity from changes in O3 and climate variables use a variety of different O3 risk assessment methods (Ronan et al., 2020) and crop models as discussed in depth in Emberson et al. (2018).'

Reply to referee comment #1

This paper presents a modeling study to evaluate the impact of ozone damage on Chinese wheat yields using the developed DO3SE-Crop model and field experimental data. Overall, this is a comprehensive study that provides valuable insights into the effects of ozone on food security. However, there are areas that could be improved.

The title of the paper could be modified to reflect a broader scope of the model's purposes and applications, rather than focusing solely on a specific site in China.

This is a really good suggestion - thanks. We have now changed the title to 'Development of the DO_3SE-crop model to assess ozone effects on wheat crop phenology, biomass and yield'.

A primary concern is that only data from 2008 was used to train the model. This raises questions about the representativeness and robustness of the model when applied to other years and regions.

We trained the model on 2008 data, and then tested the model on unseen data from 2007 and 2009 specifically to test the ability of the model to capture variability between years. Since we found that the model was able to explain 7.9-8.7% (Line 881-884) of the variation we judge this to be a good indication that the model is indeed able to capture between year variability reasonably well. However, we agree that the application of the model to other regions is untested, due to the fact that data from other sites/regions are simply not available. We have made both the description of the calibration and evaluation of the model clearer by substantially re-writing section 2.2 and also make the issues associated with the model's regional representation clearer in the text (Line 929-934) suggesting that this requires future investigation as new datasets become available.

The extremely high correlation between the modeled and observed yields (R^2 of 0.99) suggests potential overfitting, which could lead to biased prediction results. The authors should address this issue to be more convincing.

Line 256-262 We have made an update to the model which now accumulates stomatal ozone flux (acc_{fst}) from only 200°C days before the simulated timing of anthesis. This update was made to accurately capture ozone uptake at a time when the canopy was fully established. Correlations between modelled and observed yields for this updated model are now an R^2 of 0.92 (n=4) and 0.68 (n=20) for the training and test data respectively. Since we would expect a higher correlation for the training dataset (since we calibrate to fewer data points) and since the test dataset shows a reasonable fit to a larger set of unseen data, we believe this provides the evidence that that model is not over-fitting. We make this clearer in the discussion (where previously we have only reported results for the trained data) so that we now say - Line 881-884 'Importantly, when applied to the test dataset (i.e. excluding 2008 data for the Y2 and Y16 cultivar) the model, was found to simulate the grain dry matter under ambient and elevated O_3 treatments to within 7.9-8.7% of the observed values (R^2 =0.68, 76 g/m^2 see Fig. 9)'

It appears that the ozone-induced yield losses derived from field observations are significantly larger than the corresponding simulation results. The reasons behind this discrepancy should be investigated and discussed.

We find that we tend to underestimate the O_3-induced relative yield loss (RYL) by between -2.76 and 15.34 (observed less modelled RYL) across all years and cultivars. Further analysis of these data (see new tables S2a. and b.) show that average differences between observed and modelled RYL estimates for all cultivars are similar between years (ranging from 4.94 to 6.73) but that average differences between cultivars are more variable with the sensitive cultivars (Y2 and Y19) ranging between 5.02 and 9.0; and tolerant cultivars (Y16 and Y15) ranging between 2.66 and 5.54. This would suggest that O_3-induced yield losses can be more reliably modelled for tolerant cultivars which may suggest that additional processes causing O_3-induced yield losses in sensitive cultivars are not captured, such as O_3 altering the allocation of C to different plant parts (Feng et al., 2008) or O_3 inducing additional respiratory costs via the upregulation of defence mechanisms (Biswas et al., 2008).

To address this, we have added text and two tables in the supplementary (section S6; text and Table S2a. and S2b) giving the explanation above and have added a short paragraph summarising these findings to the end of the results section at Line 775-784 and the beginning of the Discussion section at Line 790-801.

In the methods section, the authors cite numerous related studies for the formulas of different modules. While the sensitivity of these empirical parameters has been evaluated, the uncertainties arising from the process need to be elaborated upon.

