



# Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces

N. Pirttioja<sup>1,\*</sup>, T. R. Carter<sup>1</sup>, S. Fronzek<sup>1</sup>, M. Bindi<sup>2</sup>, H. Hoffmann<sup>3</sup>,  
T. Palosuo<sup>4</sup>, M. Ruiz-Ramos<sup>5</sup>, F. Tao<sup>4</sup>, M. Trnka<sup>6,7</sup>, M. Acutis<sup>8</sup>, S. Asseng<sup>9</sup>, P. Baranowski<sup>10</sup>,  
B. Basso<sup>11</sup>, P. Bodin<sup>12</sup>, S. Buis<sup>13</sup>, D. Cammarano<sup>14</sup>, P. Deligios<sup>15</sup>, M.-F. Destain<sup>16</sup>,  
B. Dumont<sup>16</sup>, F. Ewert<sup>3</sup>, R. Ferrise<sup>2</sup>, L. François<sup>16</sup>, T. Gaiser<sup>3</sup>, P. Hlavinka<sup>6,7</sup>, I. Jacquemin<sup>16</sup>,  
K. C. Kersebaum<sup>17</sup>, C. Kollas<sup>17</sup>, J. Krzyszczak<sup>10</sup>, I. J. Lorite<sup>18</sup>, J. Minet<sup>16</sup>, M. I. Minguez<sup>5</sup>,  
M. Montesino<sup>19</sup>, M. Moriondo<sup>20</sup>, C. Müller<sup>21</sup>, C. Nendel<sup>17</sup>, I. Öztürk<sup>22</sup>, A. Perego<sup>8</sup>,  
A. Rodríguez<sup>5</sup>, A. C. Ruane<sup>23,24</sup>, F. Ruget<sup>13</sup>, M. Sanna<sup>8</sup>, M. A. Semenov<sup>25</sup>, C. Slawinski<sup>10</sup>,  
P. Stratonovitch<sup>25</sup>, I. Supit<sup>26</sup>, K. Waha<sup>21,27</sup>, E. Wang<sup>28</sup>, L. Wu<sup>29</sup>, Z. Zhao<sup>28,30</sup>, R. P. Rötter<sup>4</sup>

<sup>1</sup>Finnish Environment Institute (SYKE), 00251 Helsinki, Finland

<sup>2-30</sup>See title page of the Supplement at [www.int-res.com/articles/suppl/c065p087\\_supp.pdf](http://www.int-res.com/articles/suppl/c065p087_supp.pdf) for full list of author addresses

**ABSTRACT:** This study explored the utility of the impact response surface (IRS) approach for investigating model ensemble crop yield responses under a large range of changes in climate. IRSs of spring and winter wheat *Triticum aestivum* yields were constructed from a 26-member ensemble of process-based crop simulation models for sites in Finland, Germany and Spain across a latitudinal transect. The sensitivity of modelled yield to systematic increments of changes in temperature (−2 to +9°C) and precipitation (−50 to +50%) was tested by modifying values of baseline (1981 to 2010) daily weather, with CO<sub>2</sub> concentration fixed at 360 ppm. The IRS approach offers an effective method of portraying model behaviour under changing climate as well as advantages for analysing, comparing and presenting results from multi-model ensemble simulations. Though individual model behaviour occasionally departed markedly from the average, ensemble median responses across sites and crop varieties indicated that yields decline with higher temperatures and decreased precipitation and increase with higher precipitation. Across the uncertainty ranges defined for the IRSs, yields were more sensitive to temperature than precipitation changes at the Finnish site while sensitivities were mixed at the German and Spanish sites. Precipitation effects diminished under higher temperature changes. While the bivariate and multi-model characteristics of the analysis impose some limits to interpretation, the IRS approach nonetheless provides additional insights into sensitivities to inter-model and inter-annual variability. Taken together, these sensitivities may help to pinpoint processes such as heat stress, vernalisation or drought effects requiring refinement in future model development.

**KEY WORDS:** Climate · Crop model · Impact response surface · IRS · Sensitivity analysis · Wheat · Yield

Resale or republication not permitted without written consent of the publisher

## 1. INTRODUCTION

The most widely used tools available for estimating future crop yield responses to climate change are process-based crop models, and these are already

applied extensively in different parts of the world (Rosenzweig et al. 2013, Challinor et al. 2014b). Models have been constructed using a variety of approaches for structuring and parameterising the basic processes influencing crop development and growth

\*Corresponding author: nina.pirttioja@ymparisto.fi

(Asseng et al. 2013). Such differences between model approaches introduce important sources of uncertainty into crop yield estimates, even for present-day conditions (White et al. 2011). Model results are also closely related to the quality of field observations and experiments available for model calibration and testing (e.g. Craufurd et al. 2013) as well as the calibration methods themselves (Angulo et al. 2013). Moreover, there are numerous model shortcomings in representing physiological processes that are still poorly understood (Rötter et al. 2011). These include crop responses to stresses such as high temperature, drought, waterlogging, nutrient limitation, pests and diseases, stomatal control on plant photosynthesis and transpiration under changing atmospheric carbon dioxide concentrations, and impacts of surface ozone (Porter et al. 2014).

Following early model comparison exercises (e.g. Porter 1993, Jamieson et al. 1998), and successive attempts by the Intergovernmental Panel on Climate Change (IPCC) to assess climate change impacts on crops across diverse models, regions, scenarios and assumptions (Gitay et al. 2001, Easterling et al. 2007, Porter et al. 2014), there has been a resurgence of interest in and establishment of dedicated research programmes on model inter-comparison and improvement (e.g. Palosuo et al. 2011, Rosenzweig et al. 2013, Challinor et al. 2014a). There have also been calls for more systematic modelling exercises for evaluating the suitability of crop model application under projected climates, particularly climatic conditions that lie beyond the 'quantifiable uncertainty' range of conventional local application (e.g. Rötter 2014). Of the few systematic sensitivity studies that have been conducted, most focus on changes in average (mean or median) yields (e.g. Børgesen & Olesen 2011, Asseng et al. 2013, Rosenzweig et al. 2014), whereas an additional key consideration for farmers is the reliability of the harvest from year to year, which is addressed in only a few studies (Luo et al. 2007, Challinor et al. 2010, Iizumi et al. 2011, Asseng et al. 2014).

This paper presents an examination of the modelled sensitivity of winter and spring wheat *Triticum aestivum* yield response to systematic changes in temperature and precipitation, the 2 variables most commonly analysed in climate impact assessments. It applies a common impact response surface (IRS) approach for a large ensemble of process-based wheat models run across a 3-site European transect. IRSs depict the response of an impact variable to changes in 2 explanatory variables as a plotted surface. Other aspects that would need to be taken into account if making projections of yield impacts under

climate change, such as the effect of increasing atmospheric CO<sub>2</sub> concentration, are deliberately not included in the analysis. IRSs have been applied in previous sensitivity studies of climate change impacts on permafrost (Fronzek et al. 2011), hydrology (Hanasaki et al. 2007, Wetterhall et al. 2011) and crop yield (Luo et al. 2007, Børgesen & Olesen 2011, Ruane et al. 2014). The approach was found to be effective for the visualisation and rapid evaluation of model sensitivities across a wide range of combinations of climate changes that are not specific to individual projections from climate models. Further, in combination with projections of climate variables represented probabilistically, the approach can be applied to estimate the likelihood of exceeding a given impact threshold, such as a critical level of yield (Børgesen & Olesen 2011).

Previous studies were based on individual models or emulators, whereas this is the first to apply IRS methods across a large ensemble of models. The main objective of the study was to demonstrate the capabilities and limitations of the IRS approach for investigating model ensemble crop yield responses under a large range of foreseeable changes in climate. The study's specific aims, for the 2 crop varieties at each site, were (1) to explore the sensitivities of modelled yields to changes in temperature and precipitation, (2) to quantify differences in average yield responses to altered climate across models, and (3) to examine multi-model responses of inter-annual yield variability and reliability under altered climate.

## 2. MATERIALS AND METHODS

### 2.1. Crop models

An ensemble of 24 process-based wheat models run by 26 modelling groups was applied in this study (Table 1). The same model/version was calibrated separately and run by 2 modelling groups in 2 cases (CERES-wheat DSSAT ver. 4.5 and WOFOST ver. 7.1). Models are referred to in the text using a numerical identifier, with the aim of pursuing impartiality in reporting model results. Table S1 in Supplement 1 at [www.int-res.com/articles/suppl/c065p087\\_supp.pdf](http://www.int-res.com/articles/suppl/c065p087_supp.pdf) provides an overview of all models, including key references and a characterisation of how they describe selected processes and treat environmental constraints.

All models work on a daily time step, though there are some differences in their requirements for climate input variables (Table S1 in Supplement 1).

Table 1. List of wheat models applied in the present study. ID is a model identification number used in the text. Duplicate entries indicate models for which several groups provided results

ID	Model	Contact person(s)	Web documentation
1	AFRCWHEAT2	Montesino/Porter	–
2	APSIM-Nwheat 1.55	Cammarano/Asseng	www.apsim.info
3	APSIM-Wheat (modified) 7.5	Wang	www.apsim.info
4	AquaCrop 4.0	Lorite	www.fao.org/nr/water/aquacrop.html
5	ARMOSA13.04	Perego/Sanna/Acutis	–
6	CARAIB Crop	Minet/Jacquemin/François	www.umccb.ulg.ac.be/Sci/m_car_e.html
7	CERES-wheat DSSAT v.4.5	Ruiz-Ramos	http://dssat.net
8	CERES-wheat DSSAT v.4.5	Deligios	http://dssat.net
9	CERES-wheat DSSAT v.4.6	Trnka/Hlavinka	http://dssat.net
10	CropSyst 3.02	Moriondo/Ferrise/Bindi	modeling.bsyse.wsu.edu/CS_Suite_4/CropSyst
11	DNDC 9.5	Baranowski/Slawinski	www.dndc.sr.unh.edu
12	Fasset 2.5	Öztürk/Olesen	www.fasset.dk
13	HERMES V 4.26	Kersebaum/Kollas	www.zalf.de/en/forschung/institute/lisa/ forschung/ http://oekomod/hermes
14	LINTUL-4 v6	Supit/Wolf	http://models.pps.wur.nl/models
15	LPJ-GUESS	Bodin/Olin	www.nateko.lu.se/lpj-guess
16	LPJmL	Müller	www.pik-potsdam.de/research/projects/lpjml
17	MCWLA-Wheat 2.0	Tao	–
18	MONICA 1.2.5	Nendel	http://monica.agrosystem-models.com/en
19	SALUS	Basso	http://salusmodel.psm.msu.edu/
20	SIMPLACE<Lintul2, Slim>	Hoffmann/Gaiser/Ewert	www.simplace.net/, http://models.pps.wur.nl/models
21	Sirius	Semenov/Stratonovitch	www.rothamsted.ac.uk/mas-models/sirius.php
22	SiriusQuality 2.0	Ferrise/Bindi	www1.clermont.inra.fr/siriusquality
23	SPACSYS 5.0	Wu	–
24	STICS V6.9	Dumont/Ruget/Buis	www6.paca.inra.fr/stics
25	WOFOST 7.1	Palosuo/Rötter	www.wofost.wur.nl
26	WOFOST 7.1	Krzyszczak/Slawinski	www.wofost.wur.nl

Most models were developed for the field scale, while a few (i.e. CARAIB, LPJmL, LPJ-GUESS and MCWLA) have been developed for regional assessment. Models differ considerably in the way they treat factors that define, limit or reduce growth (van Ittersum & Rabbinge 1997, van Ittersum et al. 2003, Wu & Kersebaum 2008), which is reflected in contrasting structure, model parameters and associated input data requirements.

