Agricultural Systems 159 (2018) 209–224

Contents lists available at ScienceDirect

Agricultural Systems

journal homepage: www.elsevier.com/locate/agsy

# Classifying multi-model wheat yield impact response surfaces showing sensitivity to temperature and precipitation change

Stefan Fronzek[a,*], Nina Pirttioja[a], Timothy R. Carter[a], Marco Bindi[b], Holger Hoffmann[c], Taru Palosuo[d], Margarita Ruiz-Ramos[e], Fulu Tao[d], Miroslav Trnka[f,g], Marco Acutis[h], Senthold Asseng[i], Piotr Baranowski[j], Bruno Basso[k], Per Bodin[l], Samuel Buis[m], Davide Cammarano[n], Paola Deligios[o], Marie-France Destain[p], Benjamin Dumont[p], Frank Ewert[c,q], Roberto Ferrise[b], Louis François[p], Thomas Gaiser[c], Petr Hlavinka[f,g], Ingrid Jacquemin[p], Kurt Christian Kersebaum[q], Chris Kollas[q,u], Jaromir Krzyszczak[j], Ignacio J. Lorite[r], Julien Minet[p], M. Ines Minguez[e], Manuel Montesino[s], Marco Moriondo[t], Christoph Müller[u], Claas Nendel[q], Isik Öztürk[v], Alessia Perego[h], Alfredo Rodríguez[e], Alex C. Ruane[w], Françoise Ruget[m], Mattia Sanna[h], Mikhail A. Semenov[x], Cezary Slawinski[j], Pierre Stratonovitch[x], Iwan Supit[y], Katharina Waha[u,z], Enli Wang[aa], Lianhai Wu[ab], Zhigan Zhao[aa,ac], Reimund P. Rötter[ad,ae]

[a] Finnish Environment Institute (SYKE), 00251 Helsinki, Finland
[b] University of Florence, 50144 Florence, Italy
[c] INRES, University of Bonn, 53115 Bonn, Germany
[d] Natural Resources Institute Finland (Luke), 00790 Helsinki, Finland
[e] CEIGRAM-AgSystems, Universidad Politecnica de Madrid, 28040 Madrid, Spain
[f] Institute of Agrosystems and Bioclimatology, Mendel University in Brno, Brno 613 00, Czech Republic
[g] Global Change Research Centre AS CR, v.v.i., 603 00 Brno, Czech Republic
[h] University of Milan, 20133 Milan, Italy
[i] University of Florida, Gainesville, FL 32611, USA
[j] Institute of Agrophysics, Polish Academy of Sciences, 20-290 Lublin, Poland
[k] Michigan State University, East Lansing, MI 48824, USA
[l] Lund University, 223 62 Lund, Sweden
[m] INRA, UMR 1114 EMMAH, F-84914 Avignon, France
[n] James Hutton Institute, Invergowrie, Dundee, DD2 5DA, Scotland
[o] University of Sassari, 07100 Sassari, Italy
[p] Université de Liège, 4000 Liège, Belgium
[q] Leibniz Centre for Agricultural Landscape Research (ZALF), 15374 Müncheberg, Germany
[r] IFAPA Junta de Andalucia, 14004 Córdoba, Spain
[s] University of Copenhagen, 2630 Taastrup, Denmark
[t] CNR-IBIMET, 50145 Florence, Italy
[u] Potsdam Institute for Climate Impact Research, 14473 Potsdam, Germany
[v] Aarhus University, 8830 Tjele, Denmark
[w] NASA Goddard Institute for Space Studies, New York, NY 10025, USA
[x] Rothamsted Research, Harpenden, Herts, AL5 2JQ, UK
[y] Wageningen University, 6700AA Wageningen, The Netherlands
[z] CSIRO Agriculture Flagship, 4067 St Lucia, Australia
[aa] CSIRO Agriculture Flagship, 2601 Canberra, Australia
[ab] Rothamsted Research, North Wyke, Okehampton, EX20 2SB, UK
[ac] China Agricultural University, 100094 Beijing, China
[ad] University of Göttingen, Tropical Plant Production and Agricultural Systems Modelling (TROPAGS), Grisebachstraße 6, 37077 Göttingen, Germany
[ae] University of Göttingen, Centre for Biodiversity and Sustainable Land Use (CBL), Büsgenweg 1, 37077 Göttingen, Germany

* Corresponding author.
*E-mail address:* stefan.fronzek@ymparisto.fi (S. Fronzek).

http://dx.doi.org/10.1016/j.agsy.2017.08.004
Received 28 October 2016; Received in revised form 30 June 2017; Accepted 18 August 2017
Available online 05 September 2017

ARTICLE INFO

*Keywords:*
Classification
Climate change
Crop model
Ensemble
Sensitivity analysis
Wheat

ABSTRACT

Crop growth simulation models can differ greatly in their treatment of key processes and hence in their response to environmental conditions. Here, we used an ensemble of 26 process-based wheat models applied at sites across a European transect to compare their sensitivity to changes in temperature (−2 to +9°C) and precipitation (−50 to +50%). Model results were analysed by plotting them as impact response surfaces (IRSs), classifying the IRS patterns of individual model simulations, describing these classes and analysing factors that may explain the major differences in model responses.

The model ensemble was used to simulate yields of winter and spring wheat at four sites in Finland, Germany and Spain. Results were plotted as IRSs that show changes in yields relative to the baseline with respect to temperature and precipitation. IRSs of 30-year means and selected extreme years were classified using two approaches describing their pattern.

The expert diagnostic approach (EDA) combines two aspects of IRS patterns: location of the maximum yield (nine classes) and strength of the yield response with respect to climate (four classes), resulting in a total of 36 combined classes defined using criteria pre-specified by experts. The statistical diagnostic approach (SDA) groups IRSs by comparing their pattern and magnitude, without attempting to interpret these features. It applies a hierarchical clustering method, grouping response patterns using a distance metric that combines the spatial correlation and Euclidian distance between IRS pairs. The two approaches were used to investigate whether different patterns of yield response could be related to different properties of the crop models, specifically their genealogy, calibration and process description.

Although no single model property across a large model ensemble was found to explain the integrated yield response to temperature and precipitation perturbations, the application of the EDA and SDA approaches revealed their capability to distinguish: (i) stronger yield responses to precipitation for winter wheat than spring wheat; (ii) differing strengths of response to climate changes for years with anomalous weather conditions compared to period-average conditions; (iii) the influence of site conditions on yield patterns; (iv) similarities in IRS patterns among models with related genealogy; (v) similarities in IRS patterns for models with simpler process descriptions of root growth and water uptake compared to those with more complex descriptions; and (vi) a closer correspondence of IRS patterns in models using partitioning schemes to represent yield formation than in those using a harvest index.

Such results can inform future crop modelling studies that seek to exploit the diversity of multi-model ensembles, by distinguishing ensemble members that span a wide range of responses as well as those that display implausible behaviour or strong mutual similarities.

## 1. Introduction

A wide range of dynamic crop growth simulation models have been developed over the past few decades, many of which are being applied to study impacts of climate change (Asseng et al., 2014; Challinor et al., 2014; Ewert et al., 2015; Jones et al., 2017a, 2017b; White et al., 2011). These models can differ greatly in their treatment of key processes and hence in their response to environmental conditions (Asseng et al., 2013; Palosuo et al., 2011; Rötter et al., 2012). Therefore, it is of interest to examine model behaviour under changed climate in order to characterise the types of responses estimated, contrast the responses of different models and consider the reasons for these differences.