We believe this comment relates to the fact that although we focus attention on defining the values (and ranges) of key parameters, we implicitly assume that we have included all, or at least most, of the important model processes that determine O_3 effects on crop development, growth and yield. To address this, we include mention of the various ways in which O_3 can impact crop growth and yield (referencing a seminal paper by Feng et al. 2008 that performed a comprehensive meta-analysis for wheat) in the discussion to show that we are aware there are additional processes not included in DO_3SE-Crop that might influence crop responses to O_3 which could form the basis of future studies.

We also refer to some of these 'missing' processes perhaps being responsible for the underestimation in RYLs, especially for sensitive cultivars (see point above)- please see Line 796-805.

We also appreciate that although we have identified the key parameters through application of a sensitivity assessment, we have not included any uncertainty analysis (e.g. the certainty with which we can define a range of values for each parameter) and how any uncertainty in these values and ranges may influence results. We note this in the methods section at Line 562-563 (in Section 2.2) but state it is outside the scope of this paper by including the following sentence.

Line 562-563 '*We note that assessing the probability distribution of these ranges would also be useful but consider this outside the scope of the current paper due largely to data limitations*.

Uncertainties may arise from assumptions, such as "We then assume that these values are consistent across cultivars and years" (Line 580). These need to be addressed in more detail.

We now make clear in the text that assumptions that parameterisation will work well for different cultivars and years need to be made with some degree of caution by including the following text at Line 623-627 in the section on crop phenology.

Line 628-632 '*We then assume that these values are consistent across cultivars and years. Figure 3 suggests this is a reasonable assumption since the phenology module captures the timing of anthesis and harvest for unseen cultivars and years within 2 to 4 days and 1 to 6 days respectively, of the observed timings. However, we appreciate that assuming these phenology parameters will work for a wider variety of cultivar types (e.g. early or late sown and/or maturing) and years with rather different meteorological conditions needs to be done with caution*'.

The figures in the paper could be better presented to improve visualization and readability. For instance, the colors of different legends in Figure 3 are difficult to distinguish.

Line 683: This figure (now Figure 5) has been updated for better visualization and readability.

The large differences between the modeled and observed Anet and gO3 in Figure 4 need to be explained, as the model tends to underestimate these values. Additionally, correct the figure caption for panels c and d, as they appear to be swapped. Ensure consistency in font sizes, such as in Figure 5c and d.

We have provided an explanation for the discrepancy between the modelled and observed Anet and gO3 values in Figure 4 by including the following sentence at Line 651-663.

'In Fig. 4a and b, the steep decline in modelled A_{net} and g_{O3} is not seen in the observed dataset. This discrepancy may occur since the simulated A_{net} and g_{O3} values represent sunlit parts of the upper canopy which comprise both green and senesced leaf material. In contrast, observed A_{net} and g_{O3} values are measured specifically on the flag leaf and most likely for only the green parts of the leaf and most likely only for the green parts of the leaf, since the LI-6400 photosynthesis system mounted with a 6400–40 leaf chamber fluorometer (used to measure A_{net} and g_{O3} in the Xiaoji experiment, Feng et al., 2016) will not provide values for senesced leaf material. See also Figure 4 which combines A_{net} and g_{O3} with observed relative chlorophyll content and clearly shows the leaf is senescing as predicted by the model.'

Minor Comments

Correct typographical errors in the abstract, such as "...regions of ufor risk assessment."

Line 11-18 - Corrected the typo. This now reads: -

'A substantial body of empirical evidence exists to suggest that elevated O3 levels are causing significant impacts on wheat yields at sites representative of highly productive arable regions around the World. Here we extend the DO3SE model (designed to estimate total- and stomatal-O3 deposition of for risk assessment) to incorporate a coupled Anet-gsto model to estimate O3 uptake, an O3 damage module (that impacts instantaneous Anet and the timing and rate of senescence), and a crop phenology, carbon allocation, and growth model based on the JULES-Crop model. The model structure allows scaling from the leaf to the canopy to allow for multiple leaf populations and canopy layers'

In Line 50, O3 should be in subscript form.

Line 52 - Corrected the typo.

In Line 104, "Evaluation" should be in lower case.

Line 16 – Changed the heading.

Ensure references are complete and correctly formatted, including publication year, journal name, etc. For example, the reference to Yang, L. et al. should include all pertinent information.

All references have been checked for formatting.

