

Original research article

## A framework for improved predictions of the climate impacts on potential yields of UK winter wheat and its applicability to other UK crops

Garry Hayman<sup>a,\*</sup>, John W. Redhead<sup>a</sup>, Matthew Brown<sup>a</sup>, Ewan Pinnington<sup>b,1</sup>, France Gerard<sup>a</sup>, Mike Brown<sup>c</sup>, William Fincham<sup>a</sup>, Emma L. Robinson<sup>a</sup>, Chris Huntingford<sup>a</sup>, Richard F. Pywell<sup>a</sup>

<sup>a</sup> UK Centre for Ecology & Hydrology, Wallingford, Oxfordshire OX10 8BB, UK

<sup>b</sup> National Centre for Earth Observation, Department of Meteorology, University of Reading, Reading RG6 6BB, UK

<sup>c</sup> UK Centre for Ecology & Hydrology, Lancaster Environment Centre, University of Lancaster, Lancashire LA1 4AP, UK

## ARTICLE INFO

## Keywords:

Climate change adaptation

UKCP18

Data assimilation

Earth observation

Precision yield data

## ABSTRACT

Changes in the frequency of extreme weather events related to climate change potentially pose significant challenges to UK agricultural production. There is a need for improved climate change risk assessments to support adaptation strategies and to ensure security of food production in future.

We describe an innovative and practical framework for spatially explicit modelling of climate change impacts on crop yields, based on the UKCP18 climate projections. Our approach allows the integration of relatively simple crop growth models with high spatial and temporal resolution Earth Observation datasets, describing changes in crop growth parameters within year and over the longer term. We focus on modelling winter wheat, a commercially important crop. We evaluate the results of the model against precision yield data collected from 719 fields. We show that the assimilation of leaf area index data from Sentinel-2 satellite observations improves the agreement of the modelled yields with those observed. Our national-scale results indicate that wheat production initially becomes more favourable under climate change across much of the UK with the projected increase in temperature. From 2050 onwards, yields increase northwards, whilst they decline in South East England as the decrease in precipitation offsets the benefits of rising temperature.

Our framework can readily accommodate growth models for other crops and LAI retrievals from other satellite sensors. The ability to explore impacts of crop yields at fine spatial resolutions is an important part of assessing the potential risks of climate change to UK agriculture and of designing more climate resilient agricultural systems.

### Practical implications

Agricultural activities are sensitive to both the long-term climate, governing the suitability of different areas of the UK for different crops and varieties, and to short-term weather events such as floods and droughts, which can affect crop growth and the optimal timing of sowing, harvesting and the application of fertilisers and pesticides. Agricultural activities will therefore be affected by climate change, which will alter both the long-term averages of meteorological variables and the frequency of meteorological extremes. The Agricultural sector needs reliable and timely climate

information on how changes in climate and weather events impact on crop variability and yields. Such knowledge can support those working in the agricultural sector, and inform those charged with climate adaptation planning and related decision-making. This knowledge can also help to both quantify the scale of adaption required and suggest viable adaptation strategies. Since the interaction between agriculture, climate and weather is influenced by decisions made at fine spatial scales (e.g. field to farm), it is important that such information is available at sufficiently fine spatial resolution to be relevant to these scales.

To generate the most reliable forward projections of impacts in response to a changing climate requires an understanding of the processes and mechanisms linking climate to impacts. This

\* Corresponding author.

E-mail address: [garr@ceh.ac.uk](mailto:garr@ceh.ac.uk) (G. Hayman).

<sup>1</sup> Now at European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, United Kingdom.

approach is more reliable than regression-based emulators when extrapolated to new weather regimes outside those that have been experienced historically in the region of interest. Here, we develop the CropNET framework to model the yield of winter wheat, a commercially important crop in the United Kingdom. Our approach combines a relatively simple crop growth model for winter wheat with driving datasets containing key meteorological parameters that affect crop growth (surface air temperature, precipitation and solar radiation). The CropNET model accounts for the different plant growth stages and the impact on such stages of water stress and water logging from periods of low or high rainfall, respectively. The crop model also accounts for the increased radiation use efficiency at higher atmospheric concentrations of carbon dioxide (the carbon dioxide “fertilisation effect”). This model produces “potential” yield, the maximum attainable yield if all other, non-climatic factors (crop nutrition, pests and diseases, etc.), were optimal. These potential yields are thus usually higher than those observed. The framework can then integrate satellite observations of crop growth (leaf area index), now available at high spatial and temporal resolution to adjust the modelled growth of wheat, based on historic and current crop growth trajectories. This adjustment brings the estimate closer to the “actual” yield, because the observed EO data incorporates the effect of factors such as pests, diseases and variation in farm management practices that are not directly modelled. These adjusted yields are thus directly comparable with those that are observed, such as field-scale precision yield measurements.

Using meteorological parameters taken from the latest climate projections for the UK, bias-corrected to present-day observations, we run the winter wheat crop growth model for a hundred-year period from 1980 to 2080. Potential yields are derived on the 1 km x 1 km grid covering Great Britain used by the input climate data. The model results show that wheat production initially becomes more favourable across much of the UK with the projected increase in temperature. However, from year 2050 onwards, whilst conditions in Scotland and the northern part of England continue to become more favourable for wheat production, there is a decline in yields in South East England as the reduction in rain offsets the benefits of rising temperatures.

These modelled yields were then used in a crop-yield simulation demonstrator, which was co-designed with users from the arable and livestock & grassland farming sectors. The demonstrator allows stakeholders to visualise and explore the local impacts of climate change on the yields of major crops. Currently, any location in Great Britain can be selected and the demonstrator shows how the modelled yields for that location are predicted to change over time (for the given scenario of future climate change), along with uncertainty arising from the different ensemble members. Comparisons can also be made with the surrounding regional or Great Britain average, providing important contextual information. Users can also input actual yield values to provide a location-specific correction to the modelled potential yields.

The CropNET modelling framework allows integration of other crop growth models and input meteorological and satellite data. The modelling framework can be extended to other crops if relevant non-wheat crop growth parameters are available. At the time of writing in 2023, there is a strong global interest in food security matters, especially against a backdrop of geopolitical instability in the Ukraine and the impact of heavy rain and floods on rice production in India. Hence, we offer a simulation framework that is potentially amenable to application at locations beyond the UK. Simulation capability becomes of particular use in assessing risks to locations where reduced grain imports may require higher domestic crop production, against a backdrop of a simultaneously changing climate.

## 1. Introduction

Climate change poses one of the greatest risks to future food

production both globally (FAO, 2018) and in the UK (Brown et al., 2016). Around 72 % (17.5 million hectares) of the UK land area is farmed and 35 % of this is productive arable land (Defra, 2020a). In 2019, the UK agriculture sector employed 420,000 people directly and generated Gross Value Added to the economy of £10.4 billion (Defra, 2020a). Whilst a warming UK climate may benefit some crops, it is likely that more frequent extremely hot summers and changing patterns of rainfall will result in an overall negative impact on crop production (Morison and Matthews, 2016). Similarly, increasingly wet autumns are likely to constrain agricultural production by adversely affecting the timing of farming operations. These may also indirectly result in increased risk of environmental damage, such as soil compaction and erosion. The wet winter of 2019/20, followed by a drought in the Spring of 2020, led to a 38 % reduction in UK wheat production from 16.2 million tonnes in 2019 to 10.1 million tonnes in 2020 (Defra, 2020b). Further, the UK average yield of 7.2 tonnes per hectare was lower than the five year average of 8.4 tonnes per hectare (Defra, 2020b).

Crop growth models have been used since the early 1960s to simulate the relationship between plants and the environment, to predict the expected yield for applications such as crop management and agronomic decision making, as well as to study the potential impacts of climate change on food security (Kasampalis et al., 2018). Crop growth models vary in type and complexity (Basso et al., 2013; Di Paola et al., 2016). Basso et al. (2013) identify three main types of crop growth models: (1) *statistical* models, where crop yield is related to key parameter(s) using linear/non-linear regression or Bayesian approaches; (2) *mechanistic* models, which explicitly represent many of the plant and soil processes (e.g., photosynthetic processes and uptake of carbon dioxide); and (3) *functional* models, which use simplified approaches to simulate complex processes, often drawing on insight gained from mechanistic models. Nearly all crop models require meteorological inputs (Hoogenboom, 2000). Additionally, parameters describing the different processes, crop varieties, soil conditions, management practices, etc., may also be required (Huang et al., 2019). All the models give quantitative predictions of crop growth (e.g., leaf area, above ground biomass, root biomass, soil moisture, grain yield, etc.) (Huang et al., 2019). Although the functional models are considered to be generally simpler, such models usually produce reasonable results when compared to field measurements (Basso et al., 2013).

The potential of Earth observation (EO) for crop monitoring was recognised from the launch of the first Landsat satellite mission in 1972 (Hammond, 1975; Hall et al., 1981). EO multispectral crop reflectance and crop surface temperature are linked to crop growth and yield through two basic plant physiological processes (photosynthesis and evapotranspiration), which in turn are determined by the crop canopy leaf area index (LAI), development stage and stress (e.g. Idso et al. (1977); Bauer (1985)). The concept of using optical EO data to derive crop LAI in combination with meteorological and soils data to model crop growth, condition, and yield became well established by the mid 1980s (Jackson, 1984; Bauer, 1985). However, widespread application at the field scale has, until very recently, been hampered by the lack of suitable EO observations, the main limitations being slow turnaround time, coarse spatial resolution and the lack of high frequency cloud-free observations (Moran et al., 1997). Subsequent studies have demonstrated how the latter could be resolved through the combined use of optical and radar observations (e.g. Clevers and van Leeuwen (1996); Dente et al. (2008)). However, it is only since the availability of near-real-time 500 m MODIS LAI data from the NASA Terra and Aqua satellites (every 8 days, starting from 2002), and since 2017, 10 m LAI from the Multispectral Instrument (MSI) on the Copernicus Sentinel-2 satellites (every 5 days) that satellite EO has become a viable contributor to farm scale crop growth monitoring (Ma et al., 2019). The MODIS LAI time series now stretches over 20 years and is routinely used for global to regional land surface model validation and calibration (Bloom et al., 2016; Demarty et al., 2007; Raoult et al., 2016). At the time of writing, there is currently no Sentinel-2 MSI LAI product available for download,

however open-source scripts for deriving LAI from atmospherically corrected MSI are available, enabling 10 m 5 day LAI time series production from 2017 onwards (e.g. Yin (2020); Yin et al. (2022)).

For both crop growth models and the application of EO data, a remaining challenge is the access to sufficient volumes of empirical measurements of actual crop yields against which to validate and calibrate models over larger spatial or temporal extents (Hunt et al., 2019). In recent years, there has been a widespread uptake of precision farming technologies, used by farmers to provide agronomic data to improve farming efficiency. One of these technologies is the introduction of high-accuracy GPS technology onto harvesting equipment, which is equipped with meters capable of obtaining instantaneous measurements of yield. By combining these yield measurements with the position of the harvester, it is possible to create maps of yield variation at the sub-field scale. Since many of these systems upload data directly to cloud storage, it is now possible to collect very large volumes of crop yield data for large numbers of fields.

In this paper, we describe the development of a new national modelling framework to provide improved field-scale predictions of crop yields across the UK, both within year and over longer timescales under future climate scenarios. The modelling system integrates the key data elements of climate science, crop growth modelling, EO and precision agricultural data. We discuss potential improvements of this framework, and its application to provide improved climate risk assessments for agricultural production. Section 2 describes the winter wheat yield model, the climate datasets, the satellite-derived leaf area index data used in the data assimilation, and the precision yield measurements used for validation of the model. Section 3 presents results of the model validation, the effect of data assimilation on the modelled yield, and forecasts of future wheat yields. In Section 4, we discuss the results and the limitations of the modelling approach. We also describe the extension of the modelling framework to other crops and the development of a crop yield demonstrator based on the modelling framework.

## 2. Methods and materials

Our integrative modelling framework predicts the impacts of climate change on crop yield using simple, functional crop growth models, augmented by ancillary data on soil water holding capacity and with the option to assimilate within year plant growth parameters derived from EO. We focus on winter wheat as an example, which is the largest crop grown in the UK by area (1.8 million hectares, 39.8 % of all arable crops (Defra, 2019)). Fig. 1 shows a schematic of the modelling system and the data flows.

We drive the crop yield model for winter wheat (Section 2.1) with the latest climate projections for the UK (Section 2.2) providing localised ‘potential’ future crop-yield predictions. We have the option to assimilate EO data on leaf area index (Section 2.3). We evaluate the accuracy of the modelled yields against actual yield measurements collected using precision yield monitoring sensors on combine harvesters operating in 891 fields across the UK (Section 2.4). This integrated modelling framework then provides localised estimates of ‘actual’ crop yields.

### 2.1. Crop yield model for winter wheat

For this application, we need crop yield models that are responsive and efficient to run, and could be readily adapted to assimilate EO data. While there are detailed process-based crop yield models available (e.g., the Sirius wheat model of Semenov et al. (2014)), the model are often site-based and also require detailed data on crop management and variety. For our UK-scale application, such data are not readily available as maps of current patterns, let alone future projections under climate change.

We develop a crop yield model for winter wheat based on approaches developed by Sylvester-Bradley and Kindred (2014) and Lynch et al. (2017). The model simulates ‘potential’ yield. The model accounts for climatic variables, soil effects on water availability, and day length, but does not represent the impact from non-climatic, agronomic parameters such as crop nutrition, pests or diseases. In this initial model development, we did not include these biotic or management processes as we

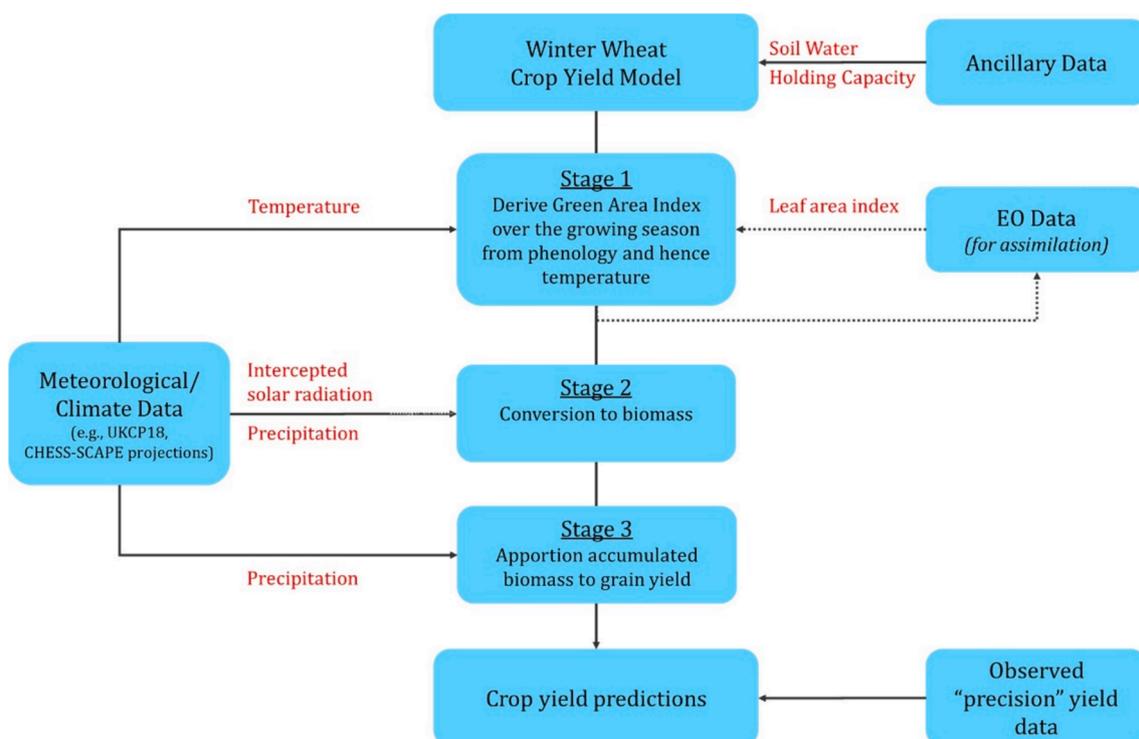


Fig. 1. Schematic of the crop yield modelling system for winter wheat and data flows.

would require information at the national scale.

The crop model for winter wheat was parametrised using benchmark values for UK wheat crops (Sylvester-Bradley et al., 2015). The key model parameters are listed in Table 1. Supporting Information, Appendix S3 of Redhead et al. (2020) provides some more details on the winter wheat model.

**Table 1**  
(a) Key parameters used in the winter wheat model.

Parameter	Units	Default
Daily Precipitation	mm	Per 1 km grid cell, values derived from CHES-met (historic) or CHES-SCAPE (future)
Maximum daily temperature	°C	Per 1 km grid cell, values derived from HadUK-Grid (historic) or CHES-SCAPE (future)
Minimum daily temperature	°C	Per 1 km grid cell, values derived from HadUK-Grid (historic) or CHES-SCAPE (future)
Mean daily temperature	°C	Per 1 km grid cell, values derived from CHES-met (historic) or CHES-SCAPE (future)
Net shortwave solar radiation	MJ	Per 1 km grid cell, values derived from downward shortwave radiation from CHES-met and HadUK-Grid (historic) or CHES-SCAPE (future) (See Section 2.2 and Appendix 1)
Required thermal time per growth stage	Growing degree days	See Table 1b
Green area index per growth stage	dimensionless	See Table 1b
Soil maximum available water content	mm	Per 1 km grid cell, values derived from Bell et al. (2018)
Atmospheric CO <sub>2</sub> concentration	ppm	Determined by year from RCP8.5
Day length threshold for start of construction phase	hours	14
Base temperature for accumulation of degree days	°C	0
Base temperature for frost damage	°C	-5
Sowing date	Julian day	1st October (274)
Harvest date	Julian day	31st August (243)
Radiation use efficiency	g MJ <sup>-1</sup>	3.1
Waterlogging penalty	kg ha <sup>-1</sup> yield loss per day of waterlogging	117.7
Base temperature for heat stress	°C	30
Heat loss penalty	% yield loss per degree day of heat stress during flowering (i.e. post anthesis)	1

(b) Growth periods used by the crop yield model for winter wheat to determine the rate of GAI development as a function of thermal time (growing degree days), based on benchmark values for UK wheat crops (Sylvester-Bradley et al., 2015).