Fundamental structural differences in the way models simulate processes such as development, assimilation, partitioning and water and nutrient uptake can be traced back to the purposes for which models were originally developed, their region of origin and the scale of their application (Challinor et al., 2009; van Ittersum et al., 2003). Given the many factors determining crop response, it is not surprising that processes are accorded variable emphasis across different models. For instance, a model developed to examine field-level processes of yield formation under well-watered conditions might focus on growth processes and the partitioning of dry matter, relying on only a simple parameterization of soil water availability. Conversely, regional yield estimates under water-limited conditions might demand a detailed representation of soil water and nutrient uptake, while adopting a simple approach to estimating yield components. Moreover, most models have not been developed independently and may share common antecedents and genealogy, which may provide clues to their comparative behaviour. Models that have evolved from a predecessor can hence exhibit many similar characteristics while including new processes or alternative descriptions of existing processes (Bouman et al., 1996; Rosenzweig et al., 2014).

However, model structure alone cannot explain all of the reported differences between model behaviour under a changing climate. Model calibration – the procedure of adjusting parameter values to obtain a good fit between model outputs and observations (Acutis and Confalonieri, 2006; Kersebaum et al., 2015) – may also play a significant role. Unless fixed calibration techniques have been pre-specified, most model inter-comparison exercises typically rely on “modellers' choice” for the techniques that are applied to the available observations. The techniques themselves can vary from trial and error methods through to optimization and Bayesian techniques (cf. Acutis and Confalonieri, 2006; Angulo et al., 2013), and must necessarily be tailored to the parameters of a given model. Even then, there may be differences in the number of parameters treated and in how the calibration data are interpreted and the techniques deployed (Confalonieri et al., 2016; Palosuo et al., 2011).

Here, a multi-model ensemble approach has been adopted to explore patterns of simulated yield response under climate change. We use an ensemble of wheat models at sites across a European transect (in Finland, Germany and Spain – Pirttioja et al., 2015) and compare their sensitivity to changes in climate by plotting simulated yield as impact response surfaces (IRS; Fronzek et al., 2010). An IRS is plotted from the results of a sensitivity analysis of an impact model with respect to changes in two key climatic variables, e.g. changes in annual mean surface temperature and annual precipitation. The observed baseline climate is adjusted with systematic increments over a range of values. Impacts are computed for each combination of changes in the two climate variables and plotted as contours on a two-dimensional IRS. Examples of IRS applications using crop models include estimates of yield response for maize (Ruane et al., 2013) and barley (Kim et al., 2013), of nitrogen leaching from wheat cultivation (Børgesen and Olesen, 2011) and of adaptation options in wheat cultivation (Ruiz-Ramos et al., 2017). The approach has also been applied in other sectors including hydrology (Holmberg et al., 2014; Prudhomme et al., 2013a; Weiß and Alcamo, 2011) and biodiversity (Fronzek et al., 2011).

IRSs of yield changes were presented in Pirttioja et al. (2015) as multi-model ensemble medians and inter-quartile ranges, focusing on long-term averages. This paper extends that work by classifying the responses of individual models and attempting to interpret differences in response between groups of models by contrasting their structure and representation of selected key processes as well as their behaviour in selected anomalous years.

For the same four sites described in this paper, Pirttioja et al. (2015) found that ensemble median yields decline with higher temperatures and decreased precipitation, but increase with higher precipitation. However, these aggregate responses disguise individual model behaviours that can depart markedly from the average. Furthermore, there are differences in responses between sites. For example, in contrast to sites in Germany and Spain, the site in Finland was more sensitive to temperature than precipitation changes.

This paper has the following objectives:

1. To classify patterns of 30-year mean yield response to changes in temperature and precipitation simulated by a multi-model ensemble of wheat models;
2. To describe these classes, their utility and differences between them;
3. To analyse factors that may explain the major differences in model responses; and
4. To examine if model behaviour in anomalous years can be related to classes of 30-year mean response.

Section 2 describes the original model simulations and introduces the classification approaches adopted in this paper. Section 3.1 presents and contrasts the results of the classification (objectives 1 and 2), which are then discussed in Sections 4.1 and 4.2, respectively. Factors explored in the paper that might explain differences in modelled yield responses (objective 3), include site locations and crop varieties selected for the simulations, model characteristics such as the description of key processes and model genealogy, and model calibration. The nature of their relationships to yield response is described in Section 3.2 and discussed in Section 4.1. Finally, patterns of model behaviour during individual, anomalous years are compared to those for modelled 30-year mean responses (objective 4) in Section 3.3.

## 2. Material and methods

The ensemble analysis presented in Pirttioja et al. (2015) underpins the work presented in this paper. Sub-section 2.1 provides a brief summary of the wheat models, data, modelling protocol and construction of impact response surfaces (IRSs) that are reported in more detail in the earlier paper. The following four sub-sections describe the new analyses undertaken for this paper. Methods of classifying IRS patterns are set out in Sub-section 2.2 and criteria for grouping wheat models according to various properties are introduced in sub-section 2.3. Procedures for relating IRS patterns to model properties are outlined in sub-section 2.4, while an analysis of IRS patterns in anomalous years is described in sub-section 2.5.

### 2.1. Crop modelling

#### 2.1.1. Crop models

An ensemble of 26 dynamic crop growth models (Table 1) has been employed in this study to simulate the growth and yield of local spring

**Table 1**
Wheat models applied in this study (identifier and genealogy), their calibration (approach and number of parameters) and their representation of some key processes (evapotranspiration, water dynamics, root distribution, water stress, heat stress and yield formation). Entries are based on Table S1 in Pirttioja et al. (2015), with some minor modifications and streamlining.

| Model identifiers | | | Calibration | | Process description | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ID | Model | Family | Approach | Number of parameters | Evapo-Transpiration | Water dynamics | Root distribution | Water stress | Heat stress | Yield formation |
| a | b | c | d | e | f | g | h | i | j | k |
| 1 | AFRCWHEAT2 | I | A | > 9 | PM | C | E | E/S | M | P |
| 2 | APSIM-Nwheat | C | – | – | PT | C | E | S | V | P |
| 3 | APSIM-Wheat | C | M | 6–9 | PT | C/R | O | E | V | P |
| 4 | AquaCrop | – | M | 6–9 | PM | C | O | E/S | R | H* |
| 5 | ARMOSA | S | M | 3–5 | PT | C | O | E/S | M | P |
| 6 | CARAIB | – | M | – | P | C/R | O | S | – | H |
| 7 | CERES-wheat DSSAT v4.5 | C | M | 6–9 | PT | C | E | E/S | – | G |
| 8 | CERES-wheat DSSAT v4.5 | C | A | 6–9 | PT | C | E | E/S | – | G |
| 9 | CERES-wheat DSSAT v4.6 | C | A | 6–9 | PT | C | E | E/S | – | P |
| 10 | CropSyst | – | M | 0–2 | PM | C/R | E | E | R | H |
| 11 | DNDC | – | M | 4 | P | C | E | E/S | P | H |
| 12 | FASSET | – | – | 0–2 | M | C | E | E/S | – | H |
| 13 | HERMES | S/C | M | 6–9 | PM | C | E | E/S | – | P |
| 14 | LINTUL-4 | L | M | > 9 | P | C | L | E | – | P |
| 15 | LPJ-GUESS | E | N | 0–2 | PT | C | L | E/S | P | H* |
| 16 | LPJmL | E | M | 3–5 | PT | C | E | E | – | H* |
| 17 | MCWLA-Wheat | – | M | – | PM | R | E | E | M | H |
| 18 | MONICA | S/C | M | 3–5 | PM | C | E | E | V | P |
| 19 | SALUS | C | – | – | PT | C | E | E | V | P |
| 20 | SIMPLACE < Lintul2, Slim > | L | M | > 9 | P | C | L | E | P | P |
| 21 | Sirius | – | M | 3–5 | PT | C | L | E | – | P |
| 22 | SiriusQuality | I | M | 6–9 | P | C | E | S | – | P |
| 23 | SPACSYS | – | A | – | PM | R | E | E | – | P |
| 24 | STICS | – | A | 6–9 | SW | C | O | E/S | M | H* |
| 25 | WOFOST | S | M | 3–5 | P | C | L | E/S | – | P |
| 26 | WOFOST | S | M | 3–5 | P | C | L | E/S | – | P |