Add missing parentheses in Figure 5b.

Line 715 - Parentheses added in the 'now Fig 6'.

Reply to referee comment #2

The manuscript "Developing the DO3Se-crop model for Xiaoji, China" written by Pande et al. used the existing measured data of winter wheat under elevated and ambient free-air chamber condition from Xiaoji from 2007 to 2009 to parametrize and develop/extend their current ozone deposition model -DO3SE.

The MS contains many error of spacing. Line justification and larger spacing would be required to support better reading and evaluation of the MS. These have now been corrected. The MS was not in balance between introduction, material and method and results sections. The balance has now been redressed. Introduction is still minimal which needs further elaboration and literature review the recent developments of crop models (which includes ozone) and it needs an explanation why DO3SE needs to include crop modeling routine. This is now included. Materials and methods were heavy with modeling description and equations which at some points one do not know which processes need to be focused. This is the first time a full version of the DO3SE_crop model has been described so all processes are important. Also, which field data supports the parameterization? The calibration and validation and sensitivity analysis were mixed up and not well described which give an impression that the model was over fitted of grain yield simulation based on many parameters. The calibration and evaluation section has been completely re-written. The results section did not follow the objectives mentioned in the introduction. The objectives of the paper are now consistent between sections.

While the MS emphasized the different cultivars responses. The simulation steps and results for these are not clear. The steps are now clarified

Many detailed comments are listed below.

Introduction

Line 55-65 needs to be more elaborated about state of art in crop modeling with ozone routine.

A short paragraph has been added at Line 66-76 describing the current status of a new generation of crop models that include ozone damage modules and the rationale for the development of the DO3SE-Crop model.

Line 94: is there any modeling comparison in the MS e.g. JULES and DO3SE-crop?

There are no direct comparisons between the JULES and DO3SE-Crop models; we make this clearer by modifying the text at Line 105-108 to that given below. The only indirect comparison is that provided in Figure 6a which compares the carbon allocation profiles simulated by each model.

Line 105-108 '*This would in the future allow comparison of the UK JULES Crop model which are based on O3 mechanisms that modify instantaneous A_net to mimic changes in yield consistent with flux-response relationships (Sitch et al., 2007) with the alternative O3 damage mechanisms used within DO3SE-crop*'.

Materials and methods

Fig1. There is no connection of input climate variables to photosynthesis and crop growth. Do ozone and thermal time drives the crop biomass growth? There is no legend or further caption to describe the information inside the diagram.

We have modified Fig1 so that meteorology and O3 data are in the same box and both feed into the resistance scheme which estimates Rsto, LAI, FO3 to calculate O3 impacted Anet and gsto.

Table 1. The table was rather simple and less informative since one does not know the real, absolute values of climatic and ozone variables. How much are different among treatments, year and growing season in term of input variables. Fumigation time? Lack of summary on climatic input and ozone that is really hard to understand the crop responses and modeling performance.

Table 1 has been updated to include minimum, average and maximum climatic variables by year and the M7 ozone metric calculated over the exposure (fumigation) period.

Line 512. Did the trial measure the gO3?

Good point, the trial only provided stomatal conductance as conductance to water vapour (i.e. gH2O), this is now made clear in the text, Line 298-301, along with a reference to the supplementary table (S1b) which describes the conversion factors from gH2O to gO3 to match the model output.

Section 2.2:

There is a mixed and unclear description how models were tested. The sensitivity analysis should be done firstly to screen the most important parameters, then coming to calibration, then validation. Where is the validation description, which treatment, modeling performance metrics, which outputs were validated?....

Many thanks for these comments on section 2.2. We agree this section was confusing and have thus re-written this entire section to clearly describe the steps for sensitivity analysis, model calibration and model evaluation. We believe this is now much clearer and addresses all the points raised below.

Line 533-543: It means that sensitivity analysis was not done for phenology? There are 11 parameters and three outputs, how the calibration is done in sequence to simulate the different developmental stages. It is not clear.

The sensitivity analysis was not performed on the phenology parameters since earlier studies have shown these parameters are relatively straightforward to calibrate using automated methods for a range of environmental conditions (Nguyen et al. 2024). We have now added text to this effect at LN 559-562.