Growth phase	GAI at start	GAI at end	Thermal time
(1) Foundation			
Sowing-GS30	0.0	1.6	1100
GS30-GS31	1.6	2.0	100
(2) Construction			
GS31-GS61	2.0	6.3	900
(3) Production			
GS61-GS69	6.3	6.3	50
GS69-GS87	6.3	1.3	750

The model has three main stages, which cover a number of growth stages.

#### Model Stage 1: Estimation of the green area index (GAI) over the growing season:

The model assumes a fixed sowing date of 1st of October, in line with the benchmark for winter wheat growth in the UK (Sylvester-Bradley et al., 2008). The same dates are used for the whole of Great Britain. Sowing dates for GB winter wheat crops do vary considerably, depending on a variety of factors, including ground conditions required for tillage and drilling, weed pressure and farmer preferences. It is therefore challenging to derive a consistent relationship with meteorological parameters. As sowing of winter wheat in GB usually takes place in a window from September (considered early sown) to November (considered late sown), we have run the model with sowing dates at these extremes to explore the sensitivity of the model to this parameter (Section 3.2 and Appendix 3).

Following sowing, the model accumulates GAI as a function of growing degree days above 0 °C through the three key phases of wheat growth, further subdivided by growth stages (GS): foundation (Sowing-GS30 and GS30-GS31), construction (GS31-GS61) and production (GS61-GS69, GS69-GS87, GS87-senescence) (Sylvester-Bradley et al., 2015).

During each growth phase, GAI is linked to growing degree days by multiplying the benchmark GAI value at the end of the growth stage by the proportional accumulation of thermal time from the end of the previous growth stage. Thus, a grid cell that has accumulated 50 % of the required thermal time to progress to the next growth stage will be assigned a GAI that is 50 % of the benchmark GAI for the end of the current growth stage.

Degree days accumulate as follows:

$$\text{If } T_{min} > 0 \text{ °C: degree days} = \frac{(T_{max} + T_{min})}{2}$$

$$\text{If } T_{min} < 0 \text{ °C: degree days} = \frac{(T_{max} - 0)}{2} - \frac{(0 - T_{min})}{4}$$

$$\text{If } T_{max} < 0 \text{ °C: degree days} = 0.$$

In order to account for vernalisation requirements, cumulative vernalisation days were also calculated as a function of mean daily temperature, following the equations in Spink et al. (2000) and Lynch et al. (2017). The crop accumulates one vernalisation day when the daily mean temperature is within the range 3–10 °C, while a proportionally lower vernalisation day is acquired if the daily mean temperature is within either of the ranges -4 to 3 °C or 10–17 °C. A crop requires 50 vernalisation days before progressing beyond GS31. When this does not occur, the estimated GAI remains constant until the crop is adequately vernalised and degree days accumulated over this period do not contribute to the following growth phases. We account for the effect of day length by limiting progression beyond GS31 until the day length has progressed beyond 14 h. We calculate the day length from the latitude and date using the *geosphere* package (Hijmans et al., 2017). Where the accumulated degree days are sufficient to progress beyond GS31 before day length > 14 h, GAI remains constant until this threshold is met.

The speed of progression through the growth stages will be quicker at higher temperatures, resulting in a reduced grain-filling period and thus a reduced amount of accumulated biomass available for yield. We also account for the effects of heat stress on the wheat plant during flowering (i.e. post anthesis); the yield is penalised by 1 % for every degree day over 30 °C, following Liu et al. (2016).

GAI development is stalled if the temperature falls below -5°C once the crop has passed GS31 (Spink et al., 2000). We assume that harvesting takes place on the 31st of August. Where the GS87-senescence phase is not complete before this date, we assume that this phase progressed more rapidly, with GAI of zero by harvest and a subsequent reduction in yield due to the decreased time to accumulate yield-contributing biomass.

We thus produce a time series of GAI for each year and each location, whether these 'locations' are the individual 1 km x 1 km grid cells for the national-scale modelling (Section 3.1) or the fields where there are

precision yield measurements, for the LAI data assimilation (Section 3.3). An example GAI time series is shown later in Fig. 6.

#### Model Stage 2: Estimation of the intercepted solar radiation and conversion to biomass:

Following the construction of the GAI time series from temperature and day length data for the grid cell, we calculate the intercepted radiation and convert this to biomass using an assumed radiation use efficiency (RUE) (Shearman et al., 2005). We assume the RUE to be  $3.1 \text{ g biomass m}^{-2} (\text{MJ intercepted light})^{-1}$ , as estimated to be maximal for UK wheat crops (Lynch et al., 2017).

We sum the rainfall from sowing to GS30 to estimate the available water before the onset of rapid growth, with the plant available water content (PAWC) at this point being the lower of the cumulative rainfall to date and the estimated soil available water content (AWC) (Bell et al., 2018), as most soils are saturated prior to this date. We define the available water content as the difference between the maximum and minimum soil moisture content (from Bell et al. (2018)). It is therefore a representation of the maximum water available to the crop when the soil is saturated. We accept that our usage may differ from other applications.

After this date, we calculate the daily PAWC ( $\text{PAWC}_d$ ) of the soil following (Lynch et al., 2017) as:

$$\text{PAWC}_d = 0.65(\text{AWC}_{d-1} + \text{Rainfall}_{d-1} - 0.2\text{Biomass}_{d-1})$$

where AWC is the available water content of the soil. When PAWC drops below the amount required for that day's biomass accumulation, no further biomass is accumulated until PAWC has increased from subsequent rainfall. This accounts for any effects of water limitation and drought.

#### Model Stage 3: Apportionment of the accumulated biomass to grain yield:

Following Lynch et al. (2017), we assume that all the biomass accumulated after GS69 (end of flowering) is transferred to the growing seeds, rather than to the vegetative parts of the plant, and thus contributes to harvestable yield. We calculate the potential yield as the sum of biomass accumulated between GS31 and GS61 multiplied by 0.3 plus the sum of biomass accumulated after GS69. The former (i.e. the accumulated sum of biomass between GS31 and GS61) accounts for the redistribution of water-soluble carbohydrates from the stem after flowering (Lynch et al., 2017; Sylvester-Bradley et al., 2008). The yield is penalised if waterlogging occurs after the onset of flowering (GS61). Waterlogging occurs when the daily rainfall exceeds the residual PAWC. If this occurs for more than 5 days, a yield penalty of  $11.77 \text{ g}^{-1} \text{ m}^{-2} \text{ d}^{-1}$  is also applied (Olgun et al., 2008), until such time as  $\text{PAWC}_d$  falls below the maximum.

We include an option to account for the fertilisation effect of higher atmospheric concentrations of carbon dioxide ( $\text{CO}_2$ ). We follow the approach used by Senapati et al. (2019) in the Sirius wheat model, where the radiation use efficiency is assumed to be proportional to the atmospheric concentration of  $\text{CO}_2$  ( $[\text{CO}_2]$ ) and increases by 30 % for a doubling of the atmospheric concentration of  $\text{CO}_2$  from its baseline concentration of 338 ppm, with RUE reaching its maximum at 750 ppm. As discussed by Senapati et al. (2019), this increase in RUE is consistent with the meta-analysis of different field-scale experiments on the effects of  $[\text{CO}_2]$  on crops (Vanuytrecht et al., 2012) and is consistent with the responses used in other wheat simulation models.

## 2.2. Climate/Meteorological data

We use daily-averaged values of the following meteorological parameters in our winter wheat crop growth and yield model: surface air temperature, net shortwave radiation and precipitation (i.e., rainfall), see Table 1a. We thereby account for impacts on crop growth resulting from waterlogging (in winter) and drought (in summer), during key periods of crop growth. We can take or derive the required

meteorological variables from (a) historical and/or current-day observed or modelled datasets, or (b) modelled future projections of climate.

### (1) Historic and Current Meteorological Datasets

We use two observationally based datasets: the Climate Hydrology and Ecology research Support System meteorology dataset for Great Britain (CHESS-met) and HadUK-Grid, as neither dataset fully provides the meteorological parameters required for the modelling (Table 1a). CHESS-met and HadUK-Grid can be used together as they are both derived from the same station observations, interpolated to the 1 km grid, adjusting for local topography. As they provide different variables, we select the most appropriate variables from each dataset.

(a) CHESS-met is an observation-based, daily meteorological dataset for Great Britain at 1 km x 1 km spatial resolution, covering the period 1961–2019 (Robinson et al., 2020; Robinson et al., 2023a). We use this dataset to provide the mean daily surface air temperature, incoming shortwave solar radiation and precipitation.

(b) the HadUK-Grid gridded and regional average climate observations for the UK (Hollis et al., 2018). HadUK-Grid is a meteorological station observation-based dataset at 1 km x 1 km spatial resolution for the UK, providing, amongst other variables, the daily minimum and maximum temperatures from 1960 and monthly sunshine hours from 1929 (Hollis et al., 2018).

### (2) Future projections of climate

The CHESS-SCAPE future projections of meteorological variables (Robinson et al., 2022; Robinson et al., 2023b) are derived from the UKCP18 climate projections for the UK (Lowe et al., 2018). The downscaling of the three meteorological variables of interest (surface air temperature, incoming short wave solar radiation and precipitation) is based on the approach used by Robinson et al. (2017), but has been adapted for the UKCP18 climate model output (Robinson et al., 2023b). The downscaled meteorological fields are then bias corrected or adjusted for each 1 km grid cell and for each season. We use all 4 members of the CHESS-SCAPE 1 km ensemble for the RCP8.5 scenario (Robinson et al., 2023b).

The winter wheat model uses net shortwave radiation as input to calculate the photosynthetically active radiation. The CHESS-met and CHESS-SCAPE datasets provide incoming or downward shortwave radiation, from which we derive the net shortwave radiation required for the modelling using surface albedo and cloud cover fraction. Appendix 1 contains more information about the meteorological datasets, the downscaling and bias correction approaches used.

We use the spread from the 4 member CHESS-SCAPE RCP8.5 ensemble as a measure of uncertainty. For the potential yields (see Section 3.1), we take the minimum and maximum of the ensemble. For the data assimilation, we use the standard deviation from the uncorrected 12-member UKCP18 RCP8.5 ensemble as a measure of the error in the modelled GAI. These are conservative estimates of the uncertainty, as the uncertainty spread of a bias corrected ensemble is expected to be smaller than an uncorrected ensemble, because by reducing biases between ensemble members and the historical data, one also reduces biases between ensemble members, and therefore reduces the overall spread. However, even in a bias-corrected ensemble there would still be uncertainty due to the different climate sensitivities of the ensemble members. However, the fact that the bias correction is applied seasonally in fact increases the amplitude of the seasonal cycle in air temperature and precipitation in this ensemble member, so increases the seasonal variability.

## 2.3. Earth observation data assimilation

The crop growth and yield models give water-limited potential yields, as the models do not account for many non-meteorological/hydrological effects on crop growth, such as local management history, fertiliser input, pests and diseases. We obtain yields closer to the actual yields through the assimilation of Earth observation data into the

crop growth model. This assimilation calibrates the growth model to local environmental conditions and management practices, without the need for more complex growth models requiring detailed local observations of both.

We assimilate EO data on leaf area index for a given growing season to adjust the modelled GAI time series for that season. Although GAI includes all green plant parts and LAI just includes leaves, the difference between the indices is very small in the case of crops. We use a Four-Dimensional Variational (4D-Var) data assimilation method (Tala-grand and Courtier, 1987), based on that used in the EO-Land Data Assimilation System (Lewis et al., 2012). The assimilation method aims to minimize a cost function (see Appendix 2 for further information on the data assimilation approach).

For the assimilation, we use LAI derived from atmospherically corrected Sentinel-2 MSI data, following approaches described in Gómez-Dans et al. (2016) and Yin et al. (2022). Sentinel-2 MSI has a revisit frequency of 5 days and spatial resolution of 10 m. The Sentinel-2 MSI data is available for download from the Sentinel hub. The Sentinel-2 LAI product has been evaluated against surface LAI measurements of winter wheat in Austria (Novelli et al., 2019), Bulgaria (Kamenova and Dimitrov, 2021), China (Pan et al., 2019) and France (Mercier et al., 2020). Novelli et al. (2019) find good agreement and low error. Pan et al. (2019) report that the retrieved LAI has a high accuracy, with a coefficient of determination of 0.892, an MRE of 10 % and an RMSE of 0.745.

We run the wheat crop model with assimilations of the 10 m MSI LAI data for those precision yield fields which have at least 7 cloud-free MSI observations between May and August inclusive. The LAI data capture criterion reduces the number of precision yield field locations from 891 to 719 (Section 2.4). For the MSI LAI, we use the average of all the cloud-free pixels within each field's boundaries. The averaging of the MSI LAI data over each field used in the validation ensures that sampling errors arising from using individual pixels within each field are removed. All MSI observations are used that have at least 70 % of the field marked as cloud-free. The number of LAI observations used in the data assimilation varied across the precision yield field sites and across years, from a minimum of 10 to a maximum of 52 observations (Table 2). The meteorological driving data is taken from the encompassing 1 km grid box. We then calculate an estimate of the yield from the optimised GAI time series.

As the Sentinel-2 data are only available from 2017, we also use MODIS LAI data, specifically the MCD15A2H product, which has a temporal resolution of 8 days and a lower spatial resolution of 500 m, downloaded from <https://modis.ornl.gov>. The MODIS LAI data are only used for the purposes of enabling projection of predicted yields into the future in section 3.3. For a given field, we take the average LAI across the encompassing MODIS pixel and use this for the assimilation.

#### 2.4. Precision yield data

Over the last 10 years, many farms across the UK have adopted precision yield monitoring and mapping services available on most modern combine harvesters. Similar capabilities to measure cut grass herbage yields are increasingly available on self-propelled forage harvesters. These systems work by linking physical sensors in the harvester that constantly measure crop weight and moisture content with accurate GPS location systems. This results in a high density of point data that can

**Table 2**  
Number of LAI observations used in the data assimilation, across the precision yield field sites and years from 2017 to 2019.

Year	Minimum Number	Mean	Maximum Number
2017	10	14.1	17
2018	16	35.9	52
2019	21	31.3	45
All years	10	27.1	52

be used to map fine-scale (<10 m) spatial variation in arable and grass crop yields within fields (Pywell et al., 2015). This has many applications for precision agriculture, including variable rate fertiliser application and pest management (Shannon et al., 2018). Over time, this approach can also be used to build up a time series of crop yields for given fields that enable analysis of temporal trends in patterns of yield. Recently, combine and forage harvesters have begun to automatically upload real-time yield data onto cloud storage platforms developed by the machinery manufacturers. These data are thus a valuable resource for the calibration and validation of crop modelling and EO monitoring (Hunt et al., 2019).

We have access to precision yield data from 891 fields, collected as part of the Achieving Sustainable Agricultural Systems (ASSIST) project, each representing the average yield for that field in a given year between 2016 and 2019. Yield data were supplied by volunteer farmers through manual data exports from farm management software or downloads from the CLAAS Telematics cloud platform ([www.claas-telematics.com](http://www.claas-telematics.com)). In addition to recording crop yields, the yield monitoring systems also provide high accuracy RTK GPS positions of the combine harvester, grain moisture content, a timestamp and machine operating metrics (e.g. speed) (Fincham et al., 2023). The fields are located across the agricultural areas of lowland England (i.e., the dominant wheat-growing areas), as shown in Fig. 2.

The calibration of the yield sensors is a known issue, but with sufficiently large sample sizes, we expect this to add 'noise' to the data (i.e., there is no systemic bias). Many other factors affect the accuracy of precision yield data, and we used a standardised data cleaning process to account for these (e.g., speed and turn angle of the combine harvester, incomplete swath width, grain moisture content), as described in Fincham et al. (2023). Each field contained several hundred individual data points. We take the mean yield of the cleaned points in each field to represent the average yield for that field. The average field size used in the validation of the winter wheat model is 13.9 ha, with a range (one standard deviation) from 2.4 to 25.4 ha.



**Fig. 2.** Map of England showing locations of the winter wheat precision yield measurements.

### 3. Results

#### 3.1. Winter wheat water-limited potential yields

We undertake runs with the winter wheat yield model using meteorological parameters from the CHES-met and HadUK-Grid datasets for the historic period (pre-2020) and from the four members of the bias-corrected CHES-SCAPE ensemble for RCP8.5 (Section 2.2) to derive modelled water-limited potential yields between 1980 and 2080 for

each 1 km x 1 km Ordnance Survey grid cell covering Great Britain. These runs include the CO<sub>2</sub> fertilisation effect (Section 2.1). We use a time series of global mean annual CO<sub>2</sub> concentrations developed for the RCP8.5 scenario by Meinshausen et al. (2011) (and available from <http://www.pik-potsdam.de/~mmalte/rcps/>).