**a Model identifier**; **b Model name**; **c Genealogy**: C, CERES; L, LINTUL; E, EPIC; S, SUCROS; I, Sirius; –, not reported; **d Calibration approaches** (based on a survey of modellers): A, automatic; M, manual; N, no calibration; –, not available; **e Number of parameters evaluated** (from survey); **f Evapotranspiration**: P, Penman; PM, Penman-Monteith; PT, Priestley–Taylor; M, Makkink; SW, Shuttleworth and Wallace; **g Water dynamics**: C, capacity approach; R, Richards approach; **h Root distribution over depth**: L, linear; E, exponential; O, other; **i Water stress**: E, actual to potential evapotranspiration ratio, S, soil available water in root zone; **j Heat stress**: V, vegetative organ; R, reproductive organ; P, phenology; M, multiple processes; –, not reported; **k Yield formation**: P, partitioning; G, grain number and biomass; H*, modified harvest index; H, harvest index.

and winter wheat cultivars at four sites in Europe. Two models have been used by different modelling groups who conducted their model calibration independently. These are regarded as separate models in the ensemble.

#### 2.1.2. Study sites and weather data

Crop model simulations were conducted with observed weather station data from the four sites across a European transect covering a wide range of climatic conditions (Table 2): Jokioinen in Finland, Nossen (only spring wheat) and Dikopshof (only winter wheat) in Germany and Lleida in Spain. Wheat cultivation at these sites ranges from being pre-dominantly temperature-limited at Jokioinen to largely water-limited at Lleida for rain-fed cultivation, while the sites in Germany experience climatically more optimal conditions. For more details see Pirttioja et al. (2015).

Daily weather variables for the period 1980–2010 obtained for the sites consist of minimum and maximum temperature, precipitation, global radiation, wind speed, minimum and maximum dew point temperature, minimum and maximum actual vapour pressure and minimum and maximum relative humidity.

Perturbations of the baseline weather data were prepared for a crop model sensitivity analysis using a simple change factor approach (e.g. Ekström et al., 2015) for combined changes in temperature and precipitation. Constant changes for all days of the baseline period were applied for temperature over the range −2 to +9 °C (at 1 °C intervals) and for precipitation over the range −50 to +50% (at 10% intervals), resulting in 132 combinations of weather perturbations. The ranges were chosen to encompass the widest available range of climate model projections by the mid-21st century at the four sites. For simplicity, all other weather variables than precipitation and minimum and maximum temperature were kept unchanged.

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

Annual sowing dates for the Finnish and German sites were calculated from observed sowing dates and used for crop model simulations, whereas a fixed sowing date was selected for the Spanish site (Pirttioja et al., 2015). A generalized soil type (clay loam) was used both for calibration and simulations, but modellers also had the option to use local soil information for calibration. Calibration data on management, plant development and yields (for Lleida, taken from Abeledo et al., 2008; Cartelle et al., 2006) are described in Pirttioja et al. (2015).

#### 2.1.4. Modelling protocol

A sensitivity analysis was carried out with the ensemble of crop models. Phenological stages and crop yield were simulated for 30 baseline years (1981–2010) and 132 perturbations of temperature and precipitation for winter and spring wheat at the four sites. This resulted in 23,760 simulated seasons per model. Sowing dates, atmospheric $CO_2$ concentrations and other weather variables than temperature and precipitation were kept at their baseline values for all perturbations. Simulations were carried out on a daily time-step for water-limited yields assuming optimal nutrients. Model outputs analysed in this paper were the grain dry matter (DM) yields for each season. To avoid unrealistically long growing periods, harvest cut-off dates (day of the year [DOY] 258 for Finland and Spain and DOY 274 for Germany) were applied if maturity was not reached earlier, and grain yields were then set to 0 kg ha$^{-1}$ (Pirttioja et al., 2015). This occurred mainly at the Finnish site with up to 39% of simulations for spring wheat and 15% for winter wheat affected under −2 °C cooling (results not shown here). Under baseline conditions the respective values were 9% and 1%. In sharp contrast, for the German and Spanish sites the cut-off was most often applied for winter wheat under +9 °C warming, with 3% and fewer than 1% of years affected, respectively (an affect attributable to failure of the modelled crop to vernalise).

#### 2.1.5. Calculation of impact response surfaces

Changes in grain yields for each perturbation of the weather input data were calculated relative to the simulated yields under unperturbed baseline conditions. Results of these were then plotted as contour lines (defining the impact response surface) by bi-linear interpolation of the yield changes at each simulated increment with respect to change in annual temperature along the x-axis and precipitation along the y-axis. This was done for the 30-year averages and for selected individual years for each model, site and crop variety.

### 2.2. Classifying crop model responses

The pattern of bivariate response depicted in an IRS plot provides a visual impression of modelled crop behaviour under climate change at a given site relative to the conditions during a reference year or averaged over a period of years. Since patterns of yield response to climate vary by site and by model, it can be instructive to group similar patterns into classes. This can help both in discriminating between types of model behaviour at different sites, but could also be used for examining links between this behaviour and different characteristics of the models themselves (or model calibrations) that might explain their varied patterns of response.

Two alternative approaches were explored to classify typical patterns of yield response, plotted as percentage changes relative to the baseline climate. To assist in discriminating between them, we have labelled them the expert diagnostic approach and the statistical diagnostic approach.

#### 2.2.1. Expert diagnostic approach (EDA)

The expert diagnostic approach (EDA) relies on an understanding of the responses being represented. This is based on classifying the IRSs according to two key aspects of the response (Fig. 1a):

1. The location of the maximum yield compared to the baseline with respect to both temperature and precipitation, Classes are labelled T−, T0 and T+ for maximum yield occurring at >1 °C below, within ±1 °C of, and at >1 °C above the baseline temperature, respectively, and correspondingly P−, P0 and P+ for maximum yield at >10% below, within ±10% of, and at >10% above the baseline precipitation. In combination this produces nine classes.
2. The strength of response, defined as the rate of change in yield expressed separately with respect to changes in precipitation and temperature relative to the location of the maximum. Classes are defined for a strong or weak response to temperature change, labelled T" (strong) and T' (weak) and for a strong (P") and weak (P')

**Table 2**
Weather stations used in this study, environmental zones as defined by Metzger et al. (2005) and wheat variety simulated. Coordinates are given in decimal degrees latitude and longitude, mean annual temperature, monthly temperature range and precipitation are for the 1981–2010 period-mean.

| Site | Country | Coordinates | Environmental zone | Mean annual temperature (and monthly range) | Annual precipitation | Wheat species |
|---|---|---|---|---|---|---|
| Jokioinen | Finland | 60.81°N, 23.5°E | Boreal | 4.7 (23.0) °C | 627 mm | Winter and spring wheat |
| Nossen | Germany | 51.06°N, 13.27°E | Continental | 9.3 (18.7) °C | 722 mm | Spring wheat |
| Dikopshof | Germany | 50.81°N, 6.95°E | Atlantic Central | 10.3 (16.2) °C | 629 mm | Winter wheat |
| Lleida | Spain | 41.63°N, 0.60°E | Mediterranean South | 15.0 (19.7) °C | 342 mm | Winter and spring wheat |