## 2.2. Study sites and key data requirements

### 2.2.1. Criteria for the selection of study sites

The models were applied using data from 4 study sites representing contrasting environmental zones (Boreal, Atlantic Central, Continental and Mediterranean South) for wheat production in Europe (our Fig. 1; Metzger et al. 2005). The study sites were selected according to multiple criteria. A primary objective was to offer a representative north-south cross-section (transect) of current agro-climatic conditions for wheat cultivation areas in Europe. Sec-

only, the sites should capture the contrast in rain-fed yield potential between favourable conditions in Central Europe and growth conditions at the northern margins (limited by temperature) and southern margins (limited by precipitation). Thirdly, options for adapting to anticipated future climate should include the possibility to substitute winter wheat types with spring wheat types and vice versa. Finally, crop data should be available for site-specific calibration of the crop models, while daily weather data for running the models were required for the baseline period of 1981 to 2010 at each site.

To keep the number of simulations manageable, one site was chosen in each of northern, central and southern Europe for each crop variety. Jokioinen in Finland was chosen for northern Europe and Lleida in Spain for southern Europe. Since sufficient calibration data for modern cultivars that would cover both winter wheat and spring wheat at a single site in central Europe were not available, 2 sites in Germany were chosen: Dikopshof, located in the west, for winter wheat and Nossen in the east, for spring wheat. The principal characteristics of the 4 sites and their agro-climatic conditions are summarised in Fig. 2.

2.2.2. Climate input data

Observed weather data for each of the 4 sites during the period 1980–2010 (including years preceding harvest) were obtained at a daily time step for the following variables: global solar radiation, minimum and maximum temperature, precipitation, wind speed and various (model-specific) measures of humidity. Supplement 2 ([www.int-res.com/articles/suppl/c065p087\\_supp.pdf](http://www.int-res.com/articles/suppl/c065p087_supp.pdf)) describes the sources of data (Table S2) as well as procedures for deriving missing values.

The sensitivity of a crop model to changes in climate was tested by systematically modifying temperature and precipitation values of baseline weather data at the chosen sites using a simple ‘change factor’ approach (e.g. Diaz-Nieto & Wilby 2005). A constant change was applied to all days of each year of the baseline time period. Observed daily minimum and maximum temperatures were modified by between

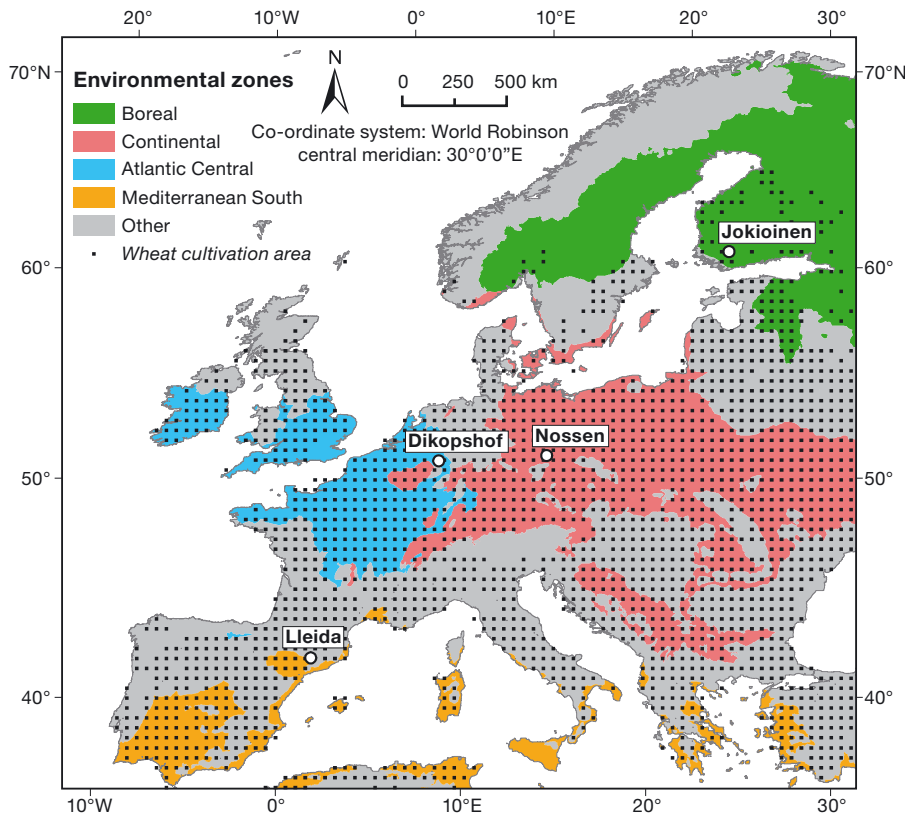


Fig. 1. Locations of weather stations used in this study superimposed on environmental zones as defined by Metzger et al. (2005). Black squares are wheat cultivation areas from Monfreda et al. (2008), re-sampled to a 0.5 degree grid

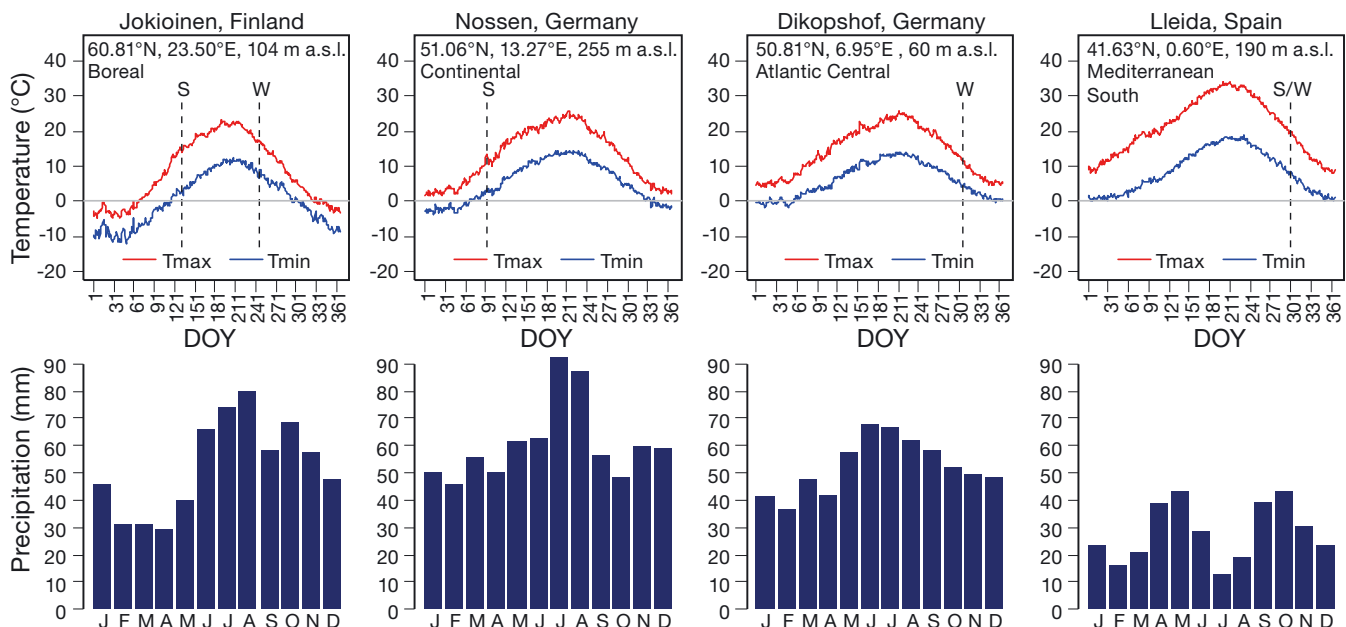


Fig. 2. Long-term (1981–2010) mean daily minimum (Tmin) and maximum (Tmax) temperature (top row) and monthly total precipitation (bottom row) for the 4 weather stations used in this study. Vertical dashed lines on the temperature plots indicate the mean sowing DOY (day of year) used for spring (S) and winter (W) wheat. Latitude, longitude and elevation (a.s.l.: above sea level) are reported for each weather station along with the environmental zone in which the station is located (cf. Fig. 1)

-2°C and + 9°C at 1°C intervals and concurrently (preserving the baseline diurnal temperature range). Daily precipitation was adjusted between -50% and +50% at 10% intervals. The ranges were selected to be large enough to encompass changes in regional climate change at the 4 sites by the mid-21st century represented in probabilistic projections (for medium emissions; Harris et al. 2010) as well as in projections from an ensemble of general circulation models that participated in the Coupled Model Intercomparison Project phase 5 (IPCC 2013a). The increments of the changes were chosen to be small enough to represent possible non-linearities in model responses to a changing climate, while at the same time ensuring a manageable number of combinations. Thus, each year of the baseline was modified according to 132 different combinations of changed temperature and precipitation and provided as input to the models.

For simplicity, all other variables were unchanged at their baseline values. Note that this procedure introduced a discrepancy between models in the way humidity was treated. This is because in order to maintain a constant level of relative humidity (*RH*) as temperature rises, vapour pressure (*e*) must also increase. Thus, by fixing *e* and *RH* at baseline values, the 7 models that required *e* as an input would have over-estimated evapotranspiration under higher temperatures compared to those models requiring *RH*. The effect of this discrepancy was evaluated using one of the models and then extrapolated to all 8 sets of runs affected. While yield declines were indeed greater under conditions of extreme warming and drying, the overall effect on the ensemble results reported below was minimal and the conclusions are unaffected.

### 2.2.3. Sowing date, soil and calibration data

For the Finnish site and the 2 sites in Germany, a specific sowing date was determined for each year of the (30 yr) baseline period based on observations. In the absence of observed sowing dates for Spain, 1 fixed sowing date (day of the year [DOY] 302), identified on the basis of local expertise, was used for all years and for both spring wheat and winter wheat. Following local practice, it was assumed that farmers at the Spanish site (Lleida) can use the same sowing date for both crop varieties and would sow as early as possible after the autumn rain, approximately at the end of October.