The calibration of the phenology parameters was performed for all 10 parameters simultaneously to find the best fit defined as the highest R2 value and lowest RMSE value when modelled estimates of flag leaf emergence, anthesis, and harvest are plotted against observations. This is now clearly explained in the text at Line 569-577.

Section 2.1. Line 504-513, it is not clear with the measurement data, was there significant impacts of ozone among ozone treatment and interactions with cultivars? Why 2008 was selected for calibration? If there is no considerable impact of ozone on crop physiological processes, the modeling parameterization is very weak. At least, this point is hardly seen in the description of data.

The description of the measurement data has been re-written and now explains why 2008 data were used to train the model, including details of availability of crop physiological data for evaluation and substantial yield effects between ambient and elevated ozone treatments for calibration of the ozone damage model. See LN 509-531.

Line 544-562: It is really unclear for the calibration process? How the calibration was done manually because there are a lot of parameters and outputs were involved. Line 549, in which steps? How parameter were changed? What are key processes should be considered for which treatment (ambient and elevated ozone). If the JULES model was already used and well calibrated why the newly crop model (DO3SE-Crop) needs to consider some other processes here (e.g. phenology, carbon allocation…). Did observed data really prove the change of carbon allocation to different organ due to ozone? (line 646-649). The description seems the many processes have been touched and kind of overfitting with using many parameters to get correctly simulated final biomass and yield?

Section 2.2 now provides a clear, step-wise, description of the calibration process.

We also make clear that the JULES model calibration of C allocation parameters has only been performed for broad, global scale application for wheat (Osborne et al., 2016) and therefore requires further calibration for application under Chinese conditions (see Line 596-598).

The observed dataset does not provide any information with regards to the change in carbon allocation parameters due to ozone. The C allocation parameters were only calibrated for ambient ozone conditions, this is now clear in the calibration description. We have also modified the text at Line 598-601 to make clear that we investigate the effect of ozone on C assimilation (not C allocation).

There is measurements which has been described from 504-518, but this was not clear how these measurements that were used for the calibration. It might be better to list out the key measured variables, which one is used for calibration. Xhu et al., 2011 is not in the literature?

As stated above section 2.2 has been rewritten so that it is now clear which measurements are used for which step of the calibration process.

Line 517 Corrected the reference typo by changing Xhu et.al., 2011 to the correct reference Zhu et.al., 2011.

Line 551: Why the vcmax and jmax need to be calibrated again? Why the measured vcmax or jmax could not be used directly for the models to simulate Anet and gs?

We calibrated Vcmax and Jmax because measurements are only provided for Y2 and Y16 cultivars and only for certain points during the growth period and we know that Vcmax and Jmax can vary seasonally in the text at Line 586 section 2.2. However, we find that the calibrated Vcmax and Jmax values match the observed values and now make this clear in the text at line 662-633.

Line 558: why the LAI needs to be between 4-7 m2 m-2. Did it mean that there is no measurement of LAI? Recent work which compared three models including from DO3SE-Crop model in Nguyen et al., (2024), it seems that the DO3SE-crop model did not perform well for the green leaf area simulation. LAI is the most important crop growth metric which influences the simulating assimilation, stomatal conductance (thus ozone fluxes, see diagram fig 1) at leaf and canopy scale which in consequence affects to leaf senescence due to ozone, then thus biomass. It is not clear how DO3SE-Crop model improves the LAI simulation.

We recently got the LAI data from the Xiaoji experimental research group. We now plotted our simulated LAI values against these observed values (see Fig S2) and can demonstrate good agreement. This improvement in LAI simulation from the DO3SE-Crop version used in Nguyen et al. (2024) is due to an updated parameterisation of the dark respiration coefficient (Rdcoeff) (which was earlier set too high since it was based on Clark et al., 2011).

Results

The result section did not follow the mentioned objectives or at least the headings were not clearly shown these. The objective 2 was not touched with regards of tolerance and sensitive cultivars. How the simulation/parameters for tolerance and sensitive cultivars were done?

We thank the reviewer for spotting this inconsistency. We have now updated the objectives for the paper given at the end of the introduction with the text below. We make clear the information that we are able to provide regarding the sensitivity of different cultivars which is supported by the improved description of the model's calibration and evaluation given in Section 2.2.