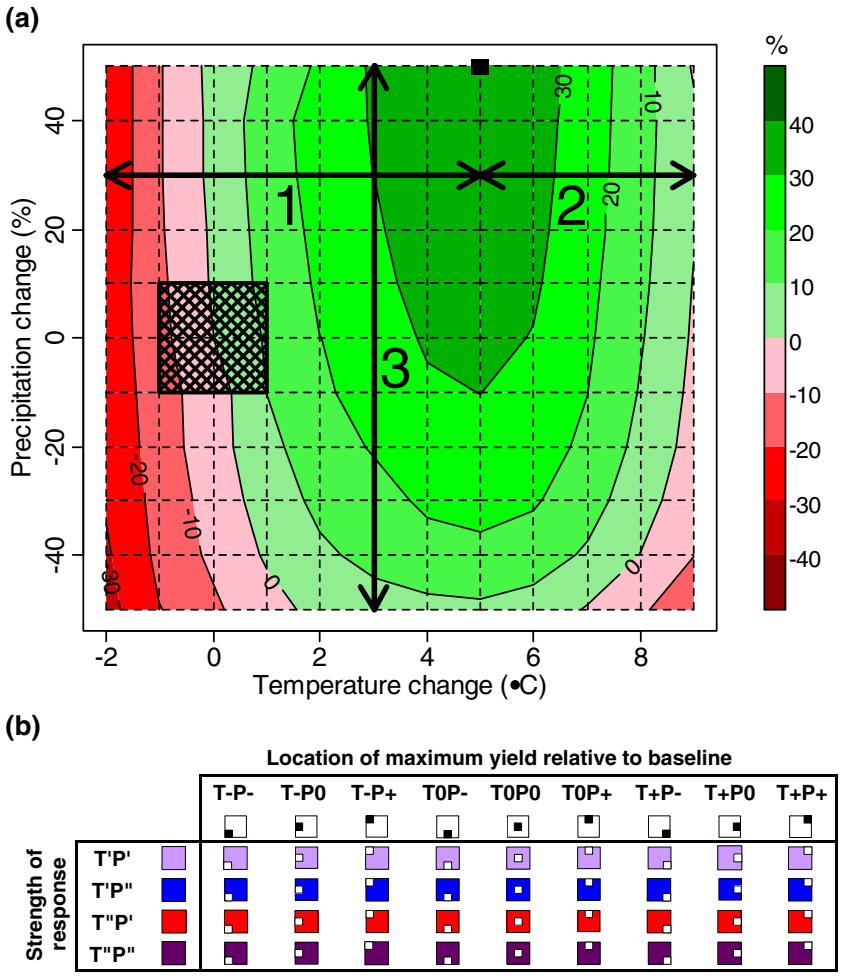


**Fig. 1.** Expert diagnostic approach for classifying impact response surfaces (IRSs) illustrating the method (upper panel, a) and full classification (lower panel, b). Upper panel illustrates the method of computing the location of maximum response (black dot in this example) and the rate of change (strength) of response (arrows) for a typical IRS depicting changes in 30-yr averaged dry matter grain yield relative to the 1981–2010 baseline (%). Maxima falling in the hatched rectangle are considered to be located at the baseline. The numbered arrows illustrate how the strength of yield response is first identified separately, for both sides of the location of the maximum with respect to both temperature (1 and 2) and precipitation (3). Here only one calculation is required for precipitation, as the maximum occurs at the top edge of the IRS. Calculations are made across all increments used to construct the IRS (dashed lines). For detailed explanation, see Supplement S1.1. The lower panel shows the full classification, which is defined according to location of the maximum yield (black squares in icons – nine columns), strength of the yield response (coloured icons – four rows) and combinations of these (36 cells). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

response to precipitation change. There are four possible combinations of these responses.

Combining classes for the location of the maximum yield with those for the strength of yield response thus produces 9 × 4 = 36 potential classes (Fig. 1b). More details of the classification can be found in Supplement S1.1. The EDA classification was performed separately for spring and winter wheat.

### 2.2.2. Statistical diagnostic approach (SDA)

The statistical diagnostic approach (SDA) groups IRSs according to a comparison of their pattern and magnitude, without attempting to interpret these features. This applies a hierarchical clustering method using a distance metric, *d*, which is the product of the spatial correlation and Euclidian distance between corresponding yield change values across IRS pairs. IRSs for each crop were grouped by hierarchical clustering that minimizes the distances between members of each cluster using the agglomerative nesting algorithm (Kaufman and Rousseeuw, 1990) implemented in the R package "cluster" (Maechler et al., 2016) with the average method to determine clusters. A dendrogram illustrating the clustering method for winter wheat is shown in Fig. 2. More details and illustrations of the clustering approach and distance metric are presented in Supplement 1.2.

## 2.3. Grouping models by their properties

Properties of different models were classified according to their genealogy, calibration and process description (Table 1). Where information was available, crop models were first grouped into broad "model families" that share a common genealogy (column c in Table 1). These expand the information provided by Pirttioja et al. (2015) and originally taken from Asseng et al. (2013), Palosuo et al. (2011) and Rötter et al. (2012).

As the modelling protocol did not specify a specific calibration procedure beyond the provision of calibration data, different approaches were applied by different modelling groups. Information about these was collected in a survey to which most groups, but not all, responded. This included text descriptions of the general approach taken and the number of model parameters tested and modified during the calibration. Three calibration approaches were identified: automatic, manual and no calibration, and for each model the number of parameters tested was allocated to one of four classes (Table 1, columns d and e).

Documentation about the description of key processes in the crop models was based on updated information from Table 1 in Pirttioja et al. (2015). The process descriptions included the methods used to estimate reference crop evapotranspiration, and water dynamics, root distribution, water stress, heat stress and yield formation (Table 1, columns f–k).

## 2.4. Linking model responses to model properties

In order to explore whether IRS patterns of modelled yield response to temperature and precipitation perturbations can be related to particular properties of the models, IRS classes based on the alternative expert and statistical diagnostic approaches were related directly to the model properties grouped in Table 1. The EDA classification (Sub-section 2.2.1) was applied by recording the frequency of IRS classes


**Fig. 2.** Dendrogram of the hierarchical clustering method used to classify IRS period-mean yield responses across all models and sites for winter wheat ($n$ = 78). Proximity in patterns for any two model/site combinations can be judged by tracing the distance along branches of the dendrogram that connect them. Close proximity indicates high similarity, with numbered C1–C8 clusters of similar patterns. Models are identified with numbers (for model names see Table 1). The equivalent dendrogram for spring wheat can be found as Fig. S1 in Supplement 1.2.

represented by those model responses sharing a given property. This was done separately for classes indicating the location of the maximum yield, the strength of response and combinations of these.

The metric of distance between IRS patterns used in defining SDA-derived clusters (Sub-section 2.2.2) was used to compare model responses that share the same property. The mean distance between pairs of IRSs across a group of models (i.e. all possible combinations) was calculated and compared to the mean distance statistics for the full ensemble of models as well as mean distances for 100 randomly selected groups of the same size to judge the significance of the result for the specific group (Supplement S1.2, Fig. S2). This provided a measure of similarity (distance ratio) between any group of models (defined by a common property) and the full IRS ensemble. The analysis was also conducted for the median distance within a group, whose results were very similar to the mean distances and are therefore not presented here.