Additional data were made available for model calibration. To avoid over-fitting, where possible, data

sets were used that had not previously been applied in model calibration and/or were not fully documented in published papers. Model users were provided with phenological observations and yields from each location and for both crop varieties. Total above-ground biomass and average grain weights were also provided for the Spanish site (Cartelle et al. 2006, Abeledo et al. 2008). As an alternative to the generalised soil type used for the model simulations (clay loam), modellers had the option to use information on the actual soil of the site for calibration. In addition to the sowing dates provided for the simulations, calibration data on management included sowing depths as well as data of varying detail on fertilisation, irrigation, tillage and residue management. Overall, calibration data on plant development included observations from between 5 and 29 seasons or treatments, depending on site and wheat variety (see Table S2 in Supplement 2).

### 2.3. Modelling protocol

A sensitivity analysis was performed by running the models for the 30 baseline years and 132 perturbations to the baseline weather for spring and winter wheat at all study sites (3 sites per crop) resulting in 23 760 simulated seasons per model. In the simulation set-up several common rules and limitations were specified. This included applying baseline sowing dates for all temperature and precipitation perturbations and assuming a generalised soil type at all sites. Atmospheric CO<sub>2</sub> concentration was kept constant at 360 ppm, representing levels observed around 1995 (IPCC 2013b, p. 1401), at the mid-point of the 1981 to 2010 baseline period. The simulations were performed on a daily time step for water-limited yields assuming optimal nutrients. Simulations were conducted as a succession of independent growing seasons with the moisture content of the top soil set at 75% of field capacity at the beginning of each season.

Model outputs were stored on a yearly resolution. For each simulation the flowering date (Zadoks stage 65; Zadoks et al. 1974), date of physiological maturity (Zadoks stage 91) and harvest date (model-specific) were stored as the day of the year (DOY). Harvested yields were represented as grain dry matter (DM), while total above ground DM production and nitrogen content of yield were also reported for the end of each simulation. Water use was stored as cumulative actual evapotranspiration (mm) from sowing to maturity.

For models that do not specify a latest date for harvest, simulations in some seasons may continue

through to a maturity date in the year subsequent to the harvest year. This unrealistic outcome was avoided by imposing a fixed maturity cut-off date, based on expert judgement, in the autumn of the harvest year (DOY 258 for Finland and Spain and DOY 274 for both German sites). If a simulation reported a maturity date that exceeded the harvest cut-off, the DM grain yield and grain nitrogen content were set to 0 ( $\text{kg ha}^{-1}$ ) and all other output variables were assigned missing values. It was assumed that before a crop reached maturity it was not fit for harvest, so that no DM yield was recorded. All modelling groups were given the chance to review initial IRS plots of the results and to iterate by refining the calibration of a model or correcting any errors, often themselves readily detectable from the IRS plots.

#### 2.4. Construction and analysis of impact response surfaces

IRSs were constructed by interpolating the results of the sensitivity analysis of each model as contour lines with respect to changes in annual temperature along the x-axis and precipitation along the y-axis. The contour lines were plotted using the contour (Becker et al. 1988) and filled.contour (Cleveland 1993) functions in the statistical software package R ver. 3.0.2 (R Core Team 2013). Individual IRSs were created for each model and combination of parameters (i.e. site, crop variety, harvest year).

In this paper we concentrated on analysing results of DM grain yields and their changes relative to the baseline for the model ensemble. For summarising average yield responses and their dispersion, we used 2 types of measures. Across different models in the ensemble we used medians for average responses, in line with earlier multi-model ensemble studies (e.g. Asseng et al. 2013, and see our Section 4.1), and the inter-quartile range (IQR; from the 25th to the 75th percentile) for the spread of responses. Over periods of time (in most cases, 30 yr) we used means, though some results from individual years are also shown. Inter-annual variability was treated in 2 ways. The year-to-year 'reliability' of yields was represented, for any individual model, as the percentage of years when DM grain yield ( $\text{kg ha}^{-1}$ ) exceeds a given threshold level. Here, reliability is defined relative to the 10th percentile of yields during the baseline period (i.e. the level of yield that was exceeded in 9 yr out of 10). The threshold is defined only across those years that have a non-zero yield for the baseline period. An alternative to focusing on reliability at

the low end of yield responses is to plot the coefficient of variation (CV) across the 30 yr, which accounts for the full distribution of yield responses. Both measures of inter-annual variability were estimated for each 30 yr simulation period under all combinations of temperature and precipitation change relative to the baseline.

### 3. RESULTS

#### 3.1. Baseline yields and standardisation of yield responses

##### 3.1.1. Baseline yields

Examination of the general magnitude of modelled yields for the baseline period of 1981 to 2010 revealed large differences both between individual models as well as between simulated yields and observed aggregate regional yields, where these were available (see Supplement 3 at [www.int-res.com/articles/suppl/c065p087\\_supp.pdf](http://www.int-res.com/articles/suppl/c065p087_supp.pdf)). This is particularly true for Lleida, where the spread across models in simulated yields of both spring and winter wheat was large, and the model ensemble median yield was considerably higher than that observed. Fig. 3a depicts results for spring wheat at Lleida (for other locations and crop varieties see Fig. S1 in Supplement 3). The ensemble median of simulated yields exceeded those observed in most years and for both varieties at Jokioinen (Fig. S1a,b), whereas observed vs. modelled yields were much closer at the 2 German sites (Fig. S1c,d).

The levels of observed and modelled baseline (1981–2010) mean yields are summarised in Table 2. Yields varied widely between models, ranging from  $1260 \text{ kg ha}^{-1}$  (Model 14 for spring wheat in Spain) to  $10484 \text{ kg ha}^{-1}$  (Model 19 for winter wheat in Germany), while the range of observed regional yields was  $2556$  to  $7529 \text{ kg ha}^{-1}$ . The highest model estimates are up to 8 times greater than the lowest estimates for the same site and crop variety, with the highest yields found at the German sites in both sets of observed statistics as well as for most models and for the ensemble medians (Table 2). The inter-annual variability of observed yields (adjusted to account for long-term trends assumed to be unrelated to weather), as represented by the CV (not shown), was lowest in the 2 German regions, higher in Finland and substantially higher in northern Spain. These observed regional differences were consistent with modelled between-site estimates of inter-annual vari-

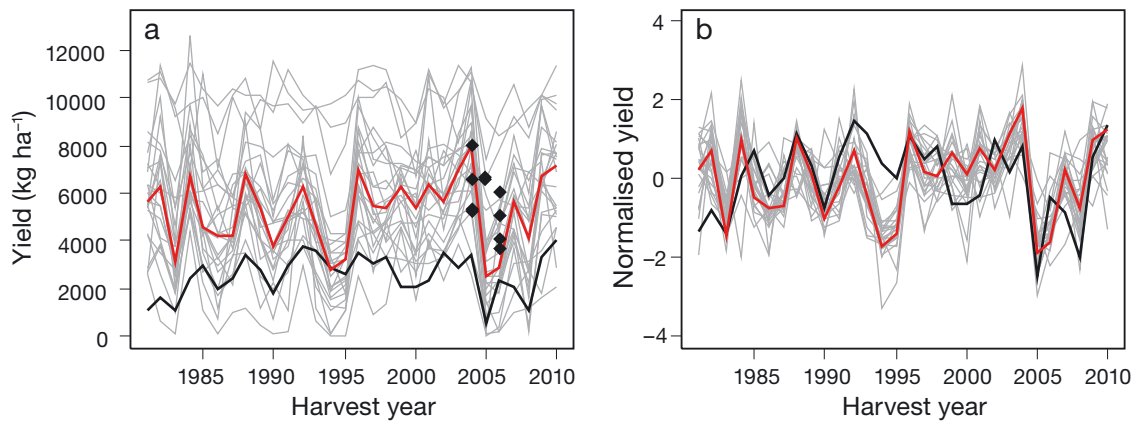


Fig. 3. Dry matter grain yields of spring wheat for the 1981–2010 baseline period at Lleida, Spain: (a) absolute and (b) normalised (standard deviations relative to the 1981–2010 mean). Grey lines: yields simulated by individual models. Red lines: multi-model ensemble median yields. Black lines: observed yields for northern Spain from the Spanish Ministry of Agriculture (de-trended in b). Black dots: site data from Lleida provided for model calibration. Multiple points for a single year are for different fertiliser and irrigation treatments

Table 2. Colour-coded classes of baseline (1981–2010) mean yields (kg ha<sup>-1</sup>). Rows: sites (Si)—Finland (FI), Germany (DE) and Spain (ES) (see Figs. 1 & 2), and crop varieties (C)—spring (S) and winter (W) wheat. Columns: model identification numbers (cf. Table 1), model ensemble medians (E50) and observations (O, where values inside the cells indicate the number of years for which observed crop yield data were available; see Supplement 3). Models for which no results for a specific site or crop variety were provided are marked 'N/A'

Si	C	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	E50	O	
FI	S												N/A											N/A					30	
FI	W																													30
DE	S												N/A																	20
DE	W																													29
ES	S												N/A																	30
ES	W																													30

kg ha <sup>-1</sup>	<2000	2000–4000	4000–6000	6000–8000	8000–10000	≥10000
---------------------	-------	-----------	-----------	-----------	------------	--------

ability shown by values of CV at the intersect of the grey lines in Fig. 7 and Fig. S3 in Supplement 3.

In order to gain an impression of how simulated and observed yields respond to annual weather, yield time series can be expressed as normalised anomalies relative to the long-term mean, with observed yields de-trended as described above. Fig. 3b illustrates this for spring wheat at Lleida, which in this case indicates that simulated yield anomalies correspond fairly well on visual inspection to those observed. No statistical measures of correspondence are provided, given that the observed yields represent much larger regions than those simulated at specific sites.

### 3.1.2. Standardising results for IRS analysis

As the first step in creating IRSs of period-mean yield responses, annual IRSs were constructed for each model, crop variety and site. An example of absolute DM grain yield responses for winter wheat at Dikopshof for each of the baseline years (1981–2010) from an arbitrarily selected model is presented in Fig. S2 in Supplement 3. The 30 yr mean IRS, constructed by averaging over the individual year responses, is shown in the larger panel at the bottom right of the figure. This illustrates how averaging can smooth out complex behaviour observed during some individual years. Only a few years (e.g. 1993

and 2007) have a response closely resembling the 30 yr mean IRS.

Given the large inter-model differences in yield levels between models (cf. Table 2), and to allow for more meaningful comparison between IRS plots of the period-mean yield responses and the inter-annual variability of yields, the 30 yr mean yields are expressed as percentage changes relative to the baseline. Fig. 4 shows an example of the relative yield responses of individual models for winter wheat at

Dikopshof. At this site, and in examples for the other sites and crop variety, there are similarities between models in the average response to climate change, with most models indicating decreases in yield for increased temperature and decreased precipitation. However, the relative strength and pattern of these relationships displays large variation between models, with a few models even depicting an opposite sign of response across the climate changes represented on the IRS.