Applying masks of the current wheat growing areas in Great Britain (i.e. all 1 km cells with greater than 0 % cover of 2015 of arable land, from the UK Land Cover Map 2015; Rowland et al. (2017)), we derive the mean annual predicted water-limited potential yields for Great

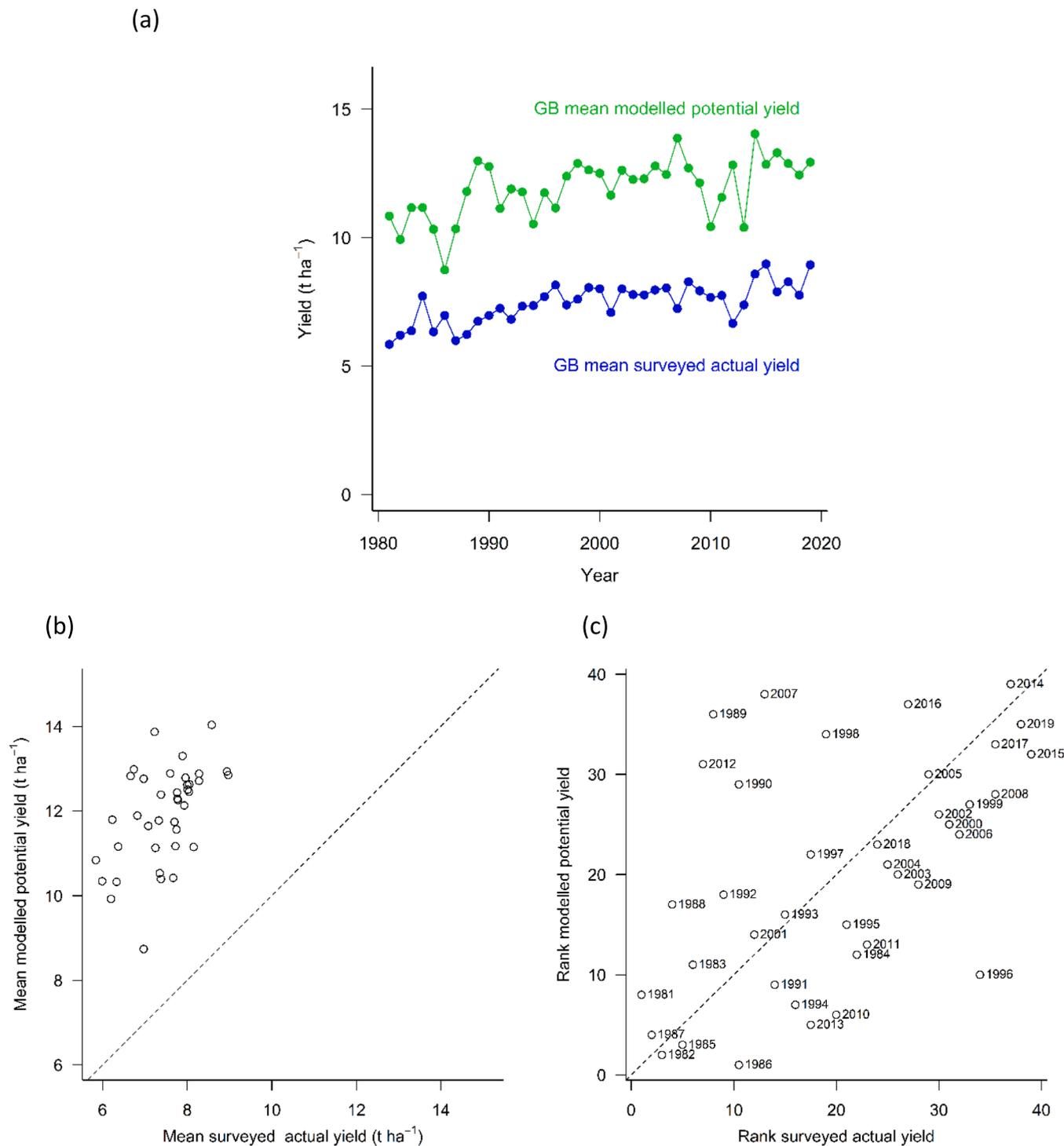


Fig. 3. Time series (panel a) and scatter plots of annual mean modelled potential winter wheat yield for Great Britain against the actual national annual yields for winter wheat, from the annual Defra Farming Statistics, as (panel b) and (panel c) ranked yields, yields in t/ha. For panels (b) and (c), the dotted line indicates a hypothetical 1:1 relationship.

Britain and compare these to the actual national yields obtained from the annual Defra Farming Statistics for the years 1990–2020 (Defra, 2020b), as a time series in Fig. 3a. The Defra Crop Survey provides nationally averaged yields, based on annual surveys of a stratified random sample of ~2000 farms each year. We use this period because prior to 1990, wheat yields showed an increasing trend due to factors other than climate (i.e. changes to farming practices and agricultural intensification). The results demonstrate that the modelled national potential yields are able to capture the overall increase in yield and the year-to-year variation in yield to some extent, with a positive relationship between the modelled potential and the observed actual yields ( $R = 0.428, 0.461$  and  $p = 0.016, 0.001$  for the raw and ranked yield data, respectively). Modelled potential yields are always higher than actual yields (as expected), but the ability of the model to use observation-based meteorological data to distinguish years of high relative yield from those with low relative yield is clear, especially when ranked yield values are compared (Fig. 3c). This gives us confidence that the model is able to capture the trends and the year-to-year variability in yields.

Fig. 4 maps the modelled water-limited potential yield averaged from 2010 to 2020 and the change in the average potential yield for 2030–2040, 2050–2060 and 2070–2080, relative to the average yield in 2010–2020. The model results show a cumulative decline in crop yields in the south eastern part of Great Britain and an increase in yields in the cooler northern parts of Great Britain, with south western areas remaining comparatively stable. We have masked out entirely non-agricultural areas using the criteria outlined above, as these are currently unsuitable for crop production and likely to remain so in the future due to constraints of topography and land use conversion.

In Fig. 5, we present three exemplar grid cells: (a) Scotland; (b) Eastern England and (c) South West England (see Fig. 4 for locations) to illustrate contrasting behaviour in the water-limited potential yields between 1980 and 2080 based on the projected climate from CHES-SCAPE (mean of the 4-member ensemble). For each of these locations, we compare the modelled annual potential yield for the grid square between 1980 and 2080, with the mean modelled yield from the current wheat growing areas in Great Britain (“national”) for the same periods. We also include a time series of the mean annual surface temperature and total precipitation for the location and for Great Britain for the same period (1980–2080; Fig. 5). The shading gives the uncertainty spread as the range (min–max) derived from the four member CHES-SCAPE ensemble.

Although we take account of the effect of heat stress on yield caused by elevated temperature during flowering, we find changes in temperature (mostly) and precipitation generally lead to increased wheat growth and yields up to around 2040, but the trends for the different locations vary for the latter part of the time series. For location (a) in Scotland, the location shows a smaller increase in temperature to that forecast nationally and precipitation is above the national mean. This led to a significant increase in wheat yield from a 2–3 t ha<sup>-1</sup> to values (14–15 t ha<sup>-1</sup>) comparable to the national mean. The large number of zero yields indicates yields in this area of Scotland remain low for the first half of the time series, with at least some ensemble members predicting yields of zero up to around 2025 (i.e. crop failure due to excessive cold or failure to reach a harvestable growth stage). By the end of the time series around 2080, the curve reaches the national average, suggesting that climate change has dramatically increased the suitability of this area for wheat. For location (b) in East Anglia, yields remain stable until around 2040, but then show a decline as decreased precipitation offsets the temperature increase. As indicated earlier, we accounted for the impact of waterlogging (in winter) and drought (in summer) during key periods of the crop growth, through the use of a

daily time step. Location (c) in South West England showed a steady increase in yield, similar to the national average, despite the similar change in temperature to location B. This is probably due to precipitation remaining slightly higher than the national average and the soils in this region being less prone to drought.

### 3.2. Sensitivity to sowing date and heat stress

The sensitivity of the model to the sowing date was investigated by running the model with three sowing dates: Early = 1st September, Benchmark = 1st October, Late = 1st November, following Sylvester-Bradley et al. (2015). Sowing date clearly affected yields, with early or late sowings yielding lower than benchmark sowings in the historic time period. For the future time period, varying sowing dates resulted in changes in projected potential yields, but this effect was inconsistent year on year, with no one sowing date consistently favouring a higher yield, and no indication that selecting the most beneficial sowing date for a given year would be sufficient to offset the effects of climate change (Appendix 3, Figure A3.1). There was also no indication of sowing date changing the spatial patterns of yield under future climate scenarios (Appendix 3, Figure A3.2).

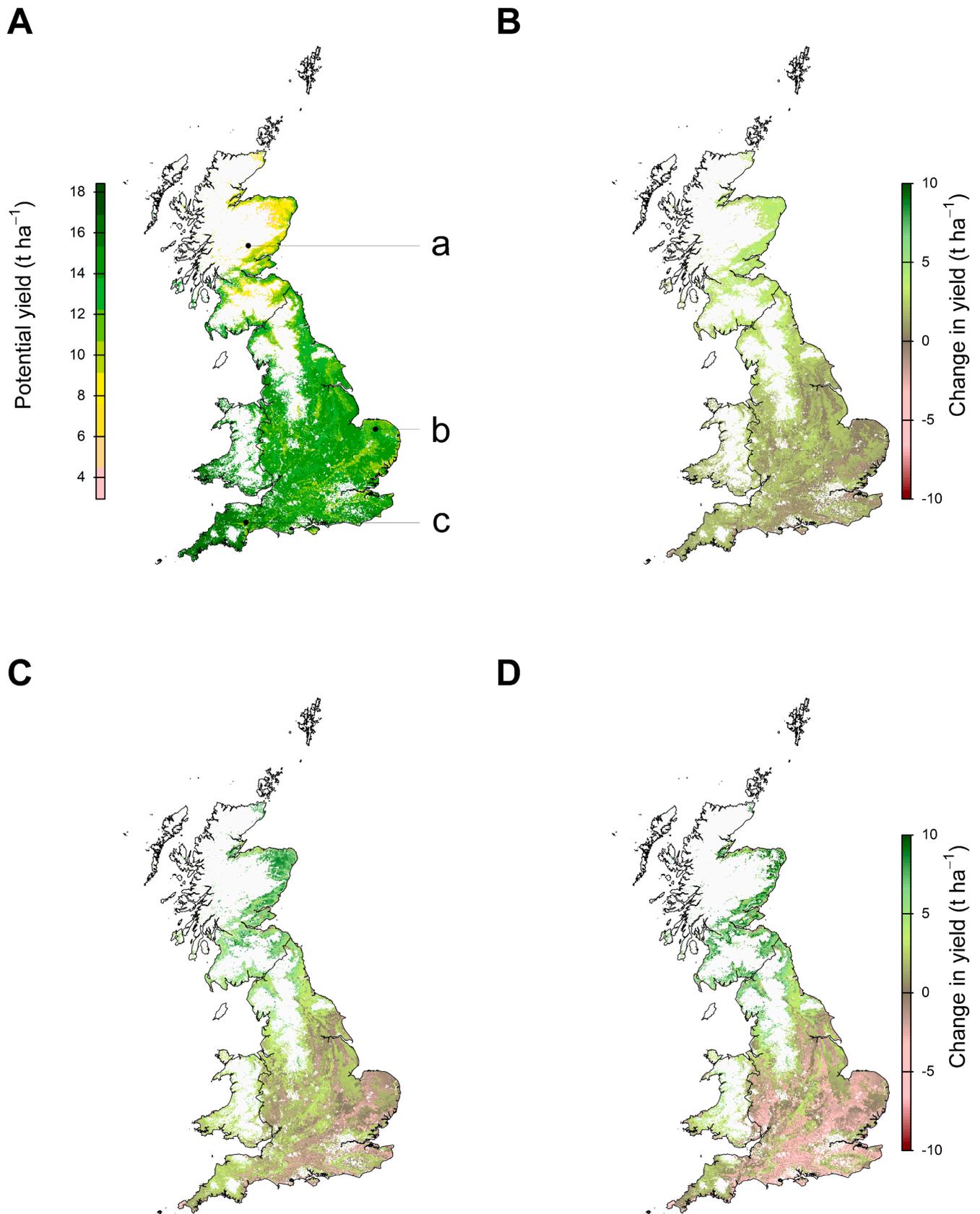
Heat stress was modelled by applying a penalty to the final yield based on the number of accumulated heat stress days over 30 °C. Application of the heat stress penalty reduced the national average yields by up to 2 % in the historic time period and 16 % in the future time period (Appendix 3, Figure A3.3). Effects on individual cells were considerably higher, with some cells predicted to experience reductions in yield of up to 60 % due to heat stress by 2080.

### 3.3. Impact of EO data assimilation on predicted yields of winter wheat

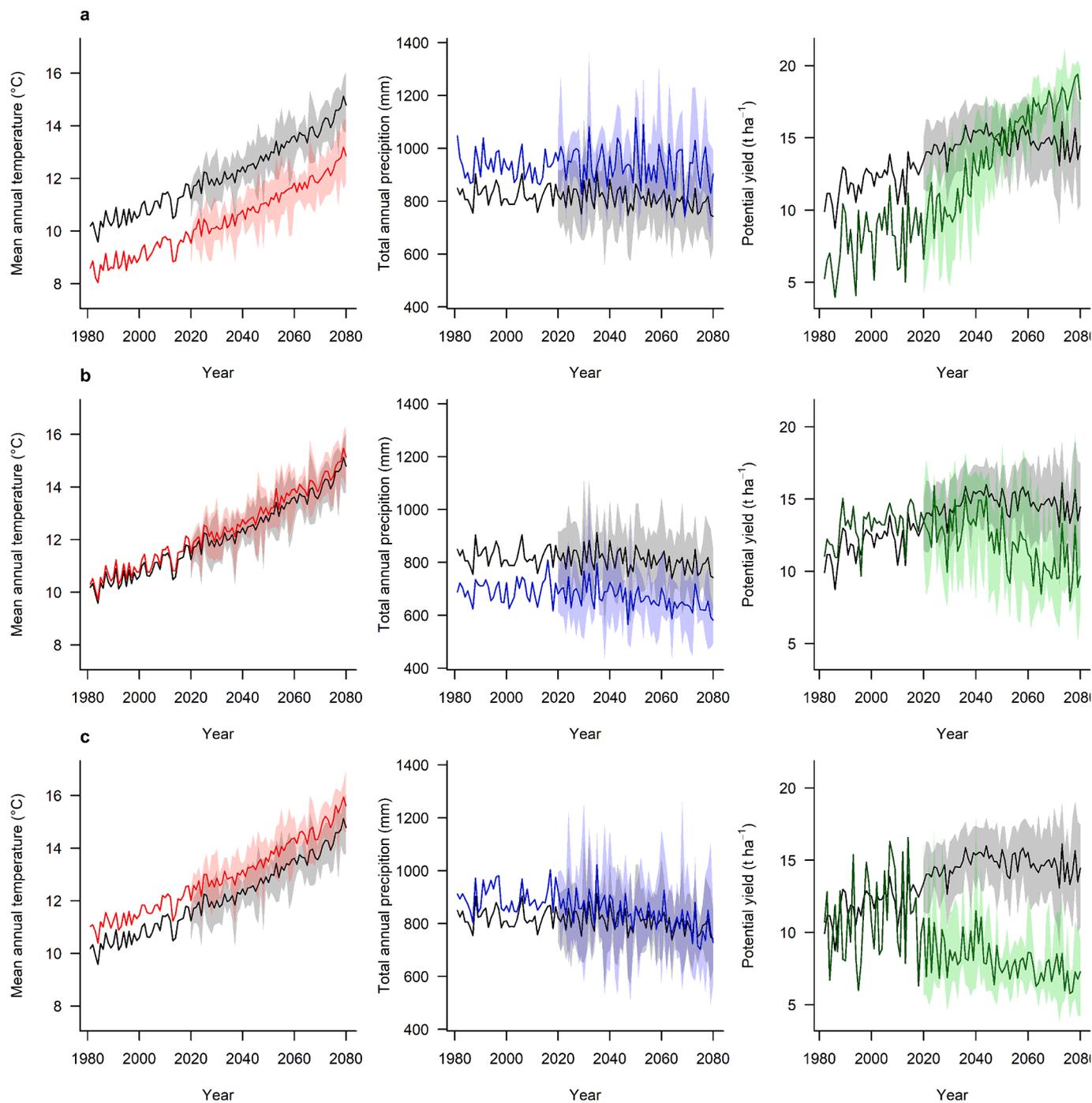
We run the wheat crop model with assimilations of the 10 m MSI LAI data for the 791 precision yield fields, where there are precision yield data and at least 7 cloud-free MSI observations between May and August inclusive (Section 2.3). The assimilation assumes a fixed sowing date of 1st October, although the sowing date could be adjusted as a possible future development, as discussed in Section 4.2 An example of the assimilation of the LAI data is shown for an illustrative field in Fig. 6. We take meteorological parameters from the CHES-met and HadUK-Grid datasets for the encompassing 1 km grid box.

The effect of assimilating the MSI LAI data into the crop model is to reduce the mean water-limited potential yield from all precision yield field sites from 13.26 to a yield of 8.99 tonnes per hectare. We show the distribution of the ensemble-mean yields with and without assimilation in Fig. 7, together with the distribution of the measured yields. The prior model before DA in Fig. 7 displays a much narrower distribution of yield values than the measured precision yields. This is because many of the precision yield sites are located close together and often within the same 1 km<sup>2</sup> grid cell. Hence, the sites share common driving meteorology. In the first instance, the model predicts the same green area index and yield for any points using the same input meteorological data. Once we update the green area index with the high-resolution Sentinel-2 observations in the DA, the model predicted yields are also updated, resulting in a much better distribution of yield values despite common meteorological forcing. In Fig. 7 (right), we show the distribution of the ensemble-mean yields with and without assimilation, as averages over the 1 km<sup>2</sup> grid cell.