### 2.5. Analysis of model responses in anomalous years

The EDA classification was also applied to IRSs for individual years, expressed relative to the yield for the unperturbed weather of a given year, to demonstrate how model responses can differ from the long-term mean. Results for spring wheat during the baseline years 1981–2010 at Nossen were selected as illustration, and two years with anomalous weather were identified for closer scrutiny based on an analysis of these 30 years.

## 3. Results

### 3.1. IRS analysis: classifying period mean responses

The full set of impact response surfaces (IRSs) of 30-year mean yield changes relative to the baseline from all model simulations is shown in Supplement 2, Fig. S3 for both spring and winter wheat at the different

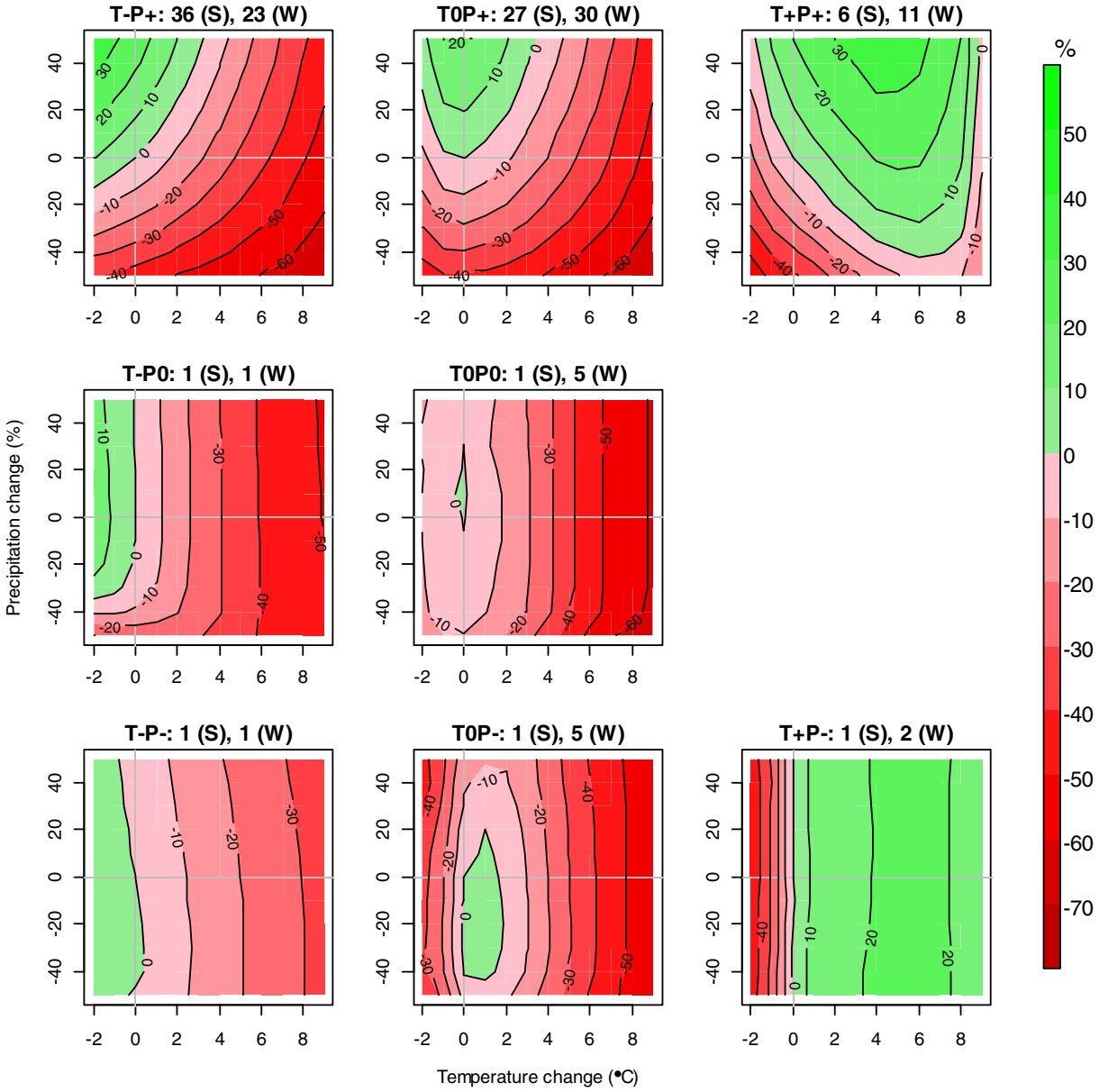


**Fig. 3.** IRS patterns of winter wheat yield change relative to the baseline (%) averaged across members of classes defined according to the location of maximum yield. Panels are organised with the location of maximum yield occurring at cooler (T−), similar (T0) or warmer (T+) temperatures than the baseline in columns from left to right, and at wetter (P+), similar (P0) or drier (P−) conditions than the baseline in rows from top to bottom. No model response was classified as T+P0. Frequencies of IRSs falling in each class for spring (S) and winter (W) wheat are shown on top of each plot. See text for a definition of the classes.

sites. These (excluding the ensemble median IRS) were classified using the expert and statistical diagnostic approaches (EDA and SDA – Subsection 2.2) and the results of the two approaches compared.

### 3.1.1. EDA classes

The location of the maximum yield varies among these IRS plots, both with respect to temperature and to precipitation. For winter wheat there are representatives of eight out of the nine possible classes depicted in Fig. 1b, with average patterns across each set shown in Fig. 3.

For an IRS with the maximum yield located at a condition that is cooler and wetter than the baseline (T−P+, Fig. 3, top-left), yield increased from the baseline towards the maximum, whereas yield decreases were found for both drier and warmer conditions. The IRS with its yield maximum close to the baseline temperature, but for increased precipitation (T0P+, Fig. 3, top-centre) has a U-shaped pattern of response, as does the T+P+ IRS, but it is shifted further to the warm side. Similarly, the patterns with the maximum yield close to the baseline precipitation (T−P0 and T0P0, Fig. 3, middle-row) are shifted towards drier conditions. The IRSs with maximum yield at drier conditions (P−, Fig. 3, bottom row) have patterns that are not U-shaped but mainly influenced by changes in temperature (T−P−, bottom-left, and T+P−, bottom-right) or show yield decreases with increases in precipitation (T0P−, bottom-centre). There was no ensemble member producing an IRS with maximum yield at warmer conditions than the baseline and with precipitation close to baseline conditions (T+P0).

The four classes with different response strengths described in Fig. 1b are illustrated in Fig. 4 by averaging across all representatives of each class for winter wheat. Patterns of response are similar in these four plots, with maximum yield located close to baseline temperature and for increased precipitation. However, the magnitude of yield change differs between the weak and strong responses. The greatest yield decline exceeds 70% in the strong response plot (T"P", Fig. 4, bottom-right), but is only slightly above 40% in the weak response plot (T'P', top-left).