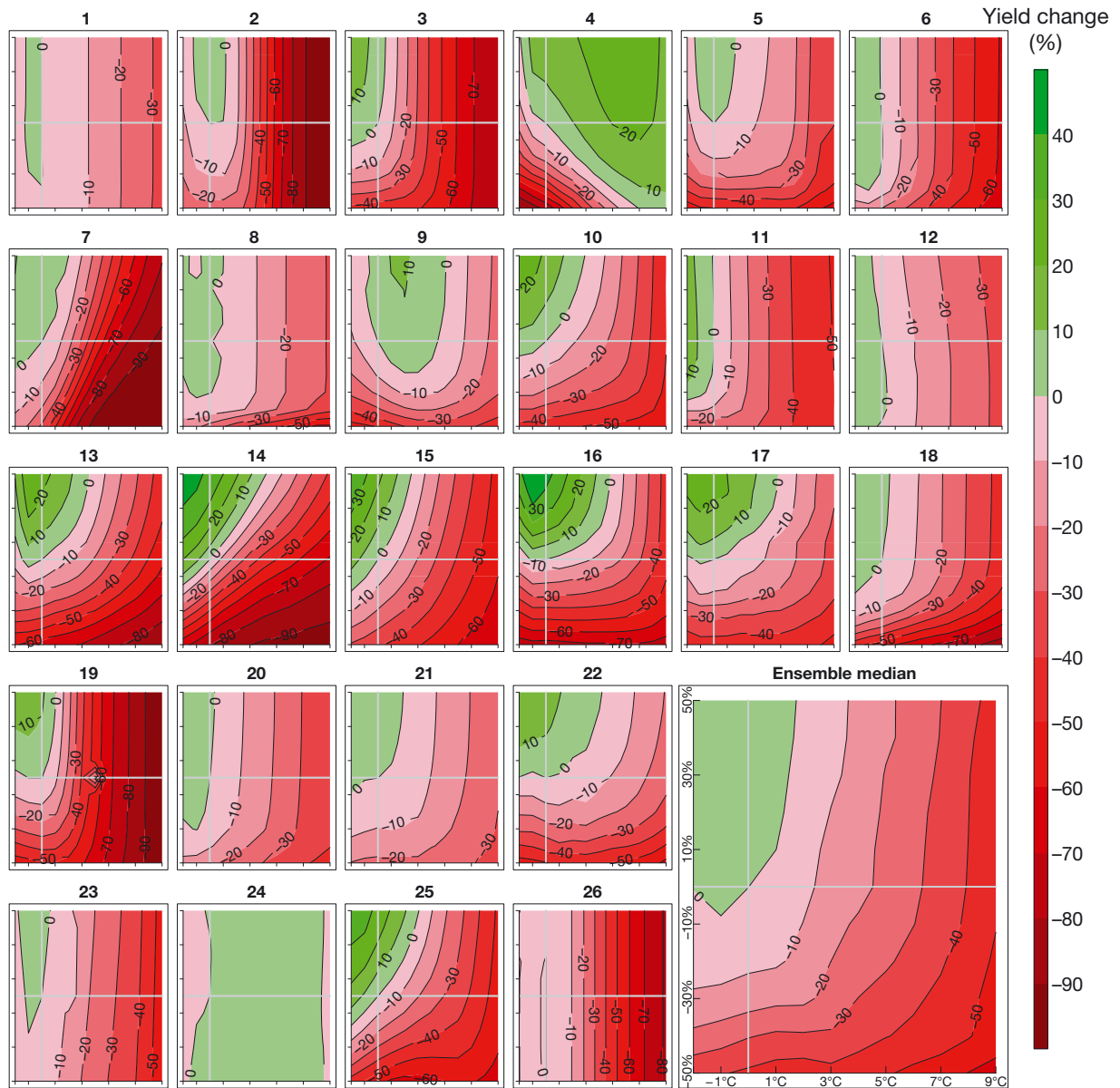


Fig. 4. Thirty yr mean changes in winter wheat dry matter grain yields simulated by 26 crop models and the ensemble median (bottom right) for changes in temperature (x-axis) and precipitation (y-axis) relative to the baseline climate (1981–2010) at Dikopshof, Germany. Axes of the smaller plots cover the same ranges as those labelled on the median plot (bottom right). By definition, the yield change is 0% for the baseline climate at the intersection of the grey lines. The number above each small plot is the model identification number (cf. Table 1)



### 3.2. IRS analysis: multi-model 30 yr mean yield responses

#### 3.2.1. Ensemble median changes

In order to summarise yield response behaviour across models, plots of the model ensemble median of the percentage change in 30 yr mean yields relative to the baseline were constructed for each site and crop variety. The pattern of median yield response at any one site appears to be very similar

between spring and winter wheat (left-hand panels in Figs. 5 & 6), though with a slightly steeper yield decline for spring than for winter wheat with respect to changes in temperature at all sites. At Jokioinen, the maximum yield for winter wheat occurred at temperatures around the baseline while for spring wheat it occurred slightly above, indicating that for this particular cultivar the optimum would be achieved with a slight warming. Cooling at Jokioinen caused a sharp decline in yield of both varieties. Conversely, maximum yields were ob-

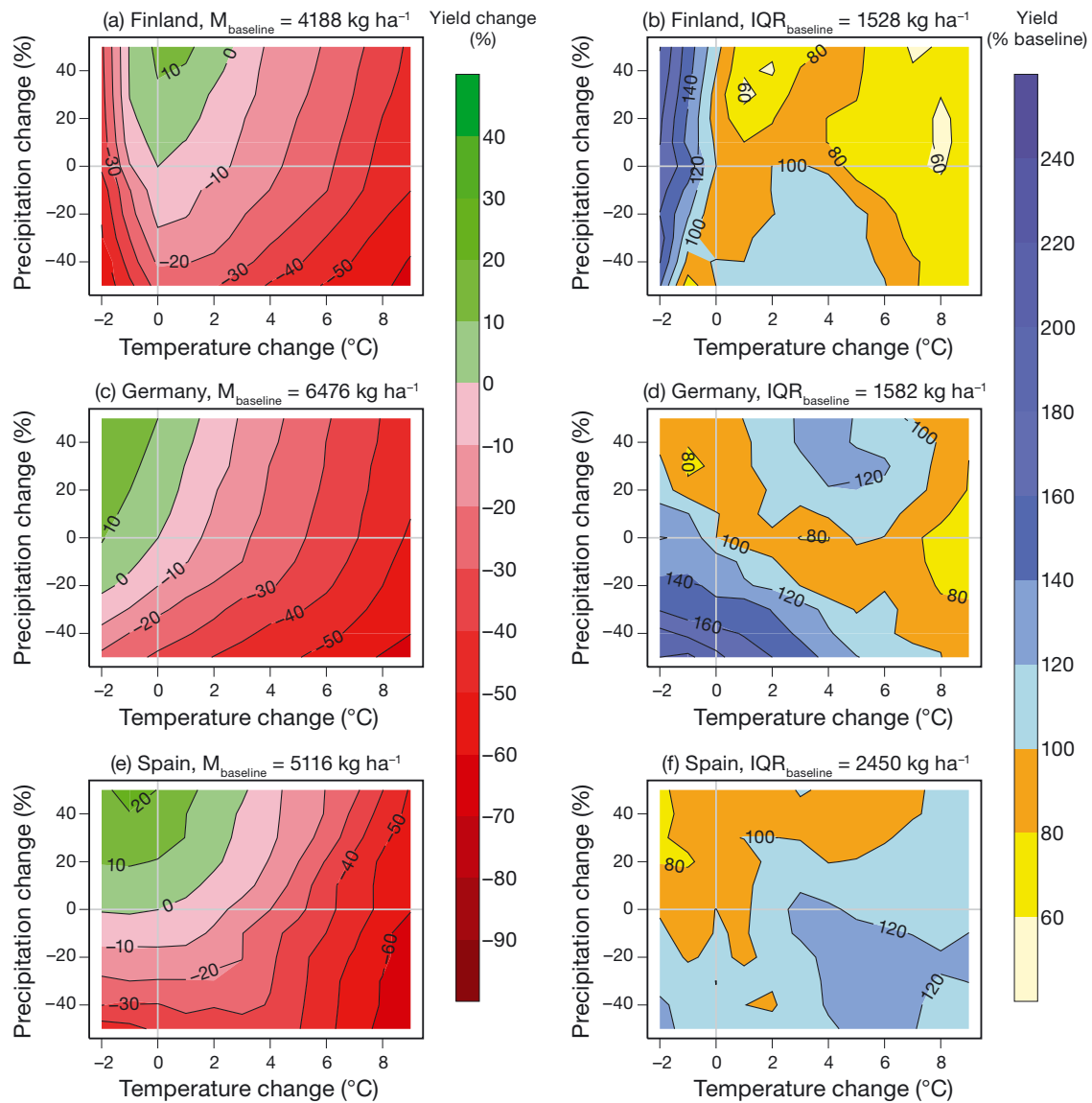


Fig 5. Spring wheat. (a,c,e) Ensemble median response (%) to changes in temperature (x-axis) and precipitation (y-axis) of period-mean dry matter grain yields relative to the baseline (1981–2010) climate and (b,d,f) ensemble inter-quartile range (%) of the relative responses across 24 crop models at (a,b) Jokioinen, Finland and 25 models at (c,d) Nossen, Germany and (e,f) Lleida, Spain. The ensemble median ( $M_{\text{baseline}}$ ) and ensemble inter-quartile range ( $IQR_{\text{baseline}}$ ) of absolute yields for the baseline are listed above each plot