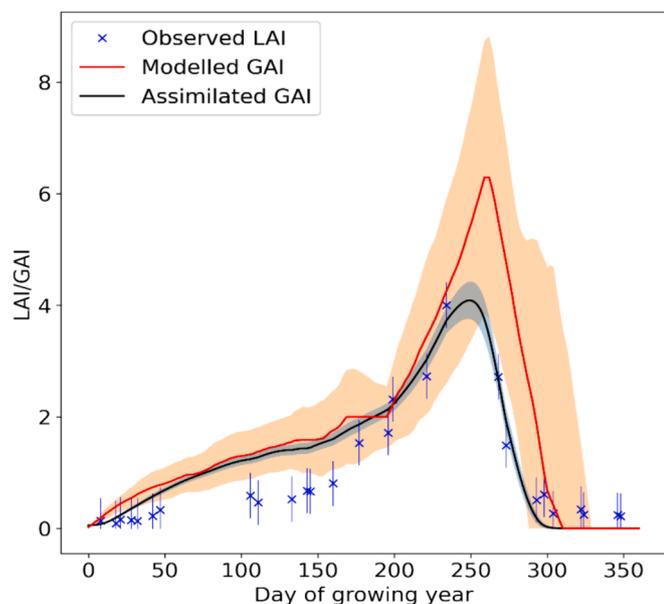
The CropNET model was developed to assess the impact of longer-term climate change on crop yields. As shown earlier in Fig. 4, the base model captures the large-scale variability from North to South and from East to West. It is less able to capture the field-scale spatial



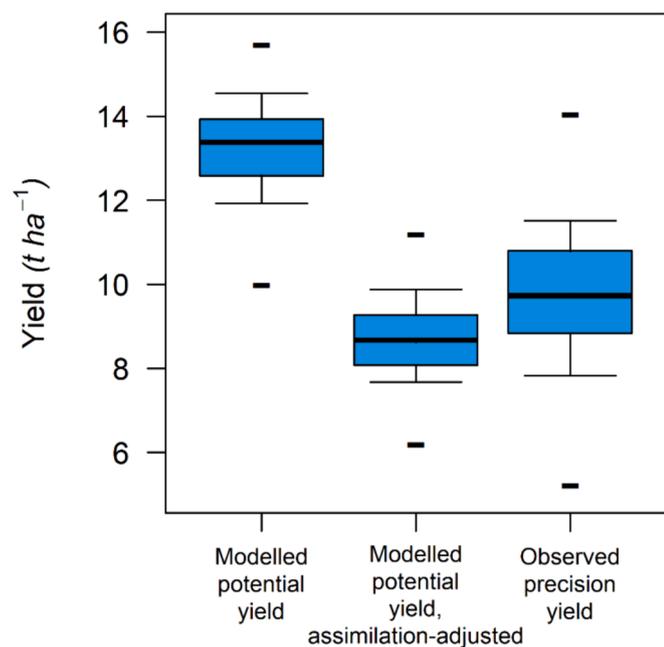
**Fig. 4.** Maps of water-limited potential yields of winter wheat modelled for Great Britain for the past decade (2010–2020) in t/ha (Map A) and of change in potential yield between the past decade and d 2030–2040 (Map B), 2050–2060 (Map C) and 2070–2080 (Map D). Climatic inputs to the model were derived from CHESM-met and HadUK-Grid for the historic period (pre-2020) and from the mean of the four bias-corrected ensemble members of the CHESM-SCAPE RCP8.5 dataset for the future time period. Points a, b and c are exemplar locations in England and Scotland for discussion.



**Fig. 5.** Time series of the mean annual air temperature at 1.5 m (in °C) (left-hand set of panels), the total annual precipitation (in mm) (middle set of panels) and the modelled annual water-limited potential yield for winter wheat (in t/ha) between 1980 and 2080 for locations in (a) Scotland; (b) Eastern England and (c) South West England (right-hand set of panels). The solid line represents the CHES-met and HadUK-Grid for the historic period (pre-2020) and the single bias-corrected CHES-SCAPE ensemble member for the future (2020–2080). The shading gives the range (min–max) from the four member CHES-SCAPE ensemble. The red, blue and green lines are for the grid cell and the black line is the ‘national’ GB average value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 6.** Example of the data assimilation using Sentinel-2 MSI LAI data. The top, red line is the modelled GAI, the lower, black line is the GAI after assimilation and the blue crosses are the Sentinel-2 MSI LAI observations. The red shaded area is the standard deviation assumed for the modelled GAI, the black shaded area is the standard deviation of the assimilated GAI, and the blue vertical lines show the standard deviation assumed for the MSI LAI observations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 7.** Box plot showing the distribution of the yield from all the precision yield fields and years run, with (middle) and without (left) assimilation. The right box plot shows the observed yields. The black horizontal line within each box plot shows the mean, and the boxes extend out to the 25th and 75th percentiles. The whiskers of the box plots extend out to the 10th and 90th percentiles, with the points marked by horizontal lines above and below these the maximum and minimum values respectively, plotted where within the axis range.

variability provided by the precision yield data. Most of the precision yield sites are in the South and East of England (see Fig. 2). That said, we show scatter plots of the actual yield against the modelled yield in Fig. 8, for the model with (blue) and without (red) assimilation for 2018 (panel a) and 2019 (panel b). We have averaged precision yield field sites in the same CHES 1 km grid.

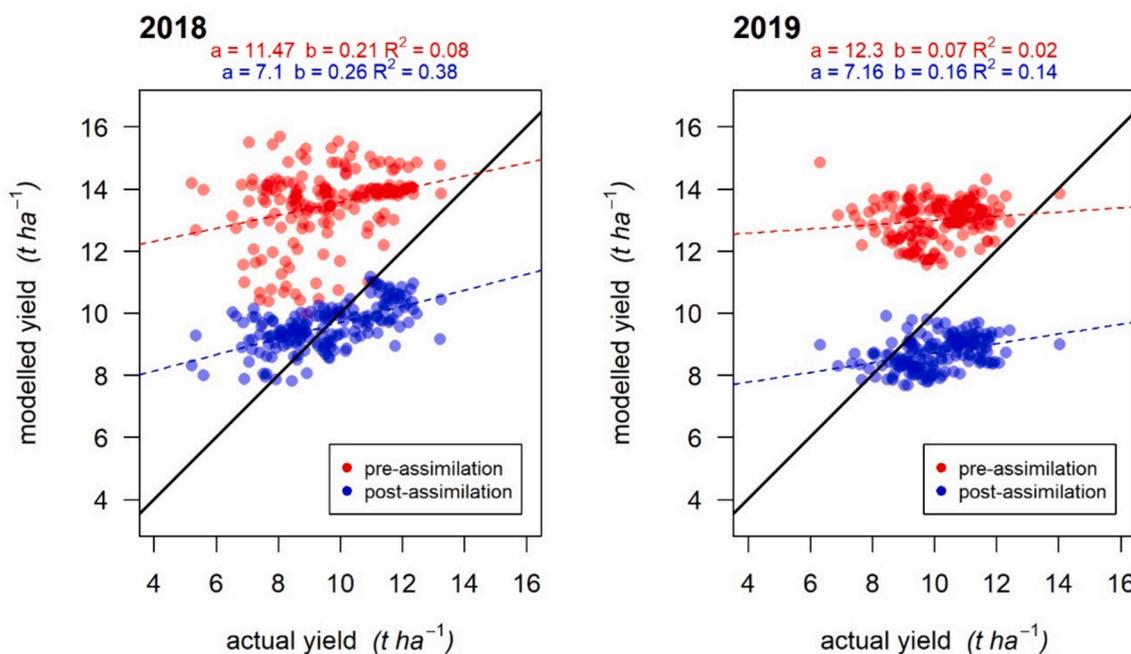
The root mean square error, unbiased root mean square error (Entekhabi et al., 2010) and correlation all show improvements (Table 3). Fig. 8 and Table 3 show an increase in the coefficient of determination ( $R^2$ ) between the observed and modelled yields after assimilation and the slope of the regression line improves slightly, indicating that the assimilation captures aspects of the crop growth that are not modelled. This, in addition to the improvement in the unbiased RMSE, shows that after assimilation the model reduces the scatter in the modelled yields, and therefore improves the ability of the model to distinguish years when there is a drought (such as 2020). We note that the model yield performance for 2018, both before and after data assimilation, is better than is the case for 2019. The model captures more of the spatial variability in the 2018 yields. 2019 was a uniformly good year in terms of wheat growing weather across much of England, with resultant high yields (see the ranked yields in Fig. 3b). To some degree, this masked much of the usual spatial variation. The model does capture this to some extent (points clustering tighter on both the X and the Y axes, with fewer low-yielding outliers).

The precision yield data against which we evaluate against our model outputs is an accurate and spatially precise measurement of actual yield harvested. There are a wide range of factors that affect the relationship between this value and the estimated yield of the wheat plant, even once data assimilation has adjusted the prediction to account for local factors affecting the green area index. These include farmer actions to help mitigate against poor harvests (e.g., adjusting timing of harvest, application of agrochemicals) and factors that impact on yield without affecting the general shape of the GAI curve (e.g., competition with other wheat plants, lodging, delayed harvest, pest damage post senescence). These can all act to suppress or mask the relationship between potential yields as modelled from climatic factors and reduce the extent to which assimilation of EO data can correct the models, thus reducing the proportion of variance explicable by the model (i.e.,  $R^2$ ). However, the modelled potential yield still provides the 'ceiling' for the yield achievable under ideal conditions. Thus, changes in potential yield over time and space are important to consider in terms of indicating the likely degree of adaptation required by farming systems to maintain current levels of productivity.

Overall, the data assimilation reduces the modelled potential yield values (based on idealised benchmark values, Section 2.1) and brings them closer to the measured actual yields, although the spatial variability is reduced. This gives us some confidence in our data assimilation approach and highlights the value of the Sentinel-2 LAI product.

#### 3.4. Winter wheat trends at a precision-yield site

The EO data assimilation approach currently considers the entire growing season within a given year. We use a simple regression approach to demonstrate how the forecast potential wheat yield could be converted into more realistic predicted yields. We take the modelled yields with and without data assimilation of MODIS LAI data for the years 2002 to 2019 (as Sentinel-2 data is only available since 2017) and performed a linear regression of the yield with data assimilation against the yield without data assimilation, independently for each location considered. We apply this regression relationship to convert the time series of water-limited potential yields into more realistic predicted yields. In Fig. 9, we show the regression fit (Fig. 9a), and the modelled



**Fig. 8.** Scatter plots of the actual yield against the modelled yield for the model with (blue) and without (red) assimilation for (a) 2018 and (b) 2019. The actual yields have been aggregated to CHES 1 km grid by averaging yields from fields in the same grid cell (which have the same climate data). The 1:1 lines are shown in black. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 3**

Statistics of the modelled yield compared against observed yield before and after assimilation of Sentinel-2 LAI.

Year	Before/After Data Assimilation	RMSE	Unbiased RMSE	Coefficient of Determination ( $R^2$ )
2018	Before	4.31	1.77	0.08
	After	1.37	1.37	0.38
2019	Before	3.21	1.31	0.02
	After	1.75	1.14	0.14
Both	Before	3.82	1.64	0.04
	After	1.57	1.44	0.11

potential yield and the yield after data assimilation for an example location in Northamptonshire, where we have precision yield measurements for the years 2016 to 2019 (Fig. 9b). We find other locations show a very similar behaviour, which is not surprising given the locations of the precision yield measurements. From Fig. 9 and for the period of the time series when data assimilation is used (2002–2019), we see that the model predicted significantly lower yields in 2004, 2012 and 2014. The yields, especially in 2012, were affected by drought. We note that the yields predicted after data assimilation shows these years as having lower yields. Ideally, we would have confirmed this by comparison against precision yields measurements. We would like to extend the analysis to include more recent measurements from the winter wheat precision yield field sites (after 2019) to provide further data-points for the evaluation of the approach. However, the present CHES-met datasets are only available to the end of 2019. The present analysis provides some reassurance regarding the performance of the crop yield model and our data assimilation method.

## 4. Discussion

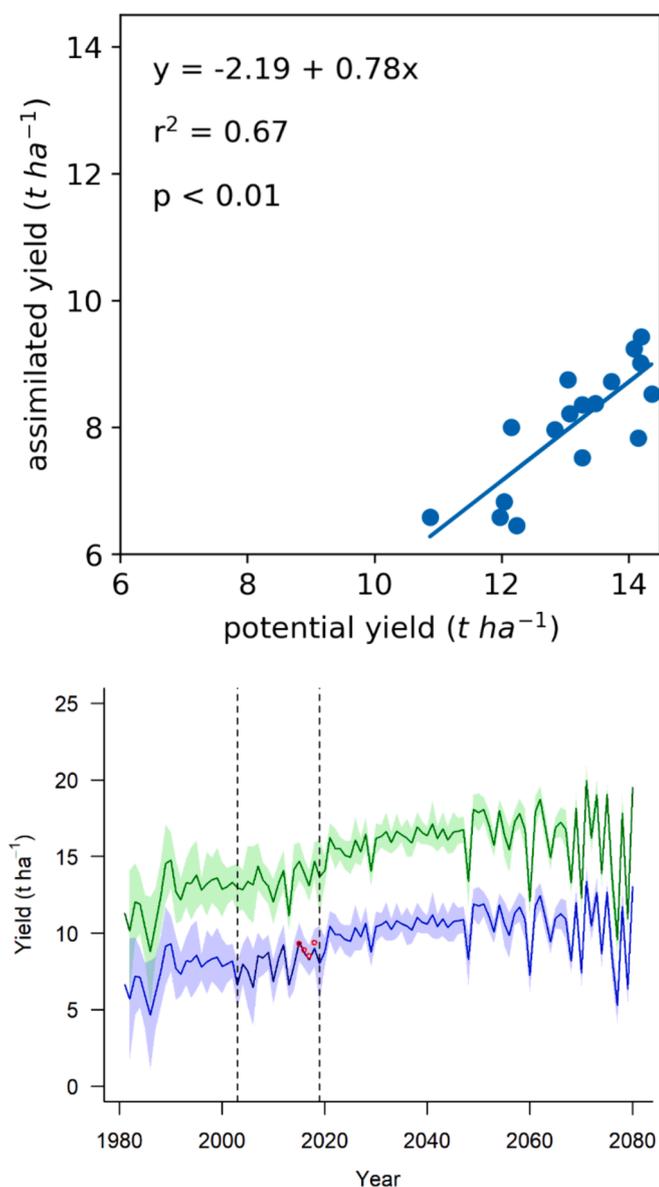
### 4.1. Winter wheat

We find that increasing temperature generally leads to increased wheat growth and yields up to around 2040, but the trends vary for the latter part of the time series at different locations. Our model suggests

that many of the key wheat growing areas of the UK will become progressively less suitable for growing this crop after 2040 due to reductions in precipitation. In terms of climate risk assessment, our results suggest that farmers and agri-businesses in these areas will need to adapt to the predicted changing climate.

CO<sub>2</sub> fertilisation acts in the model by modifying radiation use efficiency. Therefore, its impacts are greatest when the yield is not strongly limited by temperature and available water, and thus the effects of CO<sub>2</sub> fertilisation on yield vary both spatially (depending on which factors have the greatest limiting effect on yield) and temporally (as CO<sub>2</sub> rises to and beyond the point where its impact on radiation use efficiency is maximal). However, by comparing national average values, CO<sub>2</sub> fertilisation can result in an additional yield increase of up to 13 % over the change based on meteorological factors alone. However, the general shape of the yield curves over time remains similar, with an initial rise driven by the benefits of increased temperature on early growth and the effect of CO<sub>2</sub> fertilisation, followed by a decline as other factors become limiting (extreme temperature and water availability).

According to Semenov et al. (2014), the primary factors contributing to wheat yield increase are improvement in light conversion efficiency, extended duration of grain filling and optimal phenology. Maintaining and increasing yields will require resilience to adverse climatic conditions and simultaneous crop breeding to select traits for increased tolerance to heat and drought stress (in Chapter 3 of the Evidence Report for the UK second Climate Change Risk Assessment, Brown et al.



**Fig. 9.** Upper: Scatter plot of the yield before (x axis) and after (y axis) assimilation for 2002 to 2019. The equation is that of the straight blue line, derived using a least squares fit. Lower: Time series of the modelled annual mean potential winter wheat yield (in t/ha, green) and the modelled yield (in t/ha, blue) after data assimilation, for a location in Northamptonshire. The solid line represents the ensemble mean and the shading gives the minimum and maximum. The red dots are the precision yield measurements measured at the location for 2016–2019. The vertical, dashed lines indicate the years between which EO data was available and used for assimilation. Outside this range, yield after data assimilation is estimated by applying a correction factor based on the linear relationship between pre- and post- assimilation estimates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(2016)). Further, [Brown et al. \(2016\)](#) suggest that it will not be possible to adapt agriculture by simply moving southerly-adapted crop genotypes (and ‘novel’ crops) northwards. Day-length (and rate of change of day-length) plays a critical role in the timing of key developmental processes (tuberisation, bolting and flowering), which affect the yield and quality of harvested products (roots, fruits, seeds and leafy shoots). The winter wheat crop growth model used in this study includes many of these features. Additional adaptation measures might include the provision of infrastructure for water storage and crop irrigation, and management to improve soil moisture holding capacity.