The frequency of model simulations falling in different EDA classes is shown in Supplement 3, Fig. S4, across all sites and for each crop separately. Out of a total number of 36 possible combinations of classes defined by the location of the maximum yield (9) and the strength of response (4), examples of 16 were identified for spring wheat and 19 for winter wheat. For both spring and winter wheat the majority of model

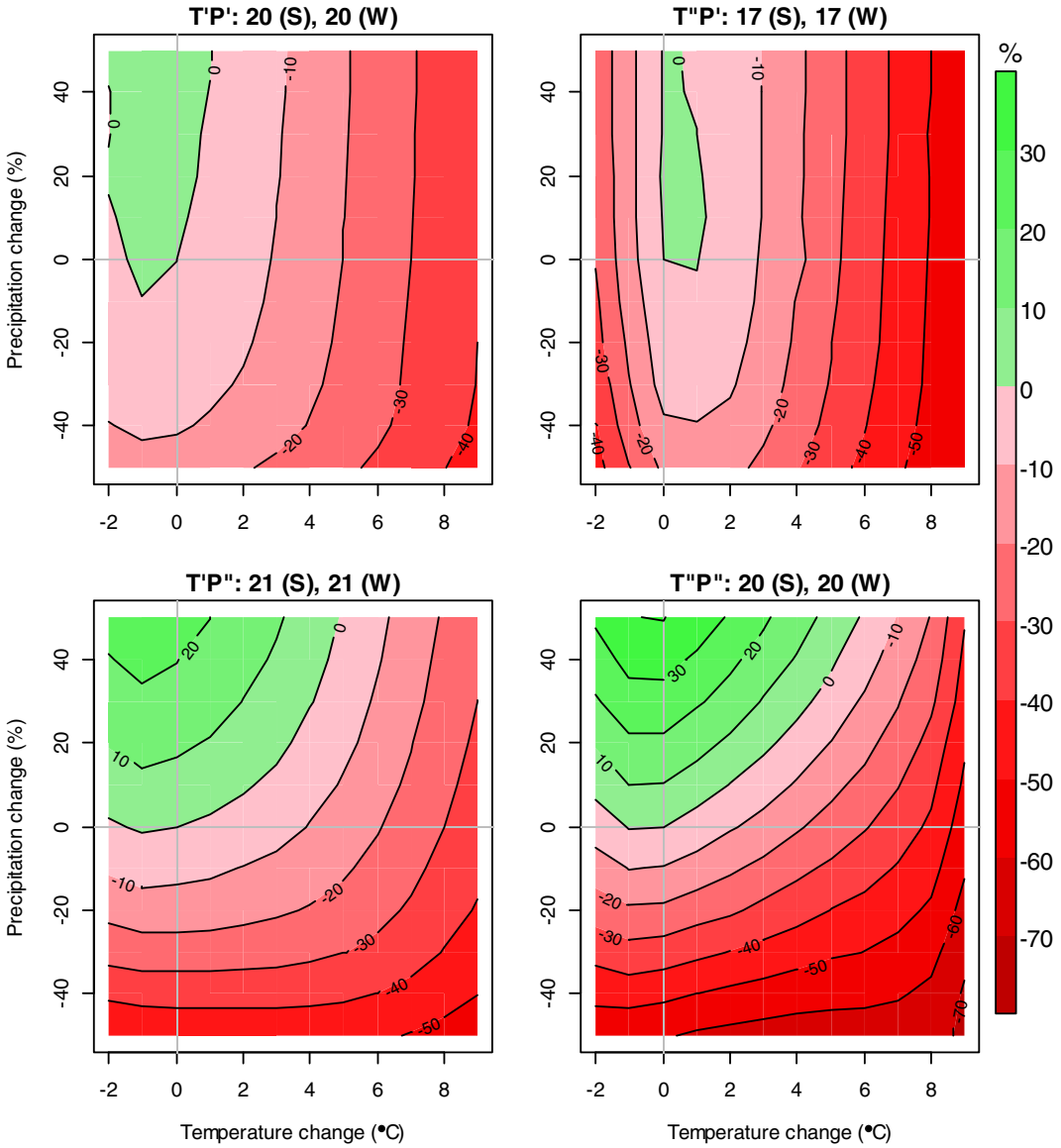


**Fig. 4.** IRS patterns of winter wheat yield change relative to the baseline (%) averaged across members of classes defined according to the strength of response. Weak (strong) responses to precipitation changes are shown in the top (bottom) row and weak (strong) responses to temperature changes in the left (right) column. Frequencies of IRSs falling in each class for spring (S) and winter (W) wheat are shown on top of each plot. See text for a definition of the classes.

simulations had their maximum yield located at wetter conditions (P+) and at cooler (T−) or similar temperatures (T0) than the baseline. For spring wheat, the location of maximum yield was at the cooler side of the baseline (T−) for nearly all ensemble members at the German site and close to the baseline (T0) at the Finnish site. The distribution between T− and T0 was more even in Spain and for winter wheat in Germany, whereas winter wheat in Finland had its maximum yield close to the baseline temperature for the majority of ensemble members. The location of maximum yield relative to the baseline precipitation was at drier (P−) or similar (P0) conditions for only five ensemble members for spring wheat and 14 for winter wheat.

The distribution of ensemble members for the strength of response were, by design of the classes, relatively evenly distributed among the strong and weak response-classes for all three sites combined (Supplement 1, S1.1). For the Finnish site, most model simulations had a strong temperature response (T"), whereas the German and Spanish sites showed a more even distribution between strong (T") and weak (T') temperature response (Fig. S4). The strength of response to changes in precipitation for winter wheat was predominantly weak in Finland and strong in Spain, whereas it was more evenly distributed for the German winter wheat site and for all sites for spring wheat.

### 3.1.2. SDA clusters and their relationship to EDA classes

The classification based on the statistical diagnostic approach (SDA) used cluster analysis to order and group crop model responses along the branches of a tree in a dendrogram (cf. Fig. 2). Eight clusters were identified for each crop and, as for the EDA classes depicted in Figs. 3 and 4, the IRS patterns have been averaged across each cluster in Fig. 5. Here both spring and winter wheat are shown, with the number of members of each cluster given in parentheses. Unlike the EDA classes, to which IRSs are allocated according to pre-defined pattern characteristics, the classes here are defined by the clustering algorithm, with dominant patterns allocated low cluster numbers and more deviant patterns allocated higher numbers. The clusters for spring and winter wheat IRSs were determined independently, so the order of the clusters is not directly comparable between the two crops. For example, the two largest clusters (C1 and C2) are quite similar for both crops, with a yield maximum near to baseline temperatures and above baseline precipitation, while C1 shows a strong temperature response and C2 a strong precipitation response. In contrast, the yield maximum for C3 is found under strong warming for winter wheat but under baseline temperatures for spring wheat, though both crops show a strong negative response to cooling. Note also that some of the clusters contain only one member, so the pattern shown in Fig. 5 simply mimics the 30-year mean IRS of one of the model simulations shown in Supplement 2, Fig. S3.

The two largest clusters obtained using this approach contained 55 of the 74 ensemble members for spring wheat and 63 of the 78 members for winter wheat (Fig. 6, column n). The two largest spring wheat clusters for the German and Spanish sites (clusters 1 and 2) included IRSs with the maximum yield located to the cooler side or close to baseline temperature and for increased precipitation. These differed in the strength of precipitation response, with cluster 1 mainly containing members with a weak response and cluster 2 with a strong response to precipitation changes. For spring wheat, most members (38%) for the Finnish site were found in cluster 3, which had no representatives for the sites in Germany and Spain (Fig. 6). Here, maximum yields were found for warmer than (T+) or close to (T0) baseline temperatures and wetter than (P+) the baseline, with strong responses to temperature change (T"). Compared to spring wheat, the three largest clusters for winter wheat had a similar relation to the classes of the location of maximum yield and the strength of response, albeit with more deviations from those descriptions and fewer members in cluster 3 for the Finnish site. The remaining clusters 4 to 8, which diverged from three major patterns, contained few (up to 4) cases for both spring and winter wheat.