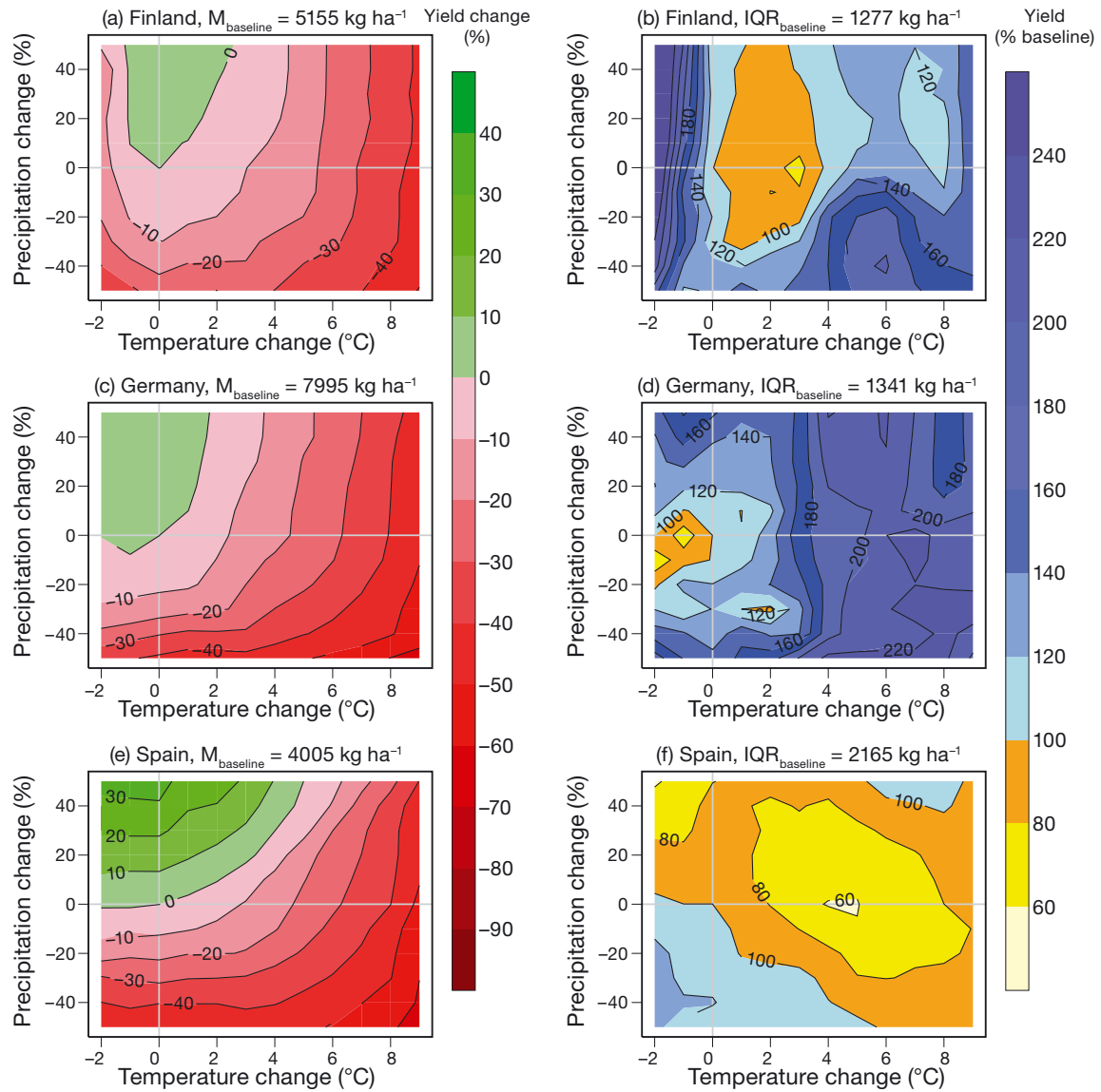


Fig. 6. As Fig. 5, but for winter wheat and for an ensemble of 26 crop models at (a,b) Jokioinen, Finland, (c,d) Dikopshof, Germany and (e,f) Lleida, Spain

tained with cooling relative to the baseline at the German sites and at Lleida.

For combinations of temperature and precipitation change across the uncertainty ranges defined for the IRS, temperature was the dominant constraint on median yield at all sites for warming above about 5°C. Under conditions of less warming, precipitation had an increasing influence across the transect from north to south, with median yields most sensitive at Lleida. Contour lines were nearly vertical in areas of the climate uncertainty space with increases in precipitation at Jokioinen and at the German sites, in particular for winter wheat, indicating that precipitation change had hardly any effect on yield in these regions of the IRS.

### 3.2.2. Ensemble inter-model variability

In contrast to the similarities of median yield response to temperature and precipitation change between crop varieties at each site and broadly across sites, there are marked differences in the inter-model variability of responses. Plots of the ensemble IQR depict the spread in the 30 yr mean response to changes in temperature and precipitation (right-hand panels, Figs. 5 & 6). The absolute IQR for the baseline (shown above each figure) is scaled to 100% at the origin of the plot for all models, and values of IQR across the IRS are expressed as percentages of the baseline. For example, values for spring wheat at Lleida are percentages of the baseline IQR of 2450 kg

ha<sup>-1</sup> (Fig. 5f). Note that variability across a wider distribution of models (10th to 90th and 5th to 95th percentiles) was also examined, with patterns found to be broadly similar to the IQR patterns reported here.

The results showing the highest consistency between sites or crop varieties are the increased IQR values for Jokioinen under cooling for both spring and winter wheat (Figs. 5b & 6b, respectively). The inter-model spread for spring wheat was lower than the baseline under most cases of warming, except for a slight increase for reduced precipitation combined with moderate warming (Fig. 5b). For winter wheat under warming, there is some similarity between IQR patterns for Jokioinen and Dikopshof, which increased under moderate to high levels of warming and reduced precipitation (Fig. 6b,d), though model spread was lowest under slightly warmer conditions at Jokioinen but cooler conditions at Dikopshof.

The inter-model variability of spring wheat yields at Nossen was lowest for combinations of temperature and precipitation change between the wettest, coolest conditions and the warmest conditions, with less agreement between modelled yields (higher IQR) for moderate warming and increased precipitation or for drying combined with cooling or moderate warming (Fig. 5d).

The IQR broadly followed the median yield response for spring wheat at Lleida, with highest yields coinciding with the lowest inter-model variability and increasing spread as yields declined with higher temperature and reduced precipitation (compare Fig. 5e and f). In contrast, for winter wheat at Lleida, the IQR was lower than the baseline for almost all cases of precipitation increase, and most cases of warming, with increased model disagreement found only for reduced precipitation combined with slight warming or cooling (Fig. 6f).

### 3.3. IRS analysis: inter-annual yield variability

Two measures were used to describe the inter-annual variability in yield response under changing climate across the IRS: yield reliability and the coefficient of variation (CV). Both measures were computed for each model and summarised as multi-model ensemble medians (Fig. 7).

Patterns of median crop yield reliability were similar at all locations, but with shifted response along both axes and with the rates of decline in reliability differing between locations. Results for spring wheat are shown in Fig. 7 (left panels); winter wheat results are given in Fig. S3 in Supplement 3. Crop yield reli-

ability declined with increasing temperature and decreasing precipitation. This effect is most notable at the German sites, for both spring and winter wheat, where the reliability declined very quickly with warming and drying. Highest reliability was achieved with a slight decrease in temperature at the German and Spanish sites, and conversely, with a slight increase at the Finnish site. By definition, the reliability should be 90% at the baseline. For spring wheat at Jokioinen, the yield threshold of some models was affected by simulated crop failure under the baseline climate occurring in >10% of the years (see threshold definition in Section 2.4), resulting in an ensemble median baseline reliability of less than 90%. Patterns of reliability appear to track those of the changes in median yields relative to the baseline (Figs. 5 & 6, left panels). Increases in median yields coincided with higher reliability, whilst reliability declined with lower median yields.

In contrast to the patterns of reliability, no clear general association between patterns of CV and median yields could be identified. The ensemble medians of the CV are depicted for spring wheat in the right panels of Fig. 7 (for winter wheat, see Fig. S3 in Supplement 3). For instance, at Jokioinen the median CV was smallest for temperature increases of 3 to 4°C in combination with increased precipitation (Fig. 7b and Fig. S3b), while the median yield already started to decrease at this level of temperature increase. The median CV increased strongly with cooling across the whole range of precipitation changes. While the median yields were strongly temperature-dependent at the German sites, the CV of annual yields was influenced mostly by precipitation, with drying leading to an increase in variability. Across the defined uncertainty ranges for spring wheat at Lleida, temperature was the dominant driving variable with respect to the inter-annual variability while the median yield was very much influenced by precipitation. However, for winter wheat the variability was affected almost exclusively by precipitation, with drying leading to more variable yields (Fig. S3f).

## 4. DISCUSSION

This study is an exploratory application of the IRS approach. In this section we attempt to summarise and interpret the results, emphasising insights gained by using the approach, outlining its shortcomings and discussing its wider applicability and utility in future studies.

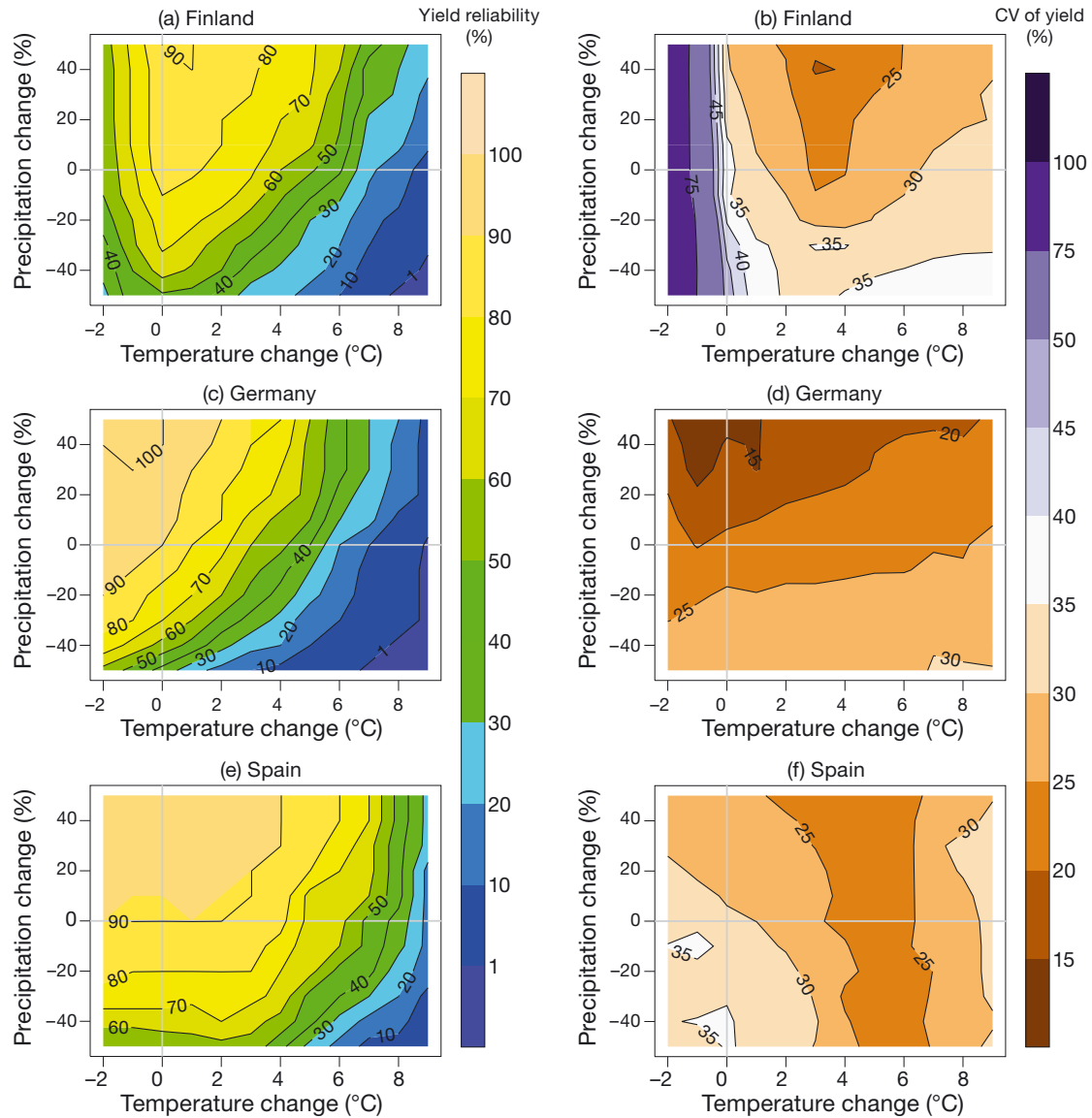


Fig. 7. Ensemble medians of yield reliability, defined as the percentage of years when DM grain yield ( $\text{kg ha}^{-1}$ ) is above the 10th percentile of the baseline yield (left-hand panels), and of coefficients of variation (CV) of annual yields (right-hand panels), for spring wheat under changes in temperature and precipitation relative to the 1981–2010 baseline for 24 crop models at (a,b) Jokioinen, Finland, and 25 models at (c,d) Nossen, Germany and (e,f) Lleida, Spain

#### 4.1. Ensemble wheat model responses

It has been argued that the use of a model ensemble increases the robustness of simulated yield estimates compared to using individual models, as the ratio of signal (average change) to noise (variation) increases with the number of models, while errors in individual models tend to cancel each other out (Asseng et al. 2013). The IRS plots offer a consensus view of how models simulate the joint effects of temperature and precipitation changes on wheat yields. Since models differ in their representation of key

processes, it can be difficult to disentangle the relative influence of these 2 climatic variables on yield response across such a large and diverse set of models. Here, the results are interpreted by treating temperature and precipitation effects in turn.