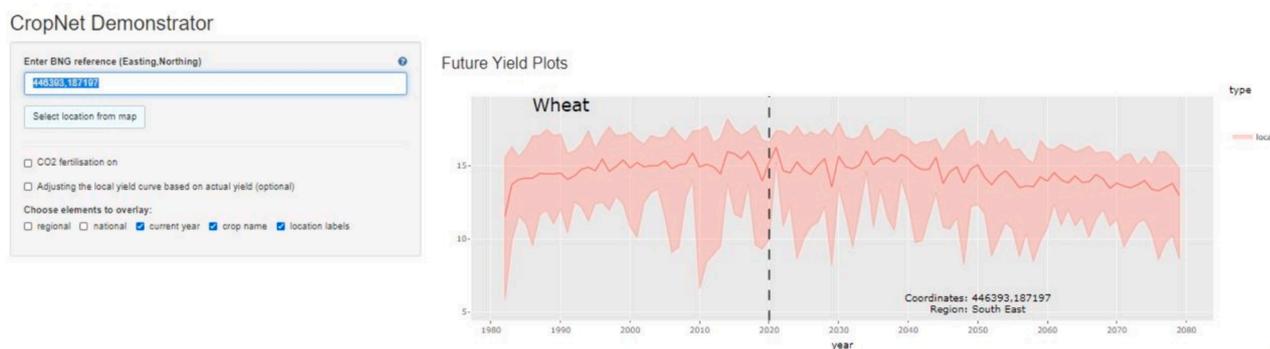
Our initial application was to consider the impact of climate change over many decades across the UK at 1 km spatial resolution. For that reason, we used a simpler modelling approach, made assumptions about sowing and harvesting dates and fixed parameters at typical values. There are a number of features of the current approach that we can improve in future developments of the model. For example, we can account for the input and impact of nutrients or allow greater flexibility in the sowing and harvesting dates. We could do this by inputting empirical or modelled data on how these vary over space and time into the model, or by using the data assimilation system to optimise these parameters. We could also use data assimilation to further increase the realism of the model by incorporating other aspects of management where relevant EO data sources exist (e.g., soil and water management). We could include other factors, such as disease incidence, which we could model independently of the climate projections (e.g., [Gouache et al. \(2013\)](#); [Juroszek and von Tiedemann \(2013\)](#)). We could adapt the model to account for their impacts on yield.

#### 4.2. Data assimilation

Previous studies have shown the ability of the coarser resolution (500 m) MODIS LAI product to improve within growing season crop yield estimates at a variety of locations ([Doraiswamy et al., 2004](#); [Wu et al., 2012](#); [Ines et al., 2013](#); [Huang et al., 2015](#)). Since 2017, the operation of the Copernicus Sentinel-1 and Sentinel-2 pairs have been providing SAR and MSI data at field-scale spatial resolutions (30 m and 10 m respectively), every 5 days, within 12 h of the acquisition. Some studies are already showing the benefit of utilising such information in combination with crop growth models ([Novelli et al., 2019](#); [Pan et al., 2019](#); [Zhuo et al., 2019](#)), to improve within growing season and long-term field-scale predictions of crop growth.

In this study, we use LAI data derived from Sentinel-2 MSI at 10 m spatial resolution and have taken a state-based approach to data assimilation, optimising the modelled LAI time series for current and past years with available Sentinel-2 MSI data. While the assimilation clearly reduces the general ‘bias’ (i.e., offset) and partially improves the regression (slope and coefficient of determination,  $R^2$ ), it does not appear to match the observed variability. The modelled yield of winter wheat before the assimilation is the maximum ‘potential’ yield based on climatic/weather factors and the soil properties only. We use the assimilation of the LAI data to account for factors, which are not modelled explicitly but which affect the crop yield, such as pest and disease, farm management practices. Strictly, the data assimilation will only capture those agronomic factors that affect LAI/GAI. Although we consider our approach to be the start of a potential operational framework, to gain full predictive capability will require knowledge of these other forcings such as pest extent and their impact on LAI. An assessment of the quality of the LAI product used for the locations studied here and its uncertainties is needed before converting the current demonstration tool into an operational UK-based framework. Another option would be to use data assimilation to optimise the parameters of our crop model to improve its predictive power of LAI, and hence yield, into the future ([Pinnington et al., 2020](#)). For example, we could adjust the thermal times and sowing dates of the crop in the model to better match the observed LAI trajectory on a field or sub-field scale.

In addition, methods of machine learning ([Suykens and Vandewalle, 1999](#)) or model emulation ([Gómez-Dans et al., 2016](#); [Fer et al., 2018](#)) could be used to improve future predicted yields on a field or sub-field scale where precision yield and Sentinel-2 data are available. This would involve training an algorithm based on the current outputs of the state-based assimilation system, precision yield data, and meteorological drivers to find an optimal estimate of yield at each individual field, given the current day covariates. Machine learning, model emulation and data assimilation for parameter estimation would allow us to predict estimates closer to the actual yield under climate change instead of the water-limited potential yield currently predicted by the model.



**Fig. 10.** Screenshot of the crop yield demonstrator showing the calculated water-limited potential yields for winter wheat between 1980 and 2080 for a location in southern England (OS grid reference 446393, 187197). The solid line is the mean yield of the ensemble and the range gives the minimum and maximum yields.

### 4.3. Extension to other crops and development of a crop-yield demonstrator

A key component of the CropNET project was to provide a demonstrator for stakeholders to visualise and explore local impacts of climate change on the yields of major crops for stakeholders. In so doing, the demonstrator aims to raise awareness of the risks posed by climate change to crop production and support the development of adaptation strategies. We co-designed the demonstrator with users from the arable and livestock & grassland farming sectors (Janes-Bassett et al., in preparation).

It uses a data cube approach to enable users to rapidly explore water-limited potential yields from the crop yield models, described in this paper, for every field in Great Britain (approximately 2 million parcels). The demonstrator is implemented on the UK JASMIN computing platform, with a web interface (<https://cropnet-demonstrator.datalabs.ceh.ac.uk/>). Fig. 10 shows a screenshot of the demonstrator, with the calculated potential yields for winter wheat between 1980 and 2080 for a location in southern England (OS grid reference 594010, 290175), using the uncorrected 12-member UKCP18 12 km RCM RCP8.5 climate projection ensemble. Users can compare local water-limited potential yields with those calculated for the surrounding region and for Great Britain. Users can also input actual yield values to provide a location-specific correction to actual yield.

Our modelling framework can readily be adapted to accommodate growth and yield models for other important crops, for which we have three key components: (a) a process-based yield model that incorporates meteorological data, (b) EO data that can be assimilated into that model; and (c) yield data of adequate resolution and accuracy for validation. We have begun to integrate crop growth models for oilseed rape and grass, which will be the subject of future publications.

## 5. Conclusions

In this paper, we describe the development and use of a crop growth model to predict climate change impacts on winter wheat. We obtained water-limited potential yields for current and future years using the latest climate projections for the UK. Our yield models suggest that much of the UK will become progressively less suitable for growing winter wheat after 2040 because of reduced summer precipitation. Although the wheat model is currently based on a general wheat crop, it could in principle be used to investigate performance of new drought-resistant varieties.

This paper provides a benchmark for future development of both the crop model and the modelling framework. Further work is needed to

improve and extend the data assimilation component (including evaluation against in situ measurements to quantify uncertainties) and to compare the data-driven approach against more process-based models to demonstrate the benefit of our approach. Our integrated approach to crop yield modelling has great potential for the application of machine learning techniques to provide continual model improvement based on the automated assimilation of field specific EO growth parameters and precision yield data.

Growth and yield models for other crops (e.g., maize, chickpeas) can readily be added to our modelling framework to give a fuller picture of the impacts of climate change on UK food production and the potential of crops not currently grown in the UK. Exploration of this framework will also provide farmers with the opportunity to design future, climate-smart crop rotations.

### Code and data availability

The crop model parameterisations and processing codes are available on request to the corresponding author. We use processing scripts developed and made available by Yin (2020) to derive leaf area index from Sentinel-2 top of atmosphere reflectance observations.

The UKCP18 data are available for download from the CEDA archive ([https://catalogue.ceda.ac.uk/?q=ukcp18&sort\\_by](https://catalogue.ceda.ac.uk/?q=ukcp18&sort_by)). The HadUK-Grid data are also available from the CEDA archive (Hollis et al., 2018). The CHES-SCAPE data are available from the CEDA archive (Robinson et al., 2022). The CHES-met data (Robinson et al., 2020) and soil data (Bell et al., 2018) are available from the NERC Environmental Information Data Centre. The precision yield data for wheat are provided to the authors on a confidential basis.

### CRedit authorship contribution statement

**Garry Hayman:** Conceptualization, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing. **John W. Redhead:** Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Matthew Brown:** Software, Validation, Visualization. **Ewan Pinnington:** Software, Writing – review & editing. **France Gerard:** Resources, Writing – review & editing. **Mike Brown:** Software, Visualization. **William Fincham:** Resources, Writing – review & editing. **Emma L. Robinson:** Resources, Writing – review & editing. **Chris Huntingford:** Resources, Writing – review & editing. **Richard F. Pywell:** Conceptualization, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The crop model parameterisations and processing codes are available on request. The precision yield data for wheat are provided to the authors on a confidential basis.

## Acknowledgements

The authors gratefully acknowledge the support of the UK Climate Resilience programme, a UK Research and Innovation's Strategic Priorities Fund programme, for the CropNET project (Monitoring and

predicting the effects of climate change on crop yields, grant NE/S01733X/2). We also acknowledge support for: (a) J.R., W.F. and R.P.: Achieving Sustainable Agricultural Systems (ASSIST, grant NE/N018125/1); (b) G.H. and M.B.: the Newton fund through the Climate Science Service Partnership – China for the VERDANT (Vegetation near real time detection and monitoring for China) project (Grant Agreement P107722); (c) E.P.: the National Centre for Earth Observation; and (d) E. R. and C.H.: The UK-SCAPE programme (award number NE/R016429/1).

We acknowledge the UK Met Office for the UKCP18 dataset used in this work and the participating farmers in the ASSIST Yield Network for access to their yield data. The CHESS-SCAPE data were produced by the SPEED project (Spatially explicit projections of environmental drivers and impacts), which is funded by the Natural Environment Research Council as part of the UK-SCAPE programme delivering National Capability (award number NE/R016429/1).

## Appendix 1. – Meteorological data

### A1.1 Observationally-based meteorological data

We use two observationally based datasets (CHESS-met and HadUK-Grid) as neither dataset fully provides the meteorological parameters required for the modelling. We require data at a spatial resolution of 1 km.

CHESS-met is an observation-based, daily meteorological dataset for Great Britain at 1 km x 1 km spatial resolution, covering the period 1961–2017 (Robinson et al., 2020) and since updated to 2019 (Robinson et al., 2023a). CHESS-met is derived from the observation-based MORECS dataset (Thompson et al., 1981; Hough and Jones, 1997), and then downscaled using information about topography. This is augmented by an independent precipitation dataset derived from station observations, Gridded Estimates of daily and monthly Areal Rainfall for the United Kingdom (CEH-GEAR; Tanguy et al. (2014); Keller et al. (2015)), along with variables from two global datasets – WATCH Forcing Data and Climate Research Unit (CRU) time series 3.21.

HadUK-Grid is a meteorological station observation-based meteorological dataset at 1 km x 1 km spatial resolution for the UK, providing, amongst other variables, daily temperature data from 1960 and monthly sunshine hours from 1929 (Hollis et al., 2018).

The daily temperature range variable (used with daily mean temperature to infer daily minimum and maximum temperature) in the CHESS-met dataset is taken from the CRU dataset, copied down to the 1 km resolution without any downscaling applied. As it is in reality at a coarser spatial resolution, we therefore chose to use the daily minimum and maximum temperature variables from the HadUK-Grid dataset.

We also require net short wave solar radiation, instead of the incoming solar radiation provided by the CHESS-met dataset. Although HadUK-Grid only provides monthly sunshine hours, we use this dataset to estimate the average cloud-cover fraction for a given month. Applying the MORECS equation 4.17 in Thompson et al. (1981), we obtain the maximum possible sunshine hours for a given month and derive the cloud fraction by dividing the actual monthly sunshine hours from HadUK-Grid by the maximum sunshine hours. We then derive the monthly average albedo, using the following equation:

$$\alpha = ((1 - dfrac) * bs\_albedo) + (dfrac * ws\_albedo)$$

where *dfrac* is the cloud cover fraction, *bs\_albedo* is the black sky albedo and *ws\_albedo* is the white-sky albedo. The black- and white-sky albedos are taken from the 1 km x 1 km GlobAlbedo dataset (Muller et al., 2012). This monthly albedo ( $\alpha$ ) was then used to convert the daily CHESS-met downwelling solar radiation to net solar radiation. Although there are other published datasets that would have all the required variables, there are none that have all the required variables at 1 km spatial resolution. Our approach here, using both HadUK-Grid and CHESS-met, allows us to obtain all the required variables at the high spatial resolution required for this study.

In summary and as indicated in Table 1, precipitation is taken from the CHESS-met dataset (note it is effectively the CEH GEAR dataset in different units). The mean air temperature is also taken from CHESS-met dataset, with the minimum and maximum daily temperature taken from the HadUK-Grid dataset. We use a combination of the HadUK-Grid and CHESS-met datasets to derive the net incoming solar radiation.

### A1.2 UKCP18 projections

The regional UKCP18 climate projections for the UK comprise a set of 12 high resolution climate projections at 12 km spatial resolution (Lowe et al., 2018). We use the ensemble for the RCP8.5 scenario, as this is the only UKCP18 scenario available at 12 km spatial resolution. Although there is a set of 12 high resolution climate projections available at 2.2 km spatial resolution, we choose not to use these as that set of projections comprises 3 time slices (1981–2000, 2021–2040 and 2061–2080), whereas the UKCP18 RCP8.5 12 km projections provide a complete time series with daily time steps from 1981 to 2080.

The RCP8.5 scenario is a higher warming scenario than the RCP2.6 scenario (Table A1.1). In the RCP8.5 scenario, the fastest rate of change in surface air temperature is at the end of the century, whereas the fastest rate of change is in the near future for the RCP2.6 scenario (Lowe et al., 2018). The rates of change over the next 20–30 years are very similar between the RCP2.6 and RCP8.5 scenarios to 2040.

**Table A1**

1. Projected change in temperature and precipitation for the UK region from 1981 to 2000 to 2041–2060, showing the 10th, 30th and 90th percentiles from the probabilistic projections. Adapted from (Lowe et al., 2018).

Scenario	Annual Temperature Change (°C)			Winter Precipitation Change (%)			Summer Precipitation Change (%)		
	10th	50th	90th	10th	50th	90th	10th	50th	90th
RCP8.5	0.9	1.8	2.7	-5	7	21	-31	-15	0
RCP2.6	0.5	2.0	2.3	-5	5	16	-24	-11	1

### A1.3 CHES-SCAPE projections and bias correction

We use all 4 members of the CHES-SCAPE 1 km ensemble for the RCP8.5 scenario (Robinson et al., 2023b). This dataset was derived from the UKCP18 RCP8.5 12 km ensemble by downscaling and has been bias corrected to the CHES-met 1 km historical climate dataset (Robinson et al., 2020). The downscaling of the three meteorological variables of interest (surface air temperature, incoming short wave solar radiation and precipitation) is based on the approach used by Robinson et al. (2017), but has been adapted for the UKCP18 climate model output (Robinson et al., 2023b). The downscaled meteorological fields are then then bias corrected to CHES-met for each 1 km grid cell and for each season (Dec-Jan-Feb (DJF), Mar-Apr-May (MAM), Jun-Jul-Aug (JJA), Sep-Oct-Nov (SON)).

#### 1 Surface temperature

**Downscaling:** Robinson et al. (2023b) first reduced the UKCP18 12 km RCM air temperature from the climate model grid box elevation (UKCP18, 2020) to mean sea level, using a lapse rate of  $-0.006 \text{ K m}^{-1}$  (Hough and Jones, 1997). They then used a bi-cubic spline to interpolate from the UKCP18 12 km resolution to the CHES-SCAPE 1 km grid. Finally, they adjusted the temperatures to the elevation of each 1 km square using the same lapse rate and grid box elevations from the Integrated Hydrological Digital Terrain Model (IHDTM) (Morris and Flavin, 1990) and the Ordnance Survey of Northern Ireland (OSNI) Open Data 50 m Digital Terrain Model (DTM) (OSNI, 2021).

**Bias-Correction:** For each CHES-SCAPE grid cell and season (DJF, MAM, JJA and SON), Robinson et al. (2023b) calculated the difference  $\mu_T = \overline{T_d} - \overline{T_m}$  between the seasonal mean daily temperature in CHES-SCAPE,  $\overline{T_d}$ , and CHES-met,  $\overline{T_m}$ , for the years 1980–2015. They removed this difference from the CHES-SCAPE data, such that the bias-corrected daily temperature is given by  $T_c = T_d - \mu_T$ , where  $T_d$  is the original CHES-SCAPE daily data. The mean of the CHES-SCAPE data is thus equal to that of the CHES-met data for the bias-correction period.

#### 2. Precipitation

**Downscaling:** Robinson et al. (2023b) downscaled the UKCP18 12 km RCM precipitation using Standardised Annual Average Rainfall 1961–90 (SAAR 61–90) (Spackman, 1993). They copied the UKCP18 12 km precipitation values to the corresponding 1 km grid box with no interpolation. They then scaled these values using the ratio of each SAAR 1 km value to its corresponding 12 km area mean (Robinson et al., 2023b).

**Bias-correction:** For each CHES-SCAPE grid cell and season, Robinson et al. (2023b) removed errors with a multiplicative scaling  $\mu_p = \overline{P_d}/\overline{P_m}$ , which is the ratio of the seasonal mean of the downscaled CHES-SCAPE precipitation to the seasonal mean CHES-met precipitation for the years 1980–2015. They then calculated the bias-corrected daily precipitation using  $P_c = P_d/\mu_p$ , where  $P_c$  is the corrected data, and  $P_d$  is the original CHES-SCAPE daily data.