## 3.2. Relationship between yield responses and model properties

The two classification systems were used to investigate whether different patterns of yield response could be related to different properties of the crop models. Fig. 7 summarises these comparisons for different types of property and for selected SDA clusters and EDA classes.

An overall observation when comparing dominant IRS patterns for the two crops is that response patterns for spring wheat appear to reflect greater sensitivity to temperature change than to precipitation change (predominant T″ class with red/purple icons), whereas sensitivity to precipitation change is more prominent for winter wheat (more representatives of P″ with blue/purple icons). However, this general observation masks deviations at some sites and for certain types of models. These are discussed in the following subsections.

To aid that discussion, the distance metric used in the clustering procedure is also presented in Fig. 7 to indicate the degree of similarity (proximity) between IRS patterns in a given group. In order to distinguish more readily those model properties having similar responses from those that are more dissimilar, Fig. 8 shows distance ratios that

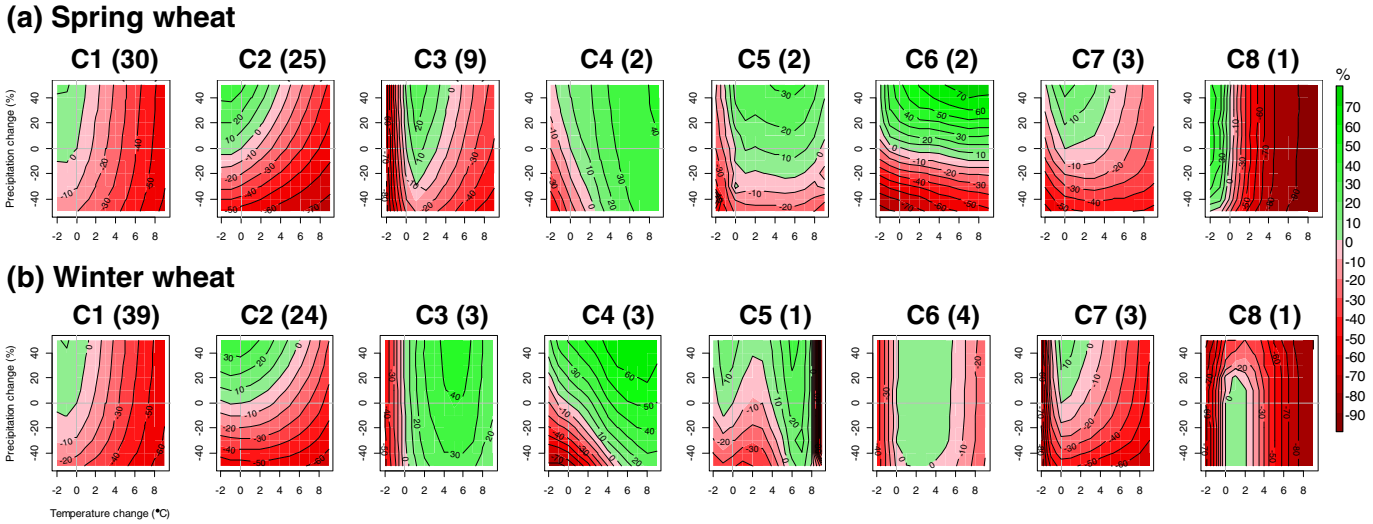


**Fig. 5.** IRS patterns of grain yield changes relative to the baseline (%) averaged across all members of clusters determined using the statistical diagnostic approach (SDA) for spring wheat (top row) and winter wheat (bottom row). A cluster identifier is indicated above each plot with the number of cluster members in parentheses. For details, see text.

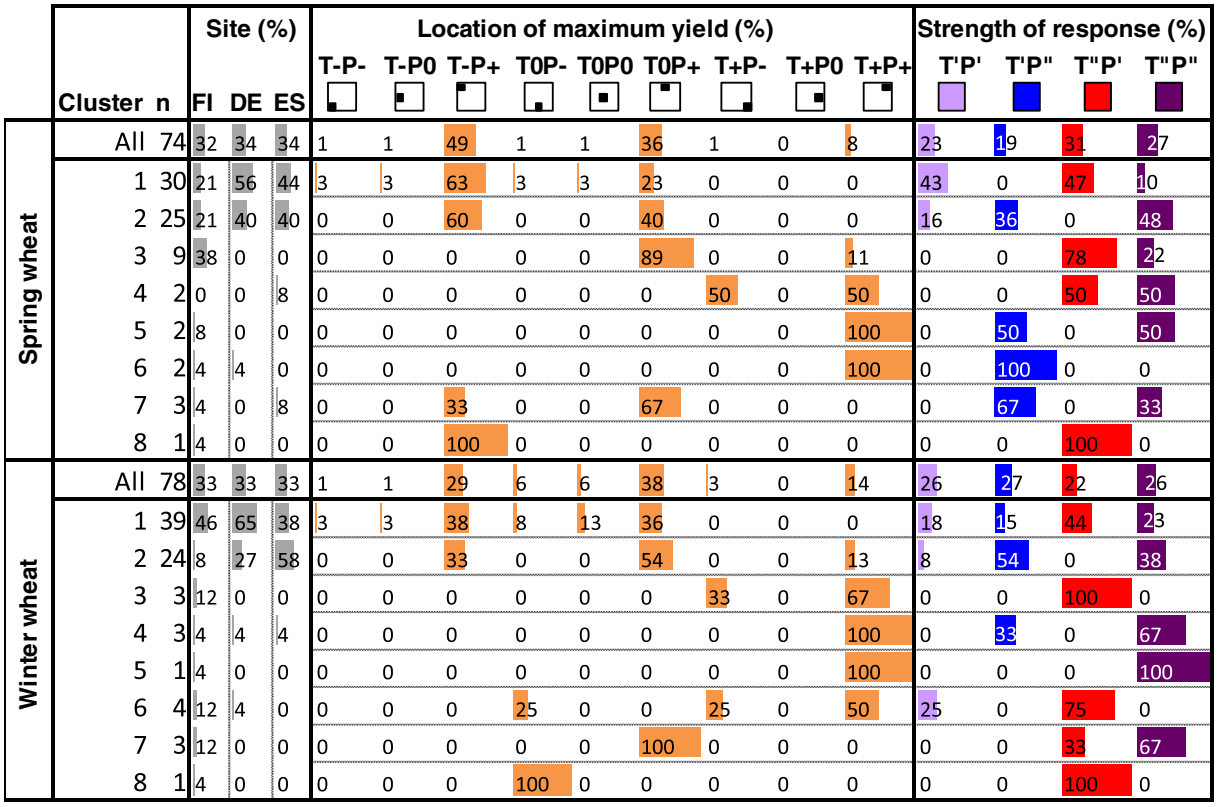


**Fig. 6.** Proportion of IRSs of each SDA cluster among the sites and the classes defining the location of maximum yield and the strength of response for spring and winter wheat. Larger percentages are indicated with proportionally wider horizontal bars. T0 denotes a location of maximum yield within 1 °C of the baseline temperature; for T− (T+) the location is outside this range. P0 denotes a location of maximum yield within 10% of the baseline precipitation; P− and P+ are outside this range. The strength of response is defined with respect to temperature (weak = T', strong = T") and precipitation (weak = P', strong = P"). Modelling sites are denoted as FI for Finland, DE for Germany and ES for Spain. Cells delineated by solid or dashed lines sum up to 100%.

have been ranked and presented separately for groups defined by site, genealogy and calibration (upper panel) and by process description (lower panel). The ranking in Fig. 8 is of distance ratios for winter wheat, with spring wheat ratios for equivalent property types also shown for comparison.