##### 4.1.1. Temperature-related effects

*Average yields and development.* Temperatures outside those typically experienced can cause significant reductions in yields through various processes

at different development stages (Asseng et al. 2014). Wheat is best suited to a cool climate with an optimum temperature range of 17 to 23°C over the entire growing season, although cultivar-specific differences exist (Porter & Gawith 1999). Fig. 8a summarises the model ensemble temperature responses for winter wheat at all 3 sites assuming baseline precipitation (combining information from IRSs in Fig. 6). The equivalent plot for spring wheat is given in Fig. S4a in Supplement 4 at [www.int-res.com/articles/suppl/c065p087\\_supp.pdf](http://www.int-res.com/articles/suppl/c065p087_supp.pdf).

Responses at all sites showed decreases in yields with warming. This is because present-day cultivars at each site have been bred and selected to develop and mature under ambient conditions. Warming accelerates plant phenology such that crops mature earlier, the grain filling period shortens and dry matter accumulation is reduced, resulting in lower yields. While baseline temperatures were close to optimal for local cultivars under Finnish conditions, yields benefited from cooling at the German and Spanish sites. This suggests that adoption of later-maturing cultivars with higher temperature requirements might already be advantageous at those sites, and increasingly so under future warming.

A decrease in yields with warming is consistent with empirical studies of observed wheat yield trends worldwide (Lobell & Field 2007). It is also in line with previous multi-model studies for constant CO<sub>2</sub> levels (Asseng et al. 2013, 2014). For example, the median 30 yr mean yield response of autumn-sown wheat to a temperature increase of 6°C relative to 1981–2010

at a site in the Netherlands was –26% (Asseng et al. 2013). This compares to –23%, –28% and –28% at Jokioinen, Dikopshof and Lleida, respectively (Fig. 8a). Spring wheat yields were more sensitive to temperature in our results (respective values are –28%, –34% and –37% for the same 6°C warming; see Fig. S4a in Supplement 4) and the largest reductions were also found in the Asseng et al. (2013) study for spring wheat yields in Argentina (–39%).

**Harvest cut-off.** Cooling had a strong effect on modelled yields at Jokioinen, due to a failure of crops to mature before the harvest cut-off date during cooler baseline years in many models. Crop failures simultaneously reduced the 30 yr mean yields while increasing inter-annual yield variability. The cut-off date was introduced to avoid model simulations from continuing unrealistically into a subsequent growing season. However, its precise timing is open to debate, as well as the assumption made to set the yield to zero during those years. In reality, under such conditions some farmers might still manage to salvage a crop.

**High temperature extremes.** The upper lethal temperature limit for wheat and its standard error is reported as  $47.5 \pm 0.5^\circ\text{C}$  (Porter & Gawith 1999). Estimates from the same source of maximum temperature above which growth ceases are considerably lower than this, varying with phenological phase ( $32.7 \pm 0.9^\circ\text{C}$  from sowing to emergence,  $31.0^\circ\text{C}$  around anthesis and  $35.4 \pm 2.0^\circ\text{C}$  during grain filling). Extreme heat events that exceed the maximum temperature limits are already observed occasionally at all sites. Moreover, with the IRS range of temperature sensi-

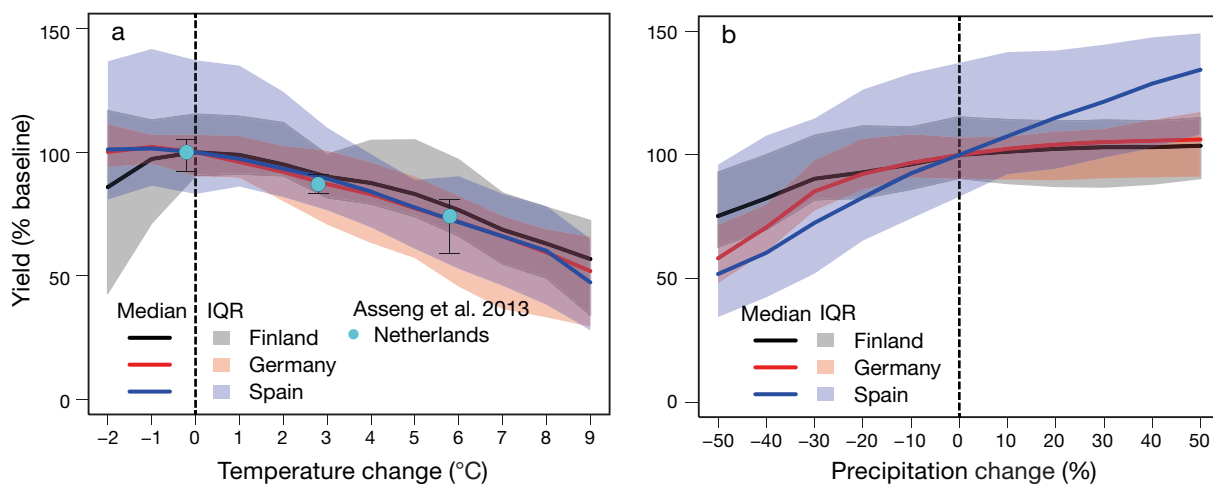


Fig. 8. Ensemble median response (solid lines) and inter-quartile range (IQR; coloured bands) of period-mean dry matter winter wheat yield (%) relative to the baseline (1981–2010) climate across 26 crop models at Jokioinen, Finland, Dikopshof, Germany and Lleida, Spain for changes in: (a) temperature with baseline precipitation, and (b) precipitation with baseline temperature. Baseline values are scaled to 100%. Coloured points and error bars are median and IQR model ensemble responses for a site in the Netherlands, from Asseng et al. (2013)

tivity extending to +9°C warming, severe restrictions on growth should be expected to show in the model results. However, not all models applied in this study account for specific heat stress impacts (Table S1 in Supplement 1), even though they may be used in climate change studies. Regardless, it is difficult to distinguish between heat-related impacts and other yield-reducing stresses from these ensemble results. At Lleida, there are some instances of daily maxima exceeding the lethal limit, though with a shortened growth period these occurred well after modelled harvest for present-day cultivars. However, temperatures of this magnitude would impose an absolute constraint on wheat viability during the summer months if longer-season cultivars were to be considered as an adaptation option.

*Vernalisation.* Increased temperature can also affect vernalisation, the chilling requirement of certain winter wheat varieties that is obligatory for flowering. The optimum temperature for vernalisation is given as  $4.9 \pm 1.1^\circ\text{C}$  with a maximum temperature of  $15.7 \pm 2.6^\circ\text{C}$ , after which vernalisation does not occur and no yield is produced (Porter & Gawith 1999). During some seasons under the high-end temperature changes, requirements for vernalisation were not met, which should result in crop failure. A few individual models showed this effect in some years for winter wheat at Lleida (not shown), though most models that simulate vernalisation (cf. Table S1 in Supplement 1) did not indicate complete crop failure.

*Inter-model variability.* Changes in inter-model variability across the IRS are determined in large measure by responses under baseline conditions. The baseline IQR at Lleida was substantially greater (by between 55 and 70%) for both crop varieties than at the other 2 sites, while median yields at Lleida were relatively low (e.g. compare the width of the coloured bands in Fig. 8a). With modelled median yields declining under warming, there is little scope for the IQR to increase significantly at Lleida, in contrast to the German and Finnish sites, where baseline IQR was lower. The spread of spring wheat yields narrowed with warming and unchanged precipitation at the Finnish and German sites (Fig. S4a in Supplement 4), while modelled winter wheat yields diverged with greater warming (Fig. 8a). This suggests differing model treatment of temperature effects on overwintering crops (e.g. through vernalisation). Divergent yield responses to cooling at Jokioinen are attributable to harvest failure occurrence in some models.

*Inter-annual variability and yield reliability.* IRS patterns of modelled yield reliability under changing climate were strongly governed by the baseline yield

response at different sites, which was used to define the threshold yield (lowest 10th percentile). Climatic conditions were most favourable, median yields highest, and inter-annual variability (based on the CV) lowest under the baseline climate at the 2 German sites. While cooling improves reliability, increments of warming reduced modelled yields below the yield threshold and reliability fell more rapidly at the German sites than at the Finnish and Spanish sites, even though absolute yield levels were still higher and the CV generally lower. Cooling reduced yield reliability for both crop varieties at Jokioinen, due to more frequent harvest failure, whilst general yield reductions due to shortened development phases explain the decline in reliability under warming climate. Warming climate had the least effect on reliability at Lleida, especially for winter wheat, even accounting for vernalisation effects. Here, the baseline modelled annual yields were already highly variable, and the simulated response of local cultivars appears to be more tolerant to high temperatures in most models than at the German and Finnish sites.

#### 4.1.2. Precipitation-related effects

*Average yields.* In the great majority of models in the ensemble, water availability had no effect on crop development rate (cf. Table S1 in Supplement 1), so the primary effect of precipitation change on grain yield is through limitations on growth. Rain-fed wheat yields were susceptible to moisture deficiency under all climatic regimes, but particularly so at the driest locations. Hence, there was a positive relationship between median yields and precipitation at all sites in the transect, with the strongest effect at Lleida (Fig. 8b). This was expressed most strongly at temperature levels close to the baseline, and was progressively reduced under increasing warming. The latter effect is due, in part, to reduced exposure to water deficits during a shortened growth period, as well as the confounding effect of higher temperatures on water deficiency through enhanced evapotranspiration.