Line 109-115 '*The key objectives of the paper are to assess the ability of DO3SE-Crop to i). simulate key phenological stages, ii) the relationship between leaf-level physiological variables and within canopy O3 concentrations, iii) simulate C allocation to different parts of the crop and iv) to simulate O3 induced grain dry matter losses for tolerant and sensitive cultivars caused due to instantaneous versus long-term senescence effects on photosynthesis*'.

Line 580-585: This is important. It is not clear how to simulate onset of leaf senescence? Did the observed result show both the early onset of senescence and increase of senescence rate due to ozone?

Line 624-628 We used the breakpoint method to determine the start of senescence and increased senescence rate due to ozone which was observed in the relative CCI data. This method is described in detail in Pande et.al., 2024 and we have now included a reference to this paper in the text along with a clear description of how this method was used in conjunction with the Xiaoji data to parameterise the phenology model. Fig 7a. and b. Clearly show the earlier drop in relative CCI for the elevated ozone treatment compared to the ambient ozone treatment.

Figure 3: It is not clear. The duration of flag leaf period means that from flag leaf appearance to dead leaf? The line 604: this Fig3c was wrong. Line 614 with Fig. 4 has the same issue. Data for which day for which cultivar? It is weak to show the simulated value and where are the measured data? Fig 4c &4d for the same AA treatment...Why the modelled and simulated were so much deviated? Where is error bar of measurement?

We have now clarified the phenological period over which the flag leaf data are shown (i.e. from the start of anthesis to maturity i.e., TT_{rep}) in (what is now) Figure 5 and Figure 4.

We have corrected the figure numbers referring to the (now) Figure 5 in the text.

We have removed the original Figure 4a. and b. (showing the diurnal values) from the text as we agree that this figure adds little without observed values (which are only included in Fig 4c. And d.).

The experimental dataset doesn't provide the hourly measured value for gO3 and Anet.

Section iv) did the trial measure the dynamic change of dry matter in different organs or only just final biomass and yield? Figure 5a, what is harvest? 5b? Result from which models? Which treatment and cultivar? Really confusing and unclear.

The trial (i.e. measurement data) only provided data on final yield (i.e. grain dry matter). Figure 5 (now Figure 6) title has been updated to clearly describe which year and cultivar the DO3SE-Crop model simulations have been performed for. We also make clear at Line 712-714 that the JULES model comparison is provided for illustrative purposes only (i.e. this model has not been calibrated with the Xiaoji data but rather is a parameterisation suggested for global application).

Figure 6. Relative to what? Was the CCI absolute? It is really hard to understand why the slope of fLS in ambient was already very steep as similar in elevated treatment? Model overestimated the drop of leaf senescence?

The relative CCI is calculated using the maximum CCI value for the ambient ozone treatment. Yes, the observed dataset gives absolute chlorophyll values. These had been converted into CCI values (after Reference), but we now realise that it is better to use normalised chlorophyll values as it will be more appropriate as the experimental study gives absolute chlorophyll content in mg/m^2. The plots have been updated to reflect this (Fig 7a and b).

We have modified the model so that ozone is accumulated (acc_{fst}) 200 degree C days before anthesis to be consistent with the UNECE Mapping Manual accumulation period (LRTAP, 2017) and to provide greater consistency in acc_{fst} when applying at regional scale by defining an easier to identify phenological accumulation period. This is now made clear at Line 256-262 in the methods section.

Line 747-750 The slope of the ambient f_{LS} is already steep since the ambient treatment already has rather high O3 levels as is now made clear in Table 1 with a value of 47ppb. According to the M7 wheat dose-response relationship this would result in a yield loss of ~ 5%.

Fig7. The section emphasized the seasonal variability but the factor of year was not shown. Why the data was not displayed for each year? Where is error bar of measurement?

We have now updated the Figure (now Figure 8) to show boxplots of grain dry matter for simulated vs observed split out by year and cultivar. The box plots now clearly show the range in the observed as well as the simulated.

Line 688-699 One could not see from the figure? What is reason why ozone fumigation was too short? Why the measured data show bigger yield reduction in ozone in 2007 compared to other years? Problem of measured data?