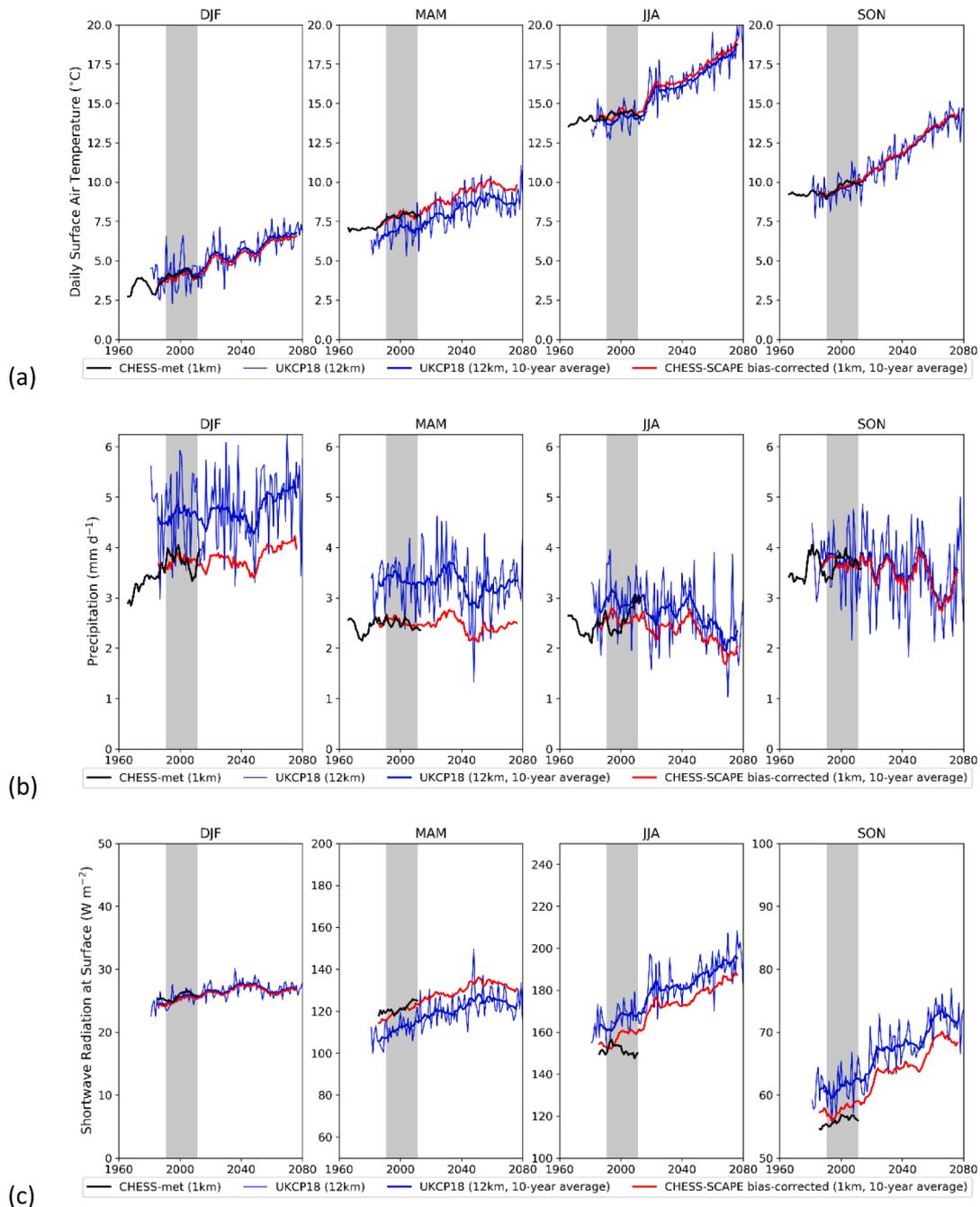
#### 3. Incoming short wave (SW) solar radiation

**Downscaling:** Robinson et al. (2023b) interpolated the UKCP18 net shortwave radiation using a bi-cubic spline. They then corrected for the average inclination and aspect of the surface, assuming that only the direct beam radiation is a function of the inclination and that the diffuse radiation is homogeneous. They assumed that the cloud cover is the dominant factor in determining the diffuse fraction (Muneer and Munawwar, 2006) and used the cloud cover grid provided by UKCP18 interpolated to 1 km using a bi-cubic spline. They calculated the aspect and slope from the IHDTM and OSNI DTM at the native resolution, following the method of Horn (1981) and then aggregated to 1 km. They scaled the direct beam radiation using the ratio of the top of atmosphere fluxes for horizontal and inclined surfaces calculated following Allen et al. (2006). They then converted this to down-welling SW radiation using the albedo calculated from the GlobAlbedo black-sky and white-sky albedo (Muller et al., 2012) and the diffuse radiation fraction.

**Bias-correction:** When considering a frequency distribution of values, Robinson et al. (2023b) found the highest attained radiation levels to be similar between CHES-met and CHES-SCAPE data. However, they found that the middle of the distribution, broadly the mean and its peak, could often be offset quite substantially. They defined two scale parameters, the first of which is the ratio of the seasonal means,  $\mu_{S1} = \overline{S_d}/\overline{S_m}$ , and the second of which is the ratio of the difference between the seasonal mean and the maximum value of CHES-SCAPE,  $\mu_{S2} = (S_d - \max(S_d))/(\overline{S_m} - \max(S_d))$ . Where the mean of CHES-SCAPE was less than the mean of CHES-met,  $\mu_{S1}$  was used to stretch the distribution such that up to the mean, CHES-SCAPE values are increased. Then, from the peak to the maximum value,  $\mu_{S2}$  was used to do the opposite and compressed the distribution into the range between the CHES-met mean and maximum value. Similarly, if the CHES-SCAPE mean was larger than that of the CHES-met data, this compresses the lower part of the distribution and stretched the upper part.

As the model requires net shortwave radiation for this study, we convert the bias-corrected CHES-SCAPE down-welling shortwave data to net shortwave radiation, using albedo data derived from GlobAlbedo (ESA, 2011) and the HadUK-Grid cloud cover fraction.

In Figure A1.1, we show, by season, the effect of the bias correction on the meteorological driving parameters (surface temperature, precipitation and the net incoming shortwave radiation at the surface).



**Fig. A1.** 1. Time series of the meteorological driving parameters used in the CropNET winter wheat model, by season (DJF, MAM, JJA, SON), for (a) surface temperature ( $^{\circ}\text{C}$ ); (b) precipitation ( $\text{mm day}^{-1}$ ) and (c) net incoming shortwave radiation at the surface ( $\text{W m}^{-2}$ ). In each plot, we show the Great Britain average of the parameters derived from the historical CHES-met data (black line), the original UKCP18 12 km data (blue), with a rolling 10-year average (solid blue line) and the bias-corrected CHES-SCAPE data (red line).

## Appendix 2. – LAI data assimilation

The cost function used in the data assimilation routine consists of three terms, each of which provides a constraint:

$$J = J_{obs} + J_{model} + J_{smooth} \quad (\text{A1})$$

### 1. The observation term

The observation term constrains the state vector to the observations and is defined as

$$J_{obs} = 0.5(\mathbf{y}_{obs} - \mathbf{H}\mathbf{x})^T \mathbf{C}_{obs}^{-1} (\mathbf{y}_{obs} - \mathbf{H}\mathbf{x}) \quad (\text{A2})$$

where  $\mathbf{y}_{obs}$  is the vector of observations (i.e., Sentinel-2 LAI). This can be of variable length, depending on the number of days with observational data. Here and in the subsequent equations,  $\mathbf{x}$  is the state vector, i.e., the time series of LAI for any particular model grid cell.  $\mathbf{H}$  is a matrix of shape [number

of days with observations by total number of days]. Each row of  $H$  will have a single 1 in it at the column index corresponding to the day through the time series at which each observation is valid. All other values in each row are zeros. Multiplying this by the state vector acts to transform the state vector into observational space, by removing the values of the state vector where there are no corresponding observations.  $C_{obs}$  is the error covariance matrix of the observations, a square matrix with the dimension length equal to the number of observations. The diagonal values correspond to the error associated with each observation value. In this application, the off-diagonal values, corresponding to correlations between observation errors, were set to zero, allowing us to rewrite the equation as:

$$J_{obs} = 0.5 \sum_{t=1}^N \frac{(y_{obs_t} - y_{mod_t})^2}{\sigma_{obs}^2} \quad (A3)$$

where  $N$  is the number of observations and  $t$  is the observation number.

From equation (A3), we see that the cost function is the sum of the squared differences between the observations and the model, weighted by the inverse of the error associated with the observations. Also note that in our application we are using a constant observation error term of  $0.4 \text{ m}^2 \text{ m}^{-2}$  that does not vary with time. This is consistent with the work of Verrelst et al. (2015), who derive an error of 0.44 for a Gaussian-process based emulator retrieval method of Sentinel-2 LAI, a similar method to that used in our study.

## 2. The model term

The model term constrains the state vector to the vector of LAI values predicted by the crop model, known as the prior, to ensure the retrieval of a physically reasonable LAI time series. This term is given by:

$$J_{model} = 0.5(\mathbf{x} - \mathbf{x}_{mod})^T C_{mod}^{-1}(\mathbf{x} - \mathbf{x}_{mod}) \quad (A4)$$

with  $\mathbf{x}_{mod}$  the model vector and  $C_{mod}$  is the error covariance matrix of the model vector. Once again this is a diagonal matrix, meaning we can rewrite the equation (A4) as:

$$J_{model} = 0.5 \sum_{t=1}^P \frac{(x_t - x_{mod_t})^2}{\sigma_{mod_t}^2} \quad (A5)$$

where  $P$  is the number of time steps in the state vector and  $\sigma_{mod}$  are the diagonal terms of  $C_{mod}$ .

For our application, we let  $\sigma_{mod}$  vary with time, as it is set to the standard deviation of the model ensemble at each time step. Where this is 0 however, such as at the end of the time series, we set it to 0.001 to avoid errors in the assimilation. This has the effect of constraining the posterior (assimilated LAI) to the prior (modelled LAI) at the start and end of the time series. We note that for both the observations and model we assume no bias in the cost function. This could be included as an additional term for the model error or within the observation error covariance matrix, as shown by Howes et al. (2017).

## 3. The smoothing term

The smoothing term constrains the state vector to a smoothed version of itself:

$$J_{smooth} = 0.5\gamma^2(\Delta\mathbf{x})^T C_{smooth}^{-1}(\Delta\mathbf{x}) \quad (A6)$$

where  $\Delta$  is a first order differential operator matrix of size  $P-1 \times P$ , formed by taking the forward difference down each separate column of an identity matrix of size  $P \times P$ ;  $C_{smooth}$  is the error covariance matrix of the smoothed state vector;  $\gamma$  is a term controlling the strength of the constraint provided by this term and thus how smooth the posterior state vector is. For our application, a value of 10 is used, as this was found to produce an optimal balance between the three terms in the cost function where the smoothing term was strong enough to smooth out all the variability arising from the observational term without dominating the minimisation routine (results not shown).  $C_{smooth}$  is once again a diagonal matrix, meaning we can write the equation:

$$J_{smooth} = 0.5\gamma^2 \sum_{t=1}^{P-1} \frac{(x_t - x_{t-1})^2}{\sigma_{smooth}^2} \quad (A7)$$

where  $\sigma_{smooth}$  are the diagonal terms of  $C_{smooth}$ . In this instance, assuming a diagonal  $C_{smooth}$ , we could combine the  $\gamma$  and  $\sigma_{smooth}$  terms as they both have the same effect on the equation. The 'error',  $\sigma_{smooth}$ , is more complex to define than the other errors but it can be understood from this equation as a parameter that constrains how much the state vector is allowed to change from one time step to the next. Here, we set it to vary according to the value of the prior state vector, with small values where this is small and vice versa, up to a maximum value of 1.5. This term therefore provides less of a constraint when the prior values of LAI are highest (which for LAI is generally when it is varying the fastest), allowing the posterior state vector to be more variable here, whilst also ensuring it is not too variable at the start of the season when LAI is low. Whilst this type of constraint could be included via the prior term, we found it difficult to achieve the desired effect using this method.

## 4. Minimisation of the cost function

From the cost function equations, it can be seen that the magnitude of each term is the sum of the squared differences between the various time series, weighted by the inverse of the errors associated with the three terms. Therefore, the terms with smaller errors contribute more to the cost function. When the cost function equation is minimised relative to the state vector, such terms provide a stronger constraint relative to the other terms. E.g., if the model (prior) term is the dominant term in the cost function, the posterior state vector, obtained by minimising the cost function, will be closer to the prior state vector than the observations, and will not be very smooth.

To minimize the cost function efficiently the gradient of the cost function with respect to the state vector is required. This can be found by differentiating the cost function,  $J$ , to obtain:

$$\nabla J = \nabla J_{obs} + \nabla J_{model} + \nabla J_{smooth} \quad (A8)$$

$$\nabla J = -\mathbf{H}^T \mathbf{C}_{obs}^{-1} (\mathbf{y}_{obs} - \mathbf{H}\mathbf{x}) + \mathbf{C}_{mod}^{-1} (\mathbf{x} - \mathbf{x}_{mod}) + \gamma^2 \Delta^T \mathbf{C}_{smooth}^{-1} \Delta \mathbf{x} \quad (\text{A9})$$

The cost function,  $J$ , and its gradient,  $\nabla J$ , are used alongside the initial guess to the LAI state vector and the MODIS LAI observations as inputs to the Broyden–Fletcher–Goldfarb–Shanno (BFGS) minimisation routine (Wright and Nocedal, 2006) in the Python package Scipy.optimize (Jones et al., 2001). The output of this minimisation routine is the optimized time series of LAI values (the posterior state vector). The crop model is then re-run with these posterior LAI values to find an improved estimate of the crop yield.

With the various assumptions made, we note that this model is linear and can be solved straightforwardly as shown. As we design the data assimilation system to be flexible for future, possibly more complex applications, we document the full system here.

The minimum occurs when  $\Delta J = 0$ .

$$0 = -\mathbf{H}^T \mathbf{C}_{obs}^{-1} (\mathbf{y}_{obs} - \mathbf{H}\mathbf{x}) + \mathbf{C}_{mod}^{-1} (\mathbf{x} - \mathbf{x}_{mod}) + \gamma^2 \Delta^T \mathbf{C}_{smooth}^{-1} \Delta \mathbf{x} \quad (\text{A10})$$

$$0 = -\mathbf{H}^T \mathbf{C}_{obs}^{-1} \mathbf{y}_{obs} - \mathbf{C}_{mod}^{-1} \mathbf{x}_{mod} + (\mathbf{H}^T \mathbf{C}_{obs}^{-1} \mathbf{H} + \mathbf{C}_{mod}^{-1} + \gamma^2 \Delta^T \mathbf{C}_{smooth}^{-1} \Delta) \mathbf{x} \quad (\text{A11})$$

$$(\mathbf{H}^T \mathbf{C}_{obs}^{-1} \mathbf{H} + \mathbf{C}_{mod}^{-1} + \gamma^2 \Delta^T \mathbf{C}_{smooth}^{-1} \Delta) \mathbf{x} = \mathbf{H}^T \mathbf{C}_{obs}^{-1} \mathbf{y}_{obs} + \mathbf{C}_{mod}^{-1} \mathbf{x}_{mod} \quad (\text{A12})$$

$$\mathbf{x} = (\mathbf{H}^T \mathbf{C}_{obs}^{-1} \mathbf{y}_{obs} + \mathbf{C}_{mod}^{-1} \mathbf{x}_{mod}) / (\mathbf{H}^T \mathbf{C}_{obs}^{-1} \mathbf{H} + \mathbf{C}_{mod}^{-1} + \gamma^2 \Delta^T \mathbf{C}_{smooth}^{-1} \Delta) \quad (\text{A13})$$

### 5 Calculation of the posterior error

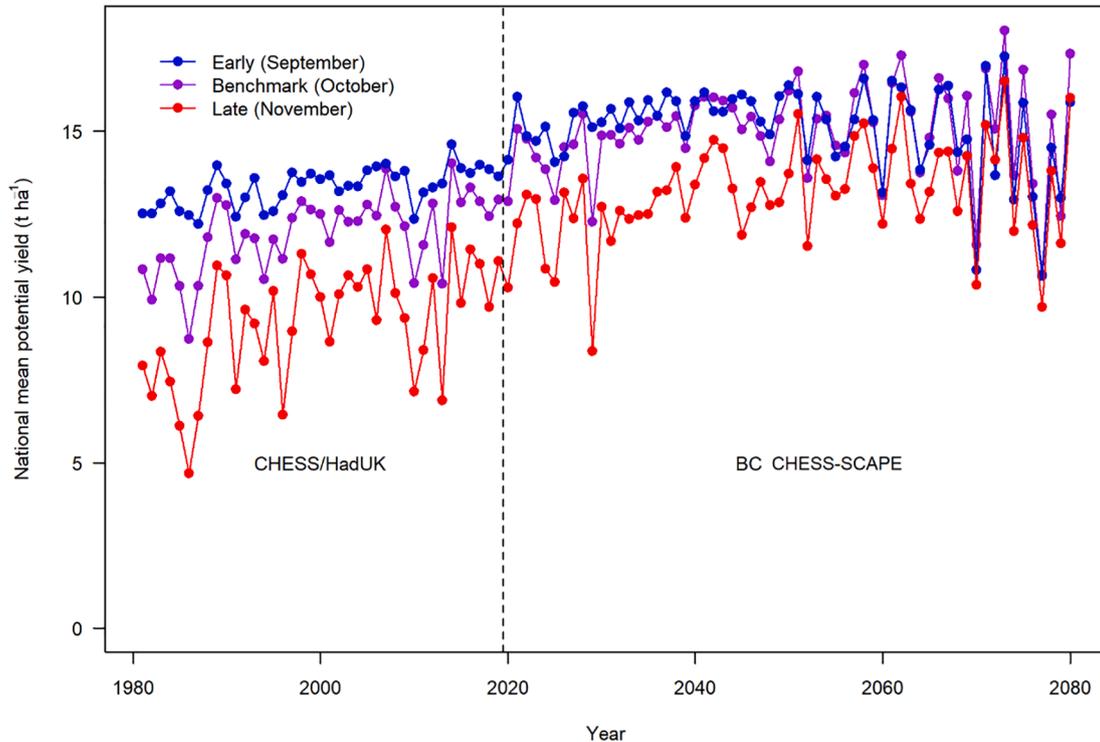
The standard deviation associated with the posterior – the resulting LAI from the assimilation – can be calculated using the second derivative of the cost function as follows:

$$\mathbf{A} = (\Delta^2 J)^{-1} = (\mathbf{H}^T \mathbf{C}_{obs}^{-1} \mathbf{H} + \mathbf{C}_{mod}^{-1} + \gamma^2 \Delta^T \mathbf{C}_{smooth}^{-1} \Delta)^{-1} \quad (\text{A14})$$