*Inter-model variability.* For both crop varieties, models appear to agree less in their responses to reductions in precipitation than to increases, suggesting a diversity of approaches being used to treat yield responses to water deficit (e.g. see Fig. 4). Assuming unchanged baseline temperatures, the level of agreement in modelled yields across the ensemble, described by the IQR, showed convergence with increasing precipitation for spring wheat at all sites

(Fig. S4b in Supplement 4), but was little affected by precipitation change for winter wheat (Fig. 8b). This latter result, considered alongside the divergence in modelled yields under warming (Fig. 8a), suggests that the process representation of high temperature responses in winter wheat may be contributing greater uncertainty to multi-model yield estimates than that of responses to water availability. For spring wheat, especially at sites in Germany and Finland where no overwintering is involved, models appear to converge in their estimates of responses to both warmer and wetter conditions (Fig. S4a,b).

*Inter-annual variability and yield reliability.* The sensitivity of yield reliability to precipitation was similar to that of median yields for both crop varieties, with reduced precipitation leading to an increased frequency of outcomes below the yield threshold in line with a general yield decline at all sites. In contrast, annual yield variability expressed by the CV varied considerably among crop varieties and sites. While the median CV for spring wheat at Lleida was dominated by temperature change, for other cases in regions of the IRS where results showed an increased CV with declining precipitation, it is likely that yield declines in dry years may be dominating the median yield response. This is especially true at Dikopshof for winter wheat, where precipitation exerted the dominant constraint on inter-annual variability regardless of temperature changes (Fig. 7d). It is worth noting in this context that there was also less agreement between models across the IRS for winter wheat yield responses at Dikopshof (Fig. 6d) than for any other crop variety or site, suggesting that more analysis is needed of responses to temperature and precipitation change during anomalous weather years in order to understand better the differences between models.

## 4.2. Applicability of the IRS approach

### 4.2.1. Shortcomings

Given the exploratory nature of this study, and the need to limit the number of model simulations required of modelling groups, numerous simplifications were required in the modelling protocol agreed for this initial study. These may have compromised the realism and real world applicability of some results.

(1) The calibration data provided for modellers comprised, in most cases, only a few data points and an insufficient level of detail to allow for rigorous cal-

ibration (van Keulen & Wolf 1986, Wallach et al. 2013). With such restricted calibration material, large uncertainties in estimates of impact variables such as yield can result (Makowski et al. 2002, Palosuo et al. 2011, Watson et al. 2014). Aside from the limitations in quantity of data, there are also numerous choices available in the approaches used to calibrate models. The effects of such calibration decisions can be compared explicitly in this study for those cases in which the same model version has been applied by different modelling groups (for example, compare results in Table 2 between ID numbers 7 and 8 and 25 and 26, respectively). Moreover, in terms of evaluation, the short time series of calibration data also made it impossible to conduct a statistically meaningful comparison of simulated versus observed yields at sites.

(2) Some of the climatic constraints to wheat production require more processes to be accounted for than are commonly treated in models. For instance only about half of the 24 models applied in this study account for specific heat stress impacts such as floret mortality at anthesis or leaf senescence (Challinor et al. 2005, Alderman et al. 2014). Surplus water is not accounted for in any of the models, though excess soil moisture can create several issues. For example, if this excess moisture occurs between sowing and the end of tillering, it can reduce the number of kernels per head and hence the number of tillers per plant and grain yield (Trnka et al. 2014), and heavy precipitation close to maturity may cause lodging of grain and yield losses (Peltonen-Sainio et al. 2011, Trnka et al. 2014). In addition, waterlogging can cause difficulties of access for machinery to farmland, affecting workability and trafficability (Jones & Thomasson 1985). Such omissions can lead to the over-estimation of yields compared to those observed and to discrepancies in capturing inter-annual variations, in particular during extreme weather years. Similarly, re-initialising the simulations at the beginning of each year, always with the same assumption of soil moisture, may affect simulation results when long-term trends in soil moisture are not accounted for.

(3) Inter-annual variability has been assessed both by presenting the full modelled yield distributions (CV) and by referencing yields against a threshold level (reliability). However, by reporting only median results of these measures, information may have been lost from the suite of multi-model outcomes. Analysis of percentiles towards the tails of the ensemble results could also be considered, though additional insights may require closer scrutiny of anomalous weather-years (e.g. cool, warm, wet or dry seasons) during the baseline period.

(4) The IRS depicts responses for temperature and precipitation changes alone. As these change, all other weather variables are assumed to remain fixed, even though the resulting combinations may be physically implausible (for example, model discrepancies in the relationship between temperature and humidity are described in Section 2.2.2). Temperature and precipitation changes were applied uniformly throughout the year, whereas in reality, climate model projections indicate that future climate changes will vary seasonally, and that this seasonal pattern of change varies by region (IPCC 2013a). Future changes of climate will also be associated with altered CO<sub>2</sub> levels, which themselves can be expected to affect crop growth and water use. CO<sub>2</sub> was fixed at 360 ppm in this study, so results for changed climate may not reflect realistic crop responses that can be anticipated for the future. Moreover, present-day management practices and current crop cultivars, that were assumed to be fixed, would in reality certainly be modified to suit the changing conditions. While the approach may have limitations for analysing and discriminating between responses to multiple predictors, CO<sub>2</sub> effects can be treated by constructing IRSs for different time periods, and other key variables can be analysed through construction of additional bivariate IRSs, or by assuming that they change concurrently with the primary variables (e.g. as indicated in climate model projections).

#### 4.2.2. Utility of the approach

In this study impact response surfaces have proven to be a useful device for analysing and comparing multiple model simulations of crop yield responses to changes in climate across a wide range of plausible future conditions. IRSs also offered a useful means for detecting modelling, data or transcription errors.

As has been demonstrated in some earlier applications (Luo et al. 2007, Fronzek et al. 2010, Wilby et al. 2014), IRS plots can readily be combined with climate change projections (here, of annual temperature and precipitation change), where responses for any scenario-based combination of changes based on climate models should logically fall somewhere within this response space. We have illustrated this by plotting recent climate change projections for the end of the century over central Europe (IPCC 2013a) onto the 30 yr mean plot of Fig. S2 in Supplement 3. This idea can be extended if projections of temperature and precipitation change are presented probabilistically (e.g. see Harris et al. 2010). By superimposing a joint probability distribution onto an IRS, it becomes

feasible to estimate the likelihood of exceeding a certain impact threshold, such as a critical level of yield as defined on the response surface (Fronzek et al. 2010, Børgesen & Olesen 2011, Ferrise et al. 2011). Such analyses are planned in ongoing studies, which also consider CO<sub>2</sub> level and seasonality that are required for more realistic projections of future yield (see Section 4.2.1).

The results presented here, along with follow-up studies, may help to identify model deficiencies that reflect shortcomings in process understanding or representation. Some of the other output variables generated in the modelling exercise, such as phenology, water use and total biomass, could be used to look for clues that might explain differences in modelled yield responses. There are several other potential applications of the IRS approach to crop modelling that also remain to be explored. IRSs can be constructed that explore within-model parameter uncertainties alongside between-model structural uncertainties of the type examined in this study (e.g. Fronzek et al. 2011). IRSs can also be used to analyse effects of seasonal weather anomalies through consideration of year-to-year responses (cf. Fronzek 2013).

Building on the experiences gained here, a follow-up study using the same data will explore methods of classifying IRS responses and how different patterns of impact response described in Section 3.2 might be related to known characteristics of the models. Another study, also for wheat, involving many of the same models, will endeavour to introduce more realism into the model simulations. The IRS approach can then be used to assess the potential effectiveness of farm-level adaptation measures, such as altered sowing date and cultivar selection.

Finally, this analysis has focused on comparing the behaviour of process-based wheat models under changed climate at sites in Europe. However, in principle the IRS approach can be applied in examining climate change impacts for any system or activity that can be represented by causal models that are sensitive to 2 dominant variables.

## 5. CONCLUSIONS

In this study we have reported a novel approach for inter-comparing simulated impacts of climate change across a large ensemble of models and a wide range of plausible changes in climate at diverse locations in Europe, using impact response surfaces (IRSs). The approach appears to offer an effective method of por-



traying model behaviour under changing climate, as well as numerous advantages for analysing, comparing and presenting results from multi-model ensemble simulations.

In spite of the simplified assumptions required for undertaking multiple simulations, some clear tendencies emerged from this analysis:

- Over the range of climate changes considered, median modelled yields were more sensitive to temperature change than to precipitation change at the Finnish site, while sensitivities were more evenly distributed between temperature and precipitation at the German and Spanish sites.
- From the model analysis, assuming current CO<sub>2</sub> levels, we can conclude that average yields of current wheat cultivars decline with higher temperatures and decreased precipitation, but benefit from increased precipitation.
- Warming alone, under baseline precipitation, induced remarkably similar rates of median yield decline (5–7% per 1°C, see Supplement 4) across sites and crop varieties.
- Corresponding responses to precipitation under baseline temperature were more varied (1–10% per 10% precipitation change).
- Individual model behaviour may depart markedly from the median response.

While IRSs are very helpful for summarising multiple model simulations, complementary approaches (e.g. focusing on individual model responses or on anomalous weather-years) are still required for gaining a fuller appreciation of the reasons for model behaviour. Furthermore, the bivariate nature of the IRS analysis may obscure responses attributable to other key explanatory variables, though these too can potentially be explored using other methods.

Finally, we have shown how the IRS approach can facilitate an examination of other statistical characteristics of the ensemble response, such as the inter-model and inter-annual variability. Plots such as the IQR can assist in highlighting aspects of the sensitivity to climate for which models exhibit divergent behaviour. The reliability and CV plots are instructive in revealing sensitivities of annual responses which may differ from period-average responses. Together, these help to pinpoint processes, such as heat stress, vernalisation or drought, that may require further attention in future model development.

*Acknowledgements.* This work was conducted in the context of CROPm, the crop modelling component of MACSUR (Modelling European Agriculture with Climate Change for Food Security), a project launched by the Joint Research

Programming Initiative (JPI) on Agriculture, Food Security and Climate Change (FACCE – Rötter et al. 2013, Ewert et al. 2014). The authors acknowledge financial support from the following sources: FACCE JPI, MACSUR, the Academy of Finland for the A-LA-CARTE (decisions: 140806 and 140870) and PLUMES (decisions: 277276 and 277403) projects, the European Commission Seventh Framework Programme IMPRESSIONS project (grant agreement no. 603416), the Italian Ministry of Agriculture (AGROSCENARI Project), the Italian Ministry of Research (FIRB 2012, RBFR12B2K4\_004), the German Federal Office for Agriculture and Nutrition, the MTT strategic project MODAGS, the MULCLIVAR project from the Spanish Ministerio de Economía y Competitividad (MINECO) CGL2012-38923-C02-02, the Belgium funding agency DGO-3(SPW–Belgium), University of Sassari, Polish National Centre for Research and Development, the German Federal Ministries of Education and Research, and Food and Agriculture (grant no. 2812ERA115), the Czech National Agency for Agricultural Research (project no. QJ1310123) and KONTAKT project no. LD13030.