As described previously we have modified the model so that the ozone accumulation period (acc_{fst}) is between 200°C days before anthesis to maturity for consistency with (LRTAP, 2017). This means that the shorter experimental fumigation period for the year 2007 (38 days in 2007 vs 92 days for 2008 and 2009) has less effect on the model results since acc_{fst} affecting senescence (the main driver of yield loss) is now far more consistent across years. These differences in RYL between years are now discussed in more detail (also in response to other reviewer comments) by adding text and two tables in the supplementary (section 6; text and Table S2a. and S2b) and a short paragraph summarising the key findings to the end of the results section at Line 775-784 and the beginning of the Discussion section at Line790-805.

Figure 8: for both ambient and elevated?

Figure 8 (now Figure 9) has been updated to show results by AA and E O3 treatment, for each year and by cultivar.

Discussion

L697: it is not clear to see the distinction of simulated tolerance and sensitive? How model has configured and related results? The range was almost similar for two groups of cultivar? Is this significant?

We now clearly describe the model results in relation to sensitive and tolerant cultivars. We explain that the model is less good at capturing some of the more extreme sensitivity to O3 that occurs in the sensitive cultivars and that this could be due to O3 effect processes that might be more common or important in O3 sensitive cultivars not being included in the current DO3SE-Crop model construct, see Line 790-801.

Line 702-703 it is not clear. Line 706: how about the role of phenology? Because model overestimated phenology?

The modified version of the DO3SE-Crop now sees much closer agreement between the fst_acc of the EE O3 treatments between years (including the shorter fumigation year 2007) therefore we no longer see the substantial model overestimates in grain dry matter for 2007, the text is changed to reflect this. We also show that phenology is modelled well using the DO3SE-Crops thermal time model (see new Figure 3; R2 of 0.95 and RMSE of 2.5 days when comparing modelled and simulated phenological stages) so this is unlikely to be a cause of any inconsistencies in results between years.

Line 710 and elsewhere, check unit

The units given (kg/ha) are all correct.

Line 721: This is not true. Model overestimated for 2008-2009 at least 15-20 days?

Please see Fig. 3; the modelled simulated the phenological stages within 2-4 and 1-6 days for the modelled anthesis and maturity when compared to the observed dataset. It has been identified that the previously reported data for anthesis and maturity contained inaccuracies.

Line 737-747. It is confusing and need clarification also for the MM and result e.g. Fig 1 & Fig2. The model simulated the ozone uptake for only flag leaf (top canopy) and divide these to different layers (Gaussian integration over flag leaf depth) or the whole canopy and divide these to different layers (Gaussian integration over LAI depth)?

The DO3SE-Crop model simulates stomatal ozone uptake by canopy layer (illustrated by a nominal 4 canopy layers in Figure 1). This is possible since the model is able to estimate changes in irradiance (described in section 1.3.1) and ozone concentration (via the resistance scheme described in section 1.3.2) with canopy depth, these estimates are then used to simulate stomatal ozone flux (as a function of stomatal conductance and boundary layer resistance) with canopy depth. The model then sums stomatal ozone flux, weighted by canopy layer sunlit and shaded LAI fractions, to estimate total canopy stomatal ozone uptake. This is described in section 1.3. The current application of the model uses 1 representative leaf population so essentially assumes this representative leaf represents the canopy average in terms of emergence, vegetation growth, anthesis and maturity as described in Figure 2. We now make this clear in the Figure 2 title as already described in section 1.1 at Line 182-193.

Line 750: this is not proved. I did not see the observation of Anet and gO3

Thanks for spotting this - we have removed this sentence.

Line 757: value of this vmax is really high while the yield was not super high. Is there any explanation for this?

The Vcmax value range of 90 to 140 umol CO_2 m^{-2} s^{-1} came from a combination of the published literature and Xiaoji experimental dataset (see AppendixA Table A2). The optimal value after calibration is 137 umol CO_2 m^{-2} s^{-1} which is in line with the observed dataset. Moreover, the complex interactions between O_3 exposure and the plants' physiological responses also play a crucial role.

Ozone significantly affected antioxidative enzymes, thereby limiting overall photosynthetic efficiency and yield, particularly in O_3-sensitive cultivars, despite their ability to maintain high carboxylation capacity. These points are at Line 863-870.