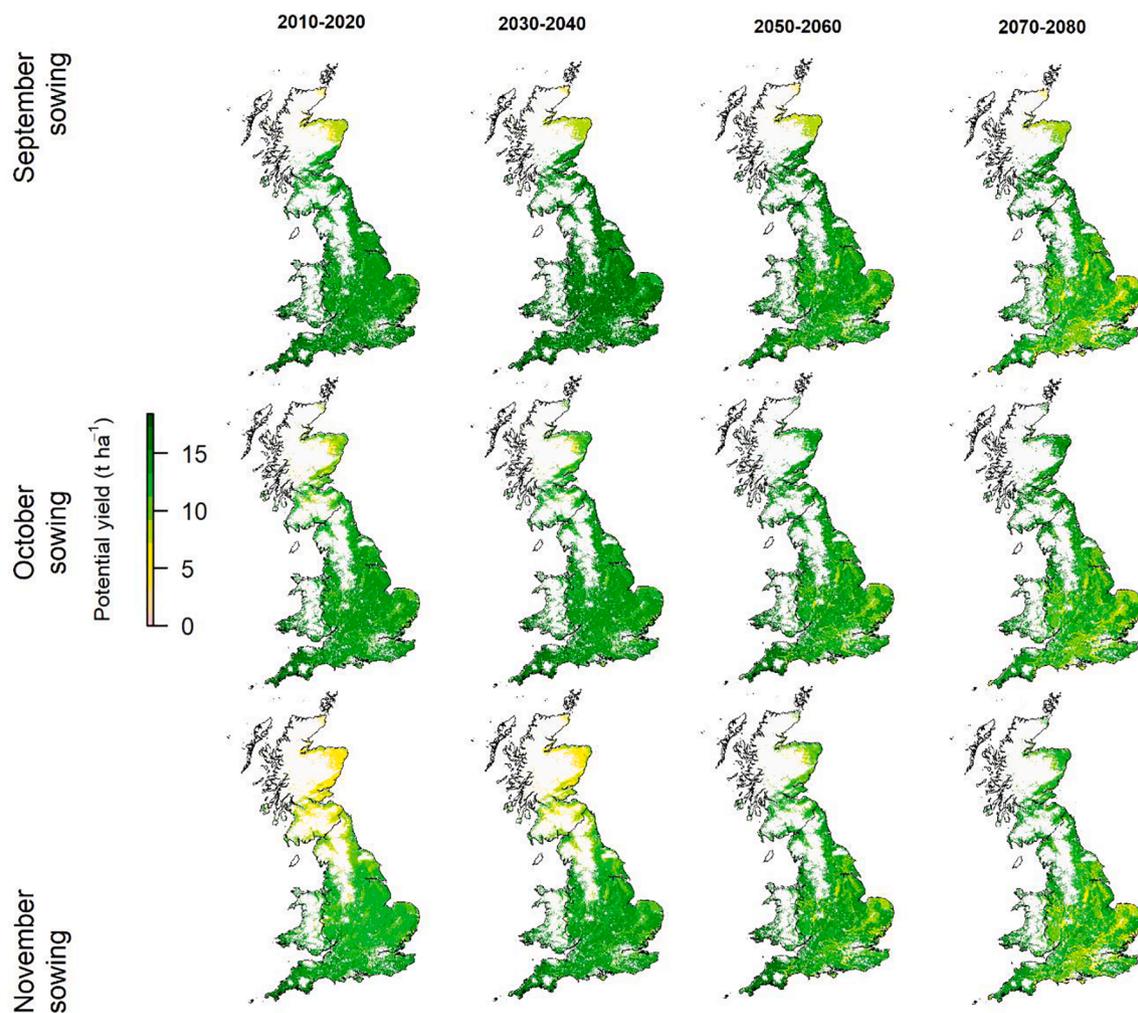
where  $\mathbf{A}$  is the analysis error covariance matrix, and all other variables are as defined previously. The diagonal of  $\mathbf{A}$  is the variance of the posterior estimate of LAI. We take the square root of this to obtain the standard deviation.

## Appendix 3. – Sensitivity to sowing date and heat stress

### 1. Sensitivity to sowing date

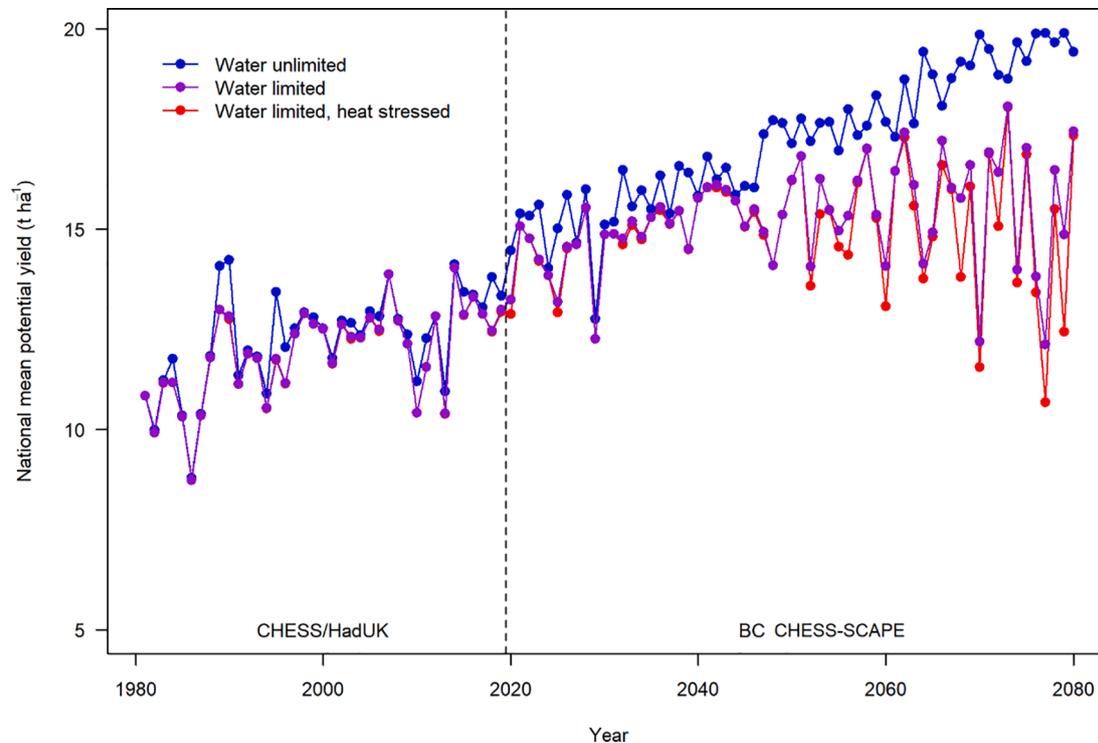


**Fig. A2.** 1. Time series of modelled national mean water-limited potential yield (in t/ha) per year, as modelled with three different sowing dates. All three model runs included waterlogging and heat stress penalties. Climatic inputs to the model were derived from CHES-met and HadUK-Grid for the historic period (pre-2020) and from bias corrected CHES-SCAPE RCP8.5 data for the future time period.



**Fig. A3.** 2. Maps of water-limited potential yields of winter wheat (in  $t/ha$ ) modelled for Great Britain for the past decade (2010–2020, first column) and for future decades: 2030–2040 (second column), 2050–2060 (third column) and 2070–2080 (final column). The upper, middle and lower panels use sowing dates of September, October and November, respectively. Climatic inputs to the model were derived from CHES-met and HadUK-Grid for the historic period (pre-2020) and from bias corrected CHES-SCAPE RCP8.5 data for the future time period.

## 2. Sensitivity to heat stress



**Fig. A4.** 3. Time series of the modelled national mean potential yield (in t/ha) per year. Modelled yields are shown with no water limitation (i.e. modelled from temperature and solar radiation only, with no effect of precipitation), water limited (i.e. including the effects of precipitation, soil water content and waterlogging) and water limited with the addition of the heat stress penalty. All three model runs used the benchmark (October) sowing date. Climatic inputs to the model were derived from CHES-met and HadUK-Grid for the historic period (pre-2020) and from bias corrected CHES-SCAPE RCP8.5 data for the future time period.

## References

- Allen, R.G., Trezza, R., Tasumi, M., 2006. Analytical integrated functions for daily solar radiation on slopes. *Agric. For. Meteorol.* 139, 55–73. <https://doi.org/10.1016/j.agrformet.2006.05.012>.
- Basso, B., Cammarano, D., Carfagna, E., 2013. Review of Crop Yield Forecasting Methods and Early Warning Systems, in Proceedings of the First Meeting of the Scientific Advisory Committee of the Global Strategy to Improve Agricultural and Rural Statistics, FAO, Rome. Available from: [http://www.fao.org/fileadmin/templates/ess/documents/meetings\\_and\\_workshops/GS\\_SAC\\_2013/Improving\\_methods\\_for\\_crops\\_estimates/Crop\\_Yield\\_Forecasting\\_Methods\\_and\\_Early\\_Warning\\_Systems\\_Lit\\_rev\\_ew.pdf](http://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/GS_SAC_2013/Improving_methods_for_crops_estimates/Crop_Yield_Forecasting_Methods_and_Early_Warning_Systems_Lit_rev_ew.pdf) (last accessed, November 2022).
- Bauer, M.E., 1985. Spectral inputs to crop identification and condition assessment. *Proc. IEEE* 73, 1071–1085. <https://doi.org/10.1109/PROC.1985.13238>.
- Bell, V.A., Rudd, A.C., Kay, A.L., Davies, H.N., 2018. Grid-to-Grid model estimates of monthly mean flow and soil moisture for Great Britain (1960 to 2015): observed driving data [MaRIUS-G2G-MORECS-monthly]. Available from NERC Environmental Information Data Centre, <https://doi.org/10.5285/e911196a-b371-47b1-968c-661eb600d83b> (last accessed November 2022).
- Bloom, A.A., Exbrayat, J.-F., van der Velde, I.R., Feng, L., Williams, M., 2016. The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proc. Natl. Acad. Sci.* 113, 1285–1290. <https://doi.org/10.1073/pnas.1515160113>.
- Brown, I., Thompson, D., Bardgett, R., Berry, P., Crute, I., Morison, J., Morecroft, M., Pinnegar, J., Reeder, T., Topp, K., 2016. UK Climate Change Risk Assessment Evidence Report: Chapter 3, Natural Environment and Natural Assets. Report prepared for the Adaptation Sub-Committee of the Committee on Climate Change, London. Available from: <https://www.theccc.org.uk/wp-content/uploads/2016/07/UK-CCRA-2017-Chapter-3-Natural-environment-and-natural-assets.pdf>, (last accessed November 2022).
- Clevers, J.G.P.W., van Leeuwen, H.J.C., 1996. Combined use of optical and microwave remote sensing data for crop growth monitoring. *Remote Sens. Environ.* 56, 42–51. [https://doi.org/10.1016/0034-4257\(95\)00227-8](https://doi.org/10.1016/0034-4257(95)00227-8).
- Defra, 2019. Farming Statistics – provisional arable crop areas, yields and livestock populations at 1 June 2019 United Kingdom, Department for Environment, Food and Rural Affairs. Available from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/865769/structure-jun2019final-uk-22jan20-rev\\_v2.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/865769/structure-jun2019final-uk-22jan20-rev_v2.pdf) (last accessed November 2022).
- Defra, 2020a. Agriculture in the United Kingdom 2019, Produced by: Department for Environment, Food and Rural Affairs, Department of Agriculture, Environment and Rural Affairs (Northern Ireland), Welsh Government, Knowledge and Analytical Services, and The Scottish Government, Rural and Environment Science and Analytical Services, Available from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/904024/AUK\\_2019\\_27Jul\\_y2020.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/904024/AUK_2019_27Jul_y2020.pdf) (last accessed November 2022).
- Defra, 2020b. Farming Statistics – provisional arable crop areas, yields and livestock populations at 1 June 2020 United Kingdom, Department for Environment, Food and Rural Affairs. Available from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/931104/structure-jun2020pro-v-UK-08oct20i.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/931104/structure-jun2020pro-v-UK-08oct20i.pdf) (last accessed November 2022).
- Demarty, J., Chevallier, F., Friend, A.D., Viovy, N., Piao, S., Ciais, P., 2007. Assimilation of global MODIS leaf area index retrievals within a terrestrial biosphere model. *Geophys. Res. Lett.* 34, L15402. <https://doi.org/10.1029/2007GL030014>.
- Dente, L., Satalino, G., Mattia, F., Rinaldi, M., 2008. Assimilation of leaf area index derived from ASAR and MERIS data into CERES-Wheat model to map wheat yield. *Remote Sens. Environ.* 112, 1395–1407. <https://doi.org/10.1016/j.rse.2007.05.023>.
- Di Paola, A., Valentini, R., Santini, M., 2016. An overview of available crop growth and yield models for studies and assessments in agriculture. *J. Sci. Food Agric.* 96, 709–714. <https://doi.org/10.1002/jsfa.7359>.
- Doraiswamy, P.C., Hatfield, J.L., Jackson, T.J., Akhmedov, B., Prueger, J., Stern, A., 2004. Crop condition and yield simulations using Landsat and MODIS. *Remote Sens. Environ.* 92, 548–559. <https://doi.org/10.1016/j.rse.2004.05.017>.
- Entekhabi, D., Reichle, R.H., Koster, R.D., Crow, W.T., 2010. Performance metrics for soil moisture retrievals and application requirements. *J. Hydrometeorol.* 11, 832–840. <https://doi.org/10.1175/2010jhm1223.1>.
- ESA, 2011. Global Albedo: Global Albedo data of the land surface. European Space Agency. Available from: <http://www.globalbedo.org> (last accessed November 2022).
- FAO, 2018. FAO'S Work on Climate Change, United Nations Climate Change Conference 2018. Available from: <http://www.fao.org/3/CA2607EN/ca2607en.pdf> (last accessed November 2022).
- Fer, I., Kelly, R., Moorcroft, P.R., Richardson, A.D., Cowdery, E.M., Dietze, M.C., 2018. Linking big models to big data: efficient ecosystem model calibration through Bayesian model emulation. *Biogeosciences* 15, 5801–5830. <https://doi.org/10.5194/bg-15-5801-2018>.
- Fincham, W.N.W., Redhead, J.W., Woodcock, B.A., Pywell, R.F., 2023. Exploring drivers of within-field crop yield variation using a national precision yield network. *J. Appl. Ecol.* 60, 319–329. <https://doi.org/10.1111/1365-2664.14323>.
- Gómez-Dans, J.L., Lewis, P.E., Disney, M., 2016. Efficient emulation of radiative transfer codes using Gaussian processes and application to land surface parameter inferences. *Remote Sens. (Basel)* 8, 119. <https://doi.org/10.3390/rs8020119>.