#### LITERATURE CITED

- Abeledo LG, Savin R, Slafer GA (2008) Wheat productivity in the Mediterranean Ebro Valley: analyzing the gap between attainable and potential yield with a simulation model. *Eur J Agron* 28:541–550
- Alderman P, Quilligan E, Asseng S, Ewert F, Reynolds M (2014) Proc Workshop Model Wheat Response High Temp. CIMMYT, El Batán, Mexico, 19–21 June 2013
- Angulo C, Rötter R, Lock R, Enders A, Fronzek S, Ewert F (2013) Implication of crop model calibration strategies for assessing regional impacts of climate change in Europe. *Agric For Meteorol* 170:32–46
- Asseng S, Ewert F, Rosenzweig C, Jones JW and others (2013) Uncertainty in simulating wheat yields under climate change. *Nat Clim Change* 3:827–832
- Asseng S, Ewert F, Martre P, Rötter RP and others (2014) Rising temperatures reduce global wheat production. *Nat Clim Change* 5:143–147
- Becker R, Chambers J, Wilks A (1988) *The new S language*. Wadsworth & Brooks/Cole, Pacific Grove, CA
- Børgesen CD, Olesen JE (2011) A probabilistic assessment of climate change impacts on yield and nitrogen leaching from winter wheat in Denmark. *Nat Hazards Earth Syst Sci* 11:2541–2553
- Cartelle J, Pedró A, Savin R, Slafer GA (2006) Grain weight responses to post-anthesis spikelet-trimming in an old and a modern wheat under Mediterranean conditions. *Eur J Agron* 25:365–371
- Challinor A, Wheeler T, Craufurd P, Slingo J (2005) Simulation of the impact of high temperature stress on annual crop yields. *Agric For Meteorol* 135:180–189
- Challinor AJ, Simelton ES, Fraser ED, Hemming D, Collins M (2010) Increased crop failure due to climate change: assessing adaptation options using models and socio-economic data for wheat in China. *Environ Res Lett* 5: 034012
- Challinor A, Martre P, Asseng S, Thornton P, Ewert F (2014a) Making the most of climate impacts ensembles. *Nat Clim Change* 4:77–80
- Challinor AJ, Watson J, Lobell DB, Howden SM, Smith DR, Chhetri N (2014b) A meta-analysis of crop yield under

- climate change and adaptation. *Nat Clim Change* 4: 287–291
- Cleveland WS (1993) *Visualizing data*. Hobart Press, Summit, NJ
- Craufurd PQ, Vadez V, Krishna Jagadish SV, Vara Prasad PV, Zaman-Allah M (2013) Crop science experiments designed to inform crop modeling. *Agric For Meteorol* 170: 8–18
- Diaz-Nieto J, Wilby RL (2005) A comparison of statistical downscaling and climate change factor methods: impacts on low flows in the River Thames, United Kingdom. *Clim Change* 69:245–268
- Easterling WE, Aggarwal PK, Batima P, Brander KM and others (2007) Food, fibre and forest products. In: Parry ML, Canziani OF, Palutikof JP, van der Linden PJ, Hanson CE (eds) *Climate change 2007: impacts, adaptation and vulnerability*. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, p 273–313
- Ewert F, Rötter R, Bindi M, Webber H and others (2014) Crop modelling for integrated assessment of risk to food production from climate change. *Environ Model Softw*, doi:10.1016/j.envsoft.2014.12.003
- Ferrise R, Moriondo M, Bindi M (2011) Probabilistic assessments of climate change impacts on durum wheat in the Mediterranean region. *Nat Hazards Earth Syst Sci* 11: 1293–1302
- Fronzek S (2013) Climate change and the future distribution of palsa mires: ensemble modelling, probabilities and uncertainties. *Monogr Boreal Environ Res No. 44*, <http://hdl.handle.net/10138/40184>
- Fronzek S, Carter TR, Raisanen J, Ruokolainen L, Luoto M (2010) Applying probabilistic projections of climate change with impact models: a case study for sub-arctic palsa mires in Fennoscandia. *Clim Change* 99:515–534
- Fronzek S, Carter TR, Luoto M (2011) Evaluating sources of uncertainty in modelling the impact of probabilistic climate change on sub-arctic palsa mires. *Nat Hazards Earth Syst Sci* 11:2981–2995
- Gitay H, Brown S, Easterling W, Jallow B and others (2001) Ecosystems and their goods and services. In: McCarthy JJ, Canziani OF, Leary NA, Dokken DJ, White KS (eds) *Climate change 2001: impacts, adaptation, and vulnerability*. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, p 235–242
- Hanasaki N, Masutomi Y, Takahashi K, Hijioka Y, Harasawa H, Matsuoka Y (2007) Development of a global water resources scheme for climate change policy support models. *Environmental systems research*. *Jpn Soc Civ Eng* 35:367–374 (in Japanese)
- Harris GR, Collins M, Sexton DMH, Murphy JM, Booth BBB (2010) Probabilistic projections for 21st century European climate. *Nat Hazards Earth Syst Sci* 10:2009–2020
- Iizumi T, Yokozawa M, Nishimori M (2011) Probabilistic evaluation of climate change impacts on paddy rice productivity in Japan. *Clim Change* 107:391–415
- IPCC (2013a) Annex I: Atlas of global and regional climate projections [van Oldenborgh GI and others]. In: Stocker TF, Quinn D, Plattner GK, Tignor M and others (eds) *Climate change 2013: the physical science basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, p 1311–1393
- IPCC (2013b) Annex II: Climate system scenario tables [Prather M and others]. In: Stocker TF, Quinn D, Plattner GK, Tignor M and others (eds) *Climate change 2013: the physical science basis* contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, p 1395–1445
- Jamieson PD, Porter JR, Goudriaan J, Ritchie JT, van Keulen H, Stol W (1998) A comparison of the models AFR-CWHEAT2, CERES-Wheat, Sirius, SUCROS2 and SWHEAT with measurements from wheat grown under drought. *Field Crops Res* 55:23–44
- Jones RJA, Thomasson AJ (1985) An agroclimatic databank for England and Wales. *Tech Monogr No. 16*, Soil survey of England and Wales, Harpenden
- Lobell DB, Field CB (2007) Global scale climate–crop yield relationships and the impacts of recent warming. *Environ Res Lett* 2:014002
- Luo Q, Bellotti W, Williams M, Cooper I, Bryan B (2007) Risk analysis of possible impacts of climate change on South Australian wheat production. *Clim Change* 85:89–101
- Makowski D, Wallach D, Tremblay M (2002) Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. *Agronomie* 22: 191–203
- Metzger M, Bunce R, Jongman R, Mürger C, Watkins J (2005) A climatic stratification of the environment of Europe. *Glob Ecol Biogeogr* 14:549–563
- Monfreda C, Ramankutty N, Foley JA (2008) Farming the planet. 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochem Cycles* 22:GB1022, doi:10.1029/2007GB002947
- Palosuo T, Kersebaum KC, Angulo C, Hlavinka P and others (2011) Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. *Eur J Agron* 35:103–114
- Peltonen-Sainio P, Jauhiainen L, Hakala K (2011) Crop responses to temperature and precipitation according to long-term multi-location trials at high-latitude conditions. *J Agric Sci* 149:49–62
- Porter J (1993) AFR-CWHEAT2: a model of the growth and development of wheat incorporating responses to water and nitrogen. *Eur J Agron* 2:69–82
- Porter JR, Gawith M (1999) Temperatures and the growth and development of wheat: a review. *Eur J Agron* 10: 23–26
- Porter JR, Xie L, Challinor A, Cochrane K and others (2014) Food security and food production systems. In: Field CB and others (eds) *Climate change 2014: impacts, adaptation and vulnerability*. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, p 485–533
- R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. [www.R-project.org/](http://www.R-project.org/)
- Rosenzweig C, Jones JW, Hatfield JL, Ruane AC and others (2013) The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. *Agric For Meteorol* 170:166–182
- Rosenzweig C, Elliott J, Deryng D, Ruane AC and others (2014) Assessing agricultural risks of climate change in the 21st century in a global gridded crop model inter-

- comparison. *Proc Natl Acad Sci USA* 111:3268–3273
- Rötter RP (2014) Agricultural impacts: robust uncertainty. *Nat Clim Change* 4:251–252
  - Rötter RP, Carter TR, Olesen JE, Porter JR (2011) Crop-climate models need an overhaul. *Nat Clim Change* 1:175–177
  - Rötter RP, Ewert F, Palosuo T, Bindi M and others (2013) Challenges for agro-ecosystem modelling in climate change risk assessment for major European crops and farming systems. *Proc Impacts World 2013 Conf*, Potsdam, 27–30 May 2013, p 555–564
  - Ruane AC, McDermaid S, Rosenzweig C, Baigorria GA and others (2014) Carbon–temperature–water change analysis for peanut production under climate change: a prototype for the AgMIP Coordinated Climate–Crop Modeling Project (C3MP). *Glob Chang Biol* 20:394–407
  - Trnka M, Rötter RP, Ruiz-Ramos M, Kersebaum KC and others (2014) Adverse weather conditions for European wheat production will become more frequent with climate change. *Nat Clim Change* 4:637–643
  - van Ittersum M, Rabbinge R (1997) Concepts in production ecology for analysis and quantification of agricultural input-output combinations. *Field Crops Res* 52:197–208
  - van Ittersum MK, Leffelaar PA, Van Keulen H, Kropff MJ, Bastiaans L, Goudriaan J (2003) On approaches and applications of the Wageningen crop models. *Eur J Agron* 18:201–234
  - van Keulen H, Wolf J (1986) Modelling of agricultural production: weather, soils and crops. PUDOC, Wageningen
  - Wallach D, Makowski D, Jones JW, Brun F (2013) Working with dynamic crop models: methods, tools and examples for agriculture and environment. Academic Press, London
  - Watson J, Challinor AJ, Fricker TE, Ferro CA (2014) Comparing the effects of calibration and climate errors on a statistical crop model and a process-based crop model. *Clim Change*, doi:10.1007/s10584-014-1264-3014-1264-3
  - Wetterhall F, Graham LP, Andreasson J, Rosberg J, Yang W (2011) Using ensemble climate projections to assess probabilistic hydrological change in the Nordic region. *Nat Hazards Earth Syst Sci* 11:2295–2306
  - White JW, Hoogenboom G, Kimball BA, Wall GW (2011) Methodologies for simulating impacts of climate change on crop production. *Field Crops Res* 124:357–368
  - Wilby RL, Dawson CW, Murphy C, O'Connor P, Hawkins E (2014) The Statistical DownScaling Model-Decision Centric (SDSM-DC): conceptual basis and applications. *Clim Res* 61:259–276
  - Wu L, Kersebaum KC (2008) Modeling water and nitrogen interaction responses and their consequences in crop models. In: Ahuja LR, Reddy VR, Saseendran SA, Yu Q (eds) *Response of crops to limited water: understanding and modeling water stress effects on plant growth processes*. ASA-CSSA-SSSA, Madison, WI, p 215–249
  - Zadoks JC, Chang TT, Konzak CF (1974) A decimal code for the growth stages of cereals. *Weed Res* 14:415–421

*Submitted: January 15, 2015; Accepted: June 22, 2015*

*Proofs received from author(s): August 6, 2015*