Line 765: There is no data of measured carbon allocation. Why one need to change the JULES configuration? Line 769: It is not clear. Did measurement include respiration, LAI, stem biomass data? Reply only the final biomass but calibrating many parameters means that this is an overfitting of the model.

The JULES C allocation was developed for global application; we trialed the JULES parameters for the Xioaji environmental conditions, and it was clear that further calibration of these parameters would be needed to allow wheat crop growth and yield to be simulated for these conditions. This was confirmed with discussion with the JULES crop model developers. In the absence of experimental data detailing key variables that would ideally have been used for this calibration we have had to define target values (as described clearly now in Section 2.2) along with the phenology and yield data that we do have from the experimental dataset. Since we only use target values (to ensure a well-proportioned crop plant; Line 877-884) we would not be overfitting the model.

Line 770-771. And mean what?

We hope we have addressed this point in the reply above.

Line 801, it is not true based on the slope in Figure 6a.

Line 651-663 We have now combined the A_{net} and g_{O3} with observed relative chlorophyll content to show that the leaf is senescing as predicted by the model. We further explained the reason behind the discrepancy between the observed and modelled A_{net} and g_{O3} . Line 652-663- 'This discrepancy may occur since the simulated A_{net} and g_{O3} values represent sunlit parts of the upper canopy which comprise both green and senesced leaf material. In contrast, observed A_{net} and g_{O3} values are measured specifically on the flag leaf and most likely for only the green parts of the leaf and most likely only for the green parts of the leaf, since the LI-6400 photosynthesis system mounted with a 6400–40 leaf chamber fluorometer (used to measure A_{net} and g_{O3} in the Xiaoji experiment, Feng et al., 2016) will not provide values for senesced leaf material.'

Line 802-803. I could not see for other validated years and cultivar. Data was only in Fig 6 which is weak.

The Chlorophyll dataset is only provided for the year 2008 and the Y16, and Y2 cultivar. We have now added the figure for Y16 (Fig.7) and Y2 (Fig. S4). This limitation is acknowledged, and the data presented in Figure 7 and Figure S4 serves as a representation of that specific dataset. Further validation with additional years and cultivars was not available in the study, which is why only the 2008 data for Y2 and Y16 was included in these figures.

Supplementary materials

Are the appendix S1a similar to Figure S1…?

No, Appendix S1a refers to a supplementary section (containing additional text, data, or methods), whereas Figure S1 refers to a supplementary figure, included in the supplementary materials.

While the calibration results looked very well, the validation of maturity was not good, mostly overestimated by model, almost which are too much. This inaccuracy of maturity might cause the simulated leaf area and biomass even more than the ozone impacts. The model must be improved phenology simulation before taking consideration of ozone impacts.

The modelled simulated the phenological stages within 2-4 and 1-6 days for the modelled anthesis and maturity when compared to the observed dataset for the test dataset (see Fig. 3). It has been identified that the previously reported data for anthesis and maturity contained inaccuracies.

Even here is supplementary material, legends were very confusing. What is all models phenology? Use consistent the term calibration for training, evaluation for testing. I did not see the emergence symbols.

We have now updated terms so that the datasets included in any analysis are clearly described. Since we would expect a higher correlation for the training dataset (since we calibrate to fewer data points) and since the test dataset shows a reasonable fit to a larger set of unseen data, we believe this provides the evidence that that model is not over-fitting. We make this clearer in the discussion (where previously we have only reported results for the trained data) so that we now say - Line 863-866 'Importantly, when applied to the test dataset (i.e. excluding 2008 data for the Y2 and Y16 cultivar) the model, was found to simulate the grain dry matter under ambient and elevated O_3 treatments to within 7.9-8.7% of the observed values (R^2 =0.68, 76 g/m^2 see Fig. S3)'

Figure S2: the model performance was perfect which raising question of overfitting (again) is or over calibration or the variation of observation was very minor. It is not clear about the units, where is root mean square error which have mentioned in the text, what are color of symbols, data from which plots?

We have updated Figure S2 (now as Fig S3 (training dataset) versus Fig 9 (Test dataset) to clearly show the distinction. This is also discussed in the text at LN 775-784 with the description of Figure 9.