- Gouache, D., Bensadoun, A., Brun, F., Pagé, C., Makowski, D., Wallach, D., 2013. Modelling climate change impact on Septoria tritici blotch (STB) in France: Accounting for climate model and disease model uncertainty. *Agric. For. Meteorol.* 170, 242–252. <https://doi.org/10.1016/j.agrformet.2012.04.019>.
- Hall, F.G., Hogg, R.C., Caudill, C.E., 1981. *AgRISTARS - Plans and First-Year Achievements*. International Geoscience and Remote Sensing Symposium, Washington, DC, USA.
- Hammond, A.L., 1975. Crop forecasting from space: Toward a global food watch. *Science* 188, 434–436. <https://doi.org/10.1126/science.188.4187.434>.
- Hijmans, R.J., Williams, E., Vennes, C., 2017. R package 'geosphere', Spherical Trigonometry, v1.5-7, Available from: <https://mran.microsoft.com/snapshot/2017-11-26/web/packages/geosphere/index.html> (accessed June 2021).
- Hollis, D., McCarthy, M., Kendon, M., Legg, T., Simpson, I., 2018. HadUK-Grid gridded and regional average climate observations for the UK. Centre for Environmental Data Analysis, <http://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb>.
- Hoogenboom, G., 2000. Contribution of agrometeorology to the simulation of crop production and its applications. *Agric. For. Meteorol.* 103, 137–157. [https://doi.org/10.1016/S0168-1923\(00\)00108-8](https://doi.org/10.1016/S0168-1923(00)00108-8).
- Horn, B.K.P., 1981. Hill shading and the reflectance map. *Proc. IEEE* 69, 14–47. <https://doi.org/10.1109/PROC.1981.11918>.
- Hough, M.N., Jones, R.J.A., 1997. The United Kingdom Meteorological Office rainfall and evaporation calculation system: MORECS version 2.0-an overview. *Hydrol. Earth Syst. Sci.* 1, 227–239. <https://doi.org/10.5194/hess-1-227-1997>.
- Howes, K.E., Fowler, A.M., Lawless, A.S., 2017. Accounting for model error in strong-constraint 4D-Var data assimilation. *Q. J. R. Meteorol. Soc.* 143, 1227–1240. <https://doi.org/10.1002/qj.2996>.
- Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., Su, W., Zhang, X., Zhu, D., Wu, W., 2015. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agric. For. Meteorol.* 204, 106–121. <https://doi.org/10.1016/j.agrformet.2015.02.001>.
- Huang, J., Gómez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., Liang, S., Chen, Z.S., Xue, J.-H., Wu, Y., Zhao, F., Wang, J., Xie, X., 2019. Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agric. For. Meteorol.* 276–277, 107609. <https://doi.org/10.1016/j.agrformet.2019.06.008>.
- Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W., Rowland, C.S., 2019. High resolution wheat yield mapping using Sentinel-2. *Remote Sens. Environ.* 233, 111410. <https://doi.org/10.1016/j.rse.2019.111410>.
- Idso, S.B., Jackson, R.D., Reginato, R.J., 1977. Remote-sensing of crop yields. *Science* 196, 19–25. <https://doi.org/10.1126/science.196.4285.19>.
- Ines, A.V.M., Das, N.N., Hansen, J.W., Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sens. Environ.* 138, 149–164. <https://doi.org/10.1016/j.rse.2013.07.018>.
- Jackson, R., 1984. Remote Sensing Of Vegetation Characteristics For Farm Management, 1984 Technical Symposium East, SPIE 0475, <https://doi.org/10.1117/12.966243> (last accessed November 2022).
- Jones, E., Oliphant, T., and Peterson, P., 2001. SciPy: Open Source Scientific Tools for Python.
- Juroszek, P., von Tiedemann, A., 2013. Climate change and potential future risks through wheat diseases: A review. *Eur. J. Plant Pathol.* 136, 21–33. <https://doi.org/10.1007/s10658-012-0144-9>.
- Kamenova, I., Dimitrov, P., 2021. Evaluation of Sentinel-2 vegetation indices for prediction of LAI, fAPAR and fCover of winter wheat in Bulgaria. *Eur. J. Remote Sens.* 54, 89–108. <https://doi.org/10.1080/22797254.2020.1839359>.
- Kasampalis, D.A., Alexandridis, T.K., Deva, C., Challinor, A., Moshou, D., Zalidis, G., 2018. Contribution of remote sensing on crop models: A review. *J. Imaging* 4, 52–70. <https://doi.org/10.3390/jimaging4040052>.
- Keller, V.D.J., Tanguy, M., Prosdociimi, I., Terry, J.A., Hitt, O., Cole, S.J., Fry, M., Morris, D.G., Dixon, H., 2015. CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications. *Earth Syst. Sci. Data* 7, 143–155. <https://doi.org/10.5194/essd-7-143-2015>.
- Lewis, P., Gómez-Dans, J., Kaminski, T., Settle, J., Quaipe, T., Gubron, N., Styles, J., Berger, M., 2012. An earth observation land data assimilation system (EO-LDAS). *Remote Sens. Environ.* 120, 219–235. <https://doi.org/10.1016/j.rse.2011.12.027>.
- Liu, B., Asseng, S., Liu, L., Tang, L., Cao, W., Zhu, Y., 2016. Testing the responses of four wheat crop models to heat stress at anthesis and grain filling. *Glob. Chang. Biol.* 22, 1890–1903. <https://doi.org/10.1111/gcb.13212>.
- Lowe, J.A., Bernie, D., Bett, P., Bricheno, L., Brown, S., Calvert, D., Clark, R., Eagle, K., Edwards, T., Fosser, G., Fung, F., Gohar, L., Good, P., Gregory, J., Harris, G., Howard, T., Kaye, N., Kendon, E., Krjinen, J., Maisey, P., McDonald, R., McInnes, R., McSweeney, C., Mitchell, J. F. B., Murphy, J., Palmer, M., Roberts, C., Robinson, J., Sexton, D., Thornton, H., Tinker, J., Tucker, S., Yamazaki, K., Belcher, S., 2018. UKCP18 Science Overview Report, UK Met Office, Available from <https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Overview-report.pdf> (last accessed November 2022).
- Lynch, J.P., Fealy, R., Doyle, D., Black, L., Spink, J., 2017. Assessment of water-limited winter wheat yield potential at spatially contrasting sites in Ireland using a simple growth and development model, *Irish Journal of Agricultural and Food Research*, 56, 65–76, Available from <https://www.jstor.org/stable/26609794> (last accessed November 2022).
- Ma, T., Duan, Z., Li, R., Song, X., 2019. Enhancing SWAT with remotely sensed LAI for improved modelling of ecohydrological process in subtropics. *J. Hydrol.* 570, 802–815. <https://doi.org/10.1016/j.jhydrol.2019.01.024>.
- Meinshausen, M., Smith, S.J., Calvin, K., Daniel, J.S., Kainuma, M.L.T., Lamarque, J.F., Matsumoto, K., Montzka, S.A., Raper, S.C.B., Riahi, K., Thomson, A., Velders, G.J.M., van Vuuren, D.P.P., 2011. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* 109, 213. <https://doi.org/10.1007/s10584-011-0156-z>.
- Mercier, A., Betbeder, J., Baudry, J., Le Roux, V., Spicher, F., Lacoux, J., Roger, D., Hubert-Moy, L., 2020. Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages. *ISPRS J. Photogramm. Remote Sens.* 163, 231–256. <https://doi.org/10.1016/j.isprsjprs.2020.03.009>.
- Moran, M.S., Inoue, Y., Barnes, E.M., 1997. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sens. Environ.* 61, 319–346. [https://doi.org/10.1016/S0034-4257\(97\)00045-X](https://doi.org/10.1016/S0034-4257(97)00045-X).
- Morison, J., Matthews, R., 2016. *Agriculture and Forestry: Climate Change Impacts, Report Card 2016*, NERC Living With Environmental Change, Available from: <http://ps://nerc.ukri.org/research/partnerships/ride/lwec/report-cards/agriculture/>.
- Morris, D.G., Flavin, R.W., 1990. A digital terrain model for hydrology, 4th International Symposium of Spatial Data Handling, Zürich, Switzerland, pp. 250–262.
- Muller, J.-P., López, G., Watson, G., Shane, N., Kennedy, T., Yuen, P., Lewis, P., Fischer, J., Gunter, L., Domench, C., Preusker, R., North, P., Heckel, A., Danne, O., Krämer, U., Zühlke, M., Brockmann, C., Pinnock, S., 2012. The ESA GlobAlbedo Project for mapping the Earth's land surface albedo for 15 Years from European Sensors, IEEE Geoscience and Remote Sensing Symposium (IGARSS), pp. 22-27.7.12, Munich, Germany.
- Munee, T., Munawwar, S., 2006. Potential for improvement in estimation of solar diffuse irradiance. *Energy Convers. Manage.* 47, 68–86. <https://doi.org/10.1016/j.enconman.2005.03.015>.
- Novelli, F., Spiegel, H., Sandén, T., Vuolo, F., 2019. Assimilation of sentinel-2 leaf area index data into a physically-based crop growth model for yield estimation. *Agronomy* 9, 255. <https://doi.org/10.3390/agronomy9050255>.
- Olgun, M., Metin Kumlay, A., Cemal Adiguzel, M., Caglar, A., 2008. The effect of waterlogging in wheat (*T. aestivum* L.). *Acta Agric. Scand., Sect. B - Soil Plant Sci.* 58, 193–198. <https://doi.org/10.1080/09064710701794024>.
- OSNI, 2021. Open Data 50m Digital Terrain Model (DTM) for Northern Ireland. Ordnance Survey of Northern Ireland. Available from: <https://www.opendatani.gov.uk/dataset/osni-open-data-50m-dtm> (data accessed April 2021).
- Pan, H., Chen, Z., de Wit, A., Ren, J., 2019. Joint assimilation of leaf area index and soil moisture from Sentinel-1 and Sentinel-2 data into the WOFOST model for winter wheat yield estimation. *Sensors* 19, 3161. <https://doi.org/10.3390/s19143161>.
- Pinnington, E., Quaipe, T., Lawless, A., Williams, K., Arkebauer, T., Scoby, D., 2020. The land variational ensemble data assimilation framework: LAVENDAR v1.0.0. *Geosci. Model Dev.* 13, 55–69. <https://doi.org/10.5194/gmd-13-55-2020>.
- Pywell, R.F., Heard, M.S., Woodcock, B.A., Hinsley, S., Ridding, L., Nowakowski, M., Bullock, J., 2015. M.: Wildlife-friendly farming increases crop yield: Evidence for ecological intensification. *Proc. R. Soc. B Biol. Sci.* 282, 20151740. <https://doi.org/10.1098/rspb.2015.1740>.
- Raoult, N.M., Jupp, T.E., Cox, P.M., Luke, C., 2016. M.: Land-surface parameter optimisation using data assimilation techniques: the adJULES system V1.0. *Geosci. Model Dev.* 9, 2833–2852. <https://doi.org/10.5194/gmd-9-2833-2016>.
- Redhead, J.W., Oliver, T.H., Woodcock, B.A., Pywell, R.F., 2020. The influence of landscape composition and configuration on crop yield resilience. *J. Appl. Ecol.* 57, 2180–2190. <https://doi.org/10.1111/1365-2664.13722>.
- Robinson, E.L., Blyth, E.M., Clark, D.B., Finch, J., Rudd, A.C., 2017. Trends in atmospheric evaporative demand in Great Britain using high-resolution meteorological data. *Hydrol. Earth Syst. Sci.* 1189–1224. <https://doi.org/10.5194/hess-21-1189-2017>.
- Robinson, E.L., Blyth, E.M., Clark, D.B., Comyn-Platt, E., Rudd, A.C., 2020. Climate hydrology and ecology research support system meteorology dataset for Great Britain (1961–2017) [CHESS-met]. NERC Environmental Information Data Centre. Available from: <https://doi.org/10.5285/2ab15bf0-ad08-415c-ba64-831168be7293> (last accessed, November 2022).
- Robinson, E.L., Huntingford, C., Shamsudheen, S.V., Bullock, J.M., 2022. CHESS-SCAPE: Future projections of meteorological variables at 1 km resolution for the United Kingdom 1980–2080 derived from UK Climate Projections 2018. The NERC EDS Centre for Environmental Data Analysis. Available from <https://doi.org/10.5285/8194b416cbee482b89e0dfbe17c5786c> (last access November 2022).
- Robinson, E.L., Blyth, E.M., Clark, D.B., Comyn-Platt, E., Rudd, A.C., Wiggins, M., 2023a. Climate hydrology and ecology research support system meteorology dataset for Great Britain (1961–2019) [CHESS-met]. NERC EDS Environmental Information Data Centre.
- Robinson, E.L., Huntingford, C., Semeena, V.S., Bullock, J.M., 2023b. CHESS-SCAPE: High-resolution future projections of multiple climate scenarios for the United Kingdom derived from downscaled United Kingdom Climate Projections 2018 regional climate model output. *Earth Syst. Sci. Data* 15, 5371–5401. <https://doi.org/10.5194/essd-15-5371-2023>.
- Rowland, C.S., Morton, R.D., Carrasco, L., McShane, G., O'Neil, A.W., Wood, C.M., 2017. Land Cover Map 2015 (1km percentage target class, GB). Available from the NERC Environmental Information Data Centre, <https://doi.org/10.5285/505d1e0c-ab60-4a60-b448-68c5bbae403e> (last accessed November 2022).
- Semenov, M.A., Stratonovitch, P., Alghabari, F., Gooding, M.J., 2014. Adapting wheat in Europe for climate change. *J. Cereal Sci.* 59, 245–256. <https://doi.org/10.1016/j.jcs.2014.01.006>.
- Senapati, N., Brown, H.E., Semenov, M.A., 2019. Raising genetic yield potential in high productive countries: Designing wheat ideotypes under climate change. *Agric. For. Meteorol.* 271, 33–45. <https://doi.org/10.1016/j.agrformet.2019.02.025>.
- Shannon, D.K., Clay, D.E., Kitchen, N.R., 2018. Precision Agriculture Basics, <https://doi.org/10.2134/precisionagbasics.2018.frontmatter> (last accessed November 2022).

- Shearman, V.J., Sylvester-Bradley, R., Scott, R.K., Foulkes, M.J., 2005. Physiological processes associated with wheat yield progress in the UK. *Crop Sci.* 45 <https://doi.org/10.2135/cropsci2005.0175a>.
- Spackman, E., 1993. Calculation and Mapping of Rainfall Averages for 1961-90. University of Salford, Manchester, 15 December 1993.
- Spink, J.H., Semere, T., Sparkes, D.L., Whaley, J.M., Foulkes, M.J., Clare, R.W., Scott, R.K., 2000. Effect of sowing date on the optimum plant density of winter wheat. *Ann. Appl. Biol.* 137, 179–188. <https://doi.org/10.1111/j.1744-7348.2000.tb00049.x>.
- Suykens, J.A.K., Vandewalle, J., 1999. Least squares support vector machine classifiers. *Neural Process. Lett.* 9, 293–300. <https://doi.org/10.1023/A:1018628609742>.
- Sylvester-Bradley, R., Berry, P., Blake, J., Jindred, D., Spink, J., Bingham, I., McVittie, J., Foulkes, J., 2008. Wheat growth guide, AHDB Cereals & Oilseeds, Stoneleigh Park, Warwickshire, Available from: [http://www.adlib.ac.uk/resources/000/265/686/WGG\\_2008.pdf](http://www.adlib.ac.uk/resources/000/265/686/WGG_2008.pdf) (last accessed November 2022), AHDB Cereals & Oilseeds, Stoneleigh Park, Warwickshire.
- Sylvester-Bradley, R., Berry, P., Blake, J., Jindred, D., Spink, J., Bingham, I., McVittie, J., Foulkes, J., 2015. Wheat growth guide. AHDB Cereals & Oilseeds, Stoneleigh Park, Warwickshire.
- Sylvester-Bradley, R., Kindred, D., 2014. The Yield Enhancement Network: Philosophy and results from the first season. *Asp. Appl. Biol.* 125, 53–62.
- Talagrand, O., Courtier, P., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Q. J. R. Meteorolog. Soc.* 113, 1311–1328. <https://doi.org/10.1002/qj.49711347812>.
- Tanguy, M., Dixon, H., Prosdocimi, I., Morris, D.G., Keller, V.D.J., 2014. Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890-2012) [CEH-GEAR]. Available from the NERC Environmental Information Data Centre, <https://doi.org/10.5285/5dc179dc-f692-49ba-9326-a6893a503f6e> (last accessed November 2022).
- Thompson, N., Barrie, I.A., Ayles, M., 1981. The Meteorological Office Rainfall and Evaporation Calculation System (MORECS), Hydrological Memorandum 45. Met. Office, Bracknell, U.K.
- UKCP18, 2020. UKCP18 land projections: 12km regional climate model, surface altitude (m) over Europe. Centre for Environmental Data Analysis. Available from: [http://datap.ceda.ac.uk/badc/ukcp18/data/land-rcm/ancil/orog/orog\\_land-rcm\\_uk\\_12km\\_osgb.nc](http://datap.ceda.ac.uk/badc/ukcp18/data/land-rcm/ancil/orog/orog_land-rcm_uk_12km_osgb.nc) (last accessed November 2022).
- Vanuytrecht, E., Raes, D., Willems, P., Geerts, S., 2012. Quantifying field-scale effects of elevated carbon dioxide concentration on crops. *Climate Res.* 54, 35–47. <https://doi.org/10.3354/cr01096>.
- Verrelst, J., Rivera, J.P., Veroustraete, F., Muñoz-Marí, J., Clevers, J.G.P.W., Camps-Valls, G., Moreno, J., 2015. Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods – A comparison. *ISPRS J. Photogramm. Remote Sens.* 108, 260–272. <https://doi.org/10.1016/j.isprsjprs.2015.04.013>.
- Wright, S.J., Nocedal, J., 2006. Numerical optimization, Springer Science, Second Edition, ISBN-10: 0-387-30303-0.
- Wu, S., Huang, J., Liu, X., Fan, J., Ma, G., Zou, J., 2012. Assimilating MODIS-LAI into Crop Growth Model with EnKF to Predict Regional Crop Yield. In: *Computer and Computing Technologies in Agriculture V* (Editors: D. Li and Y. Chen), 410-418, Berlin, Heidelberg, Available from [https://doi.org/10.1007/978-3-642-27275-2\\_46](https://doi.org/10.1007/978-3-642-27275-2_46) (last accessed November 2022).
- Yin, F., Lewis, P.E., Gómez-Dans, J.L., 2022. Bayesian atmospheric correction over land: Sentinel-2/MSI and Landsat 8/OLI. *Geosci. Model Dev.* 15, 7933–7976. <https://doi.org/10.5194/gmd-15-7933-2022>.
- Yin, F., 2020. From Sentinel 2 TOA reflectance to LAI, Available from: [https://github.com/MarcYin/S2\\_TOA\\_TO\\_LAI](https://github.com/MarcYin/S2_TOA_TO_LAI) (last accessed November 2022).
- Zhuo, W., Huang, J., Li, L., Zhang, X., Ma, H., Gao, X., Huang, H., Xu, B., Xiao, X., 2019. Assimilating soil moisture retrieved from Sentinel-1 and Sentinel-2 data into WOFOST model to improve winter wheat yield estimation. *Remote Sens. (Basel)* 11, 1618. <https://doi.org/10.3390/rs11131618>.