### 3.2.1. Model genealogy

Models that share a common genealogy were grouped into model families (see Bouman et al., 1996; Rosenzweig et al., 2014; van Ittersum et al., 2003). These showed smaller values of the distance metric than random samples for all five model families identified for spring wheat and three of the five families for winter wheat (Fig. 7). In all cases, family members shared dominant IRS clusters across both crops, while classes defining the location of the maximum yield were also fairly consistent between crops, though the strength of response usually differs between dominant spring and winter wheat classes. Note that the same model independently calibrated and used by different researchers showed similar responses for the WOFOST and CERES v4.5 models (not shown).

### 3.2.2. Model calibration

IRSs of models calibrated using automatic methods displayed patterns for both crops that differed more than would be expected for a random sample (positive distance metric), while models calibrated using manual methods showed similar patterns of response for winter wheat but not spring wheat (Fig. 7). Similarity in IRS patterns was observed among models for which few parameters were tested (≤2 for spring wheat; ≤5 for winter wheat; Fig. 8).

### 3.2.3. Process descriptions

Distance ratios among models were used to distinguish those process descriptions giving similar IRS patterns. Models using the Priestley-Taylor equation to simulate evapotranspiration show similar IRS patterns for both spring and winter wheat. Similar patterns are also found among models using the Penman-Monteith equation for spring wheat and the Penman equation for winter wheat (Figs. 7 and 8). Ratios among the few models that used descriptions of the water dynamics solving the Richards equation alone or as part of a mixed approach indicated similar patterns of yield response for both crops, whereas responses from the majority of models that relied on the capacity approach were more dissimilar. Models with simpler process descriptions for the root distribution (linear) and water stress (only one process) showed greater similarity in IRS patterns (smaller distance ratios) than models with more complex descriptions. Models simulating heat stress processes on the vegetative organs gave most similar responses, whereas those representing heat processes on the reproductive organs and on phenology (for spring wheat) were dissimilar (Figs. 7 and 8).

Those models that applied dry matter partitioning schemes to simulate yield formation and some others that estimated grain number and biomass were found to produce similar IRS patterns for both crops (low distance ratios). Models that made use of a modified harvest index exhibited large differences in patterns of response in both crops, while those that used an unmodified harvest index showed large differences for spring wheat but similarities for winter wheat (Figs. 7 and 8).

### 3.2.4. Site influence on response patterns

IRS patterns across all model simulations resembled each other more closely than for random samples (low distance ratios) for the German sites (both crops) and the Spanish site (winter wheat). For these

| | | Spring wheat | | | | | Winter wheat | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | SDA | | EDA classes | | | SDA | | EDA classes | |
| Model property | | n | cluster | distance | Ymax loc | Strength | n | cluster | distance | Ymax loc | Strength |
| Location | FI | 24 | C3 | + | ■ | ■ | 26 | C1 | ++ | ■ | ■ |
| | DE | 25 | C1/C2 | -- | ■ | ■ | 26 | C1 | - | ■ | ■ |
| | ES | 25 | C1/C2 | + | ■■ | × | 26 | C1/C2 | - | ■■ | ■■ |
| ***Genealogy:*** | C | 24 | C1 | - | ■■ | × | 24 | C1 | - | ■ | × |
| | L | 6 | C2 | -- | ■ | ■ | 6 | C2 | + | × | ■ |
| | E | 6 | C2 | - | ■■ | ■ | 6 | C2 | - | ■■ | ■ |
| Family | S | 15 | C1 | - | ■ | × | 15 | C1 | -- | ■ | ■ |
| | I | 6 | C2 | - | ■■ | ■ | 6 | C2 | + | ■ | ■ |
| | - | 23 | C1 | + | ■ | ■ | 27 | C1 | + | × | ■ |
| ***Calibration:*** | A | 14 | C1 | + | ■ | ■ | 15 | C1 | + | × | ■ |
| Approach | M | 48 | C2 | + | ■■ | ■ | 48 | C1 | - | ■ | × |
| Number of parameters | 0-2 | 6 | C1 | - | ■ | ■ | 9 | C1 | -- | ■ | × |
| | 3-5 | 21 | C1 | + | ■ | × | 21 | C1 | - | ■ | ■ |
| | 6-9 | 21 | C1 | + | ■ | ■ | 21 | C2 | ++ | × | × |
| | >9 | 9 | C2 | + | ■ | × | 9 | C2 | ++ | × | ■ |
| ***Processes:*** | P | 21 | C1 | + | ■ | ■ | 21 | C1 | - | ■ | ■ |
| | PM | 20 | C2 | - | ■■ | × | 21 | C1 | + | ■ | ■ |
| Evapo-transpiration | PT | 30 | C1 | - | ■■ | ■ | 30 | C1 | - | ■ | × |
| | M | 0 | | | | | 3 | C1 | -- | × | ■ |
| | SW | 3 | C1 | - | ■ | ■ | 3 | × | + | × | × |
| Water dynamics | C | 60 | C2/C3 | + | ■■ | × | 63 | C2 | + | ■ | × |
| | R | 5 | C1/C2 | -- | ■■ | ■ | 6 | C1/C2 | -- | ■ | ■ |
| | C/R | 9 | C1 | -- | ■ | ■ | 9 | C1 | - | ■ | ■ |
| Root distribution | L | 18 | C1 | -- | ■ | ■ | 18 | C1 | -- | ■ | × |
| | E | 39 | C2 | + | ■■ | × | 42 | C1/C2 | - | ■ | × |
| | O | 15 | C1 | + | ■■ | ■ | 15 | × | ++ | × | × |
| Water stress | E | 29 | C1 | -- | ■■ | × | 30 | C1 | -- | ■ | × |
| | S | 9 | C2 | -- | ■■ | ■ | 9 | C1/C2 | - | ■ | ■ |
| | E/S | 36 | C1 | + | ■ | × | 39 | C4 | + | × | × |
| Heat stress | V | 12 | C1 | -- | ■■ | ■ | 12 | C1 | -- | ■ | ■■ |
| | R | 6 | × | + | ■ | ■ | 6 | C4 | ++ | ■ | ■ |
| | P | 9 | × | ++ | ■ | ■■ | 9 | C1 | - | ■ | ■ |
| | M | 12 | C2 | - | ■ | ■ | 12 | C2 | + | ■ | ■ |
| Yield formation | P | 44 | C1/C2 | - | ■■ | ■ | 45 | C1/C2 | - | ■ | × |
| | G | 6 | C1 | - | ■ | ■ | 6 | C1 | - | × | ■ |
| | H* | 12 | C2 | + | ■ | × | 12 | × | ++ | × | ■ |
| | H | 12 | × | ++ | × | × | 15 | C1 | - | × | ■ |

**Fig. 7.** Relating model properties to IRS classes for simulations with spring and winter wheat. From left to right columns show: property type; property class (locations: FI (Finland), DE (Germany), ES (Spain), for other codes, see Table 1 footnotes); five columns for spring wheat: number of models (n); statistical diagnostic approach (SDA) clusters with frequency above 33.3% and distance metric (relative to mean distance of 100 randomly selected IRS groups of the same size): >150%(++), 100%–150%(+), 50%–100% (-), ≤50% (–); classes with frequency above 33.3% from the expert diagnostic approach (EDA – see Fig. 1 for explanation) for location of maximum yield (Ymax loc); ditto for strength of response (Strength): × no class/cluster >33.3% frequency; equivalent five columns for winter wheat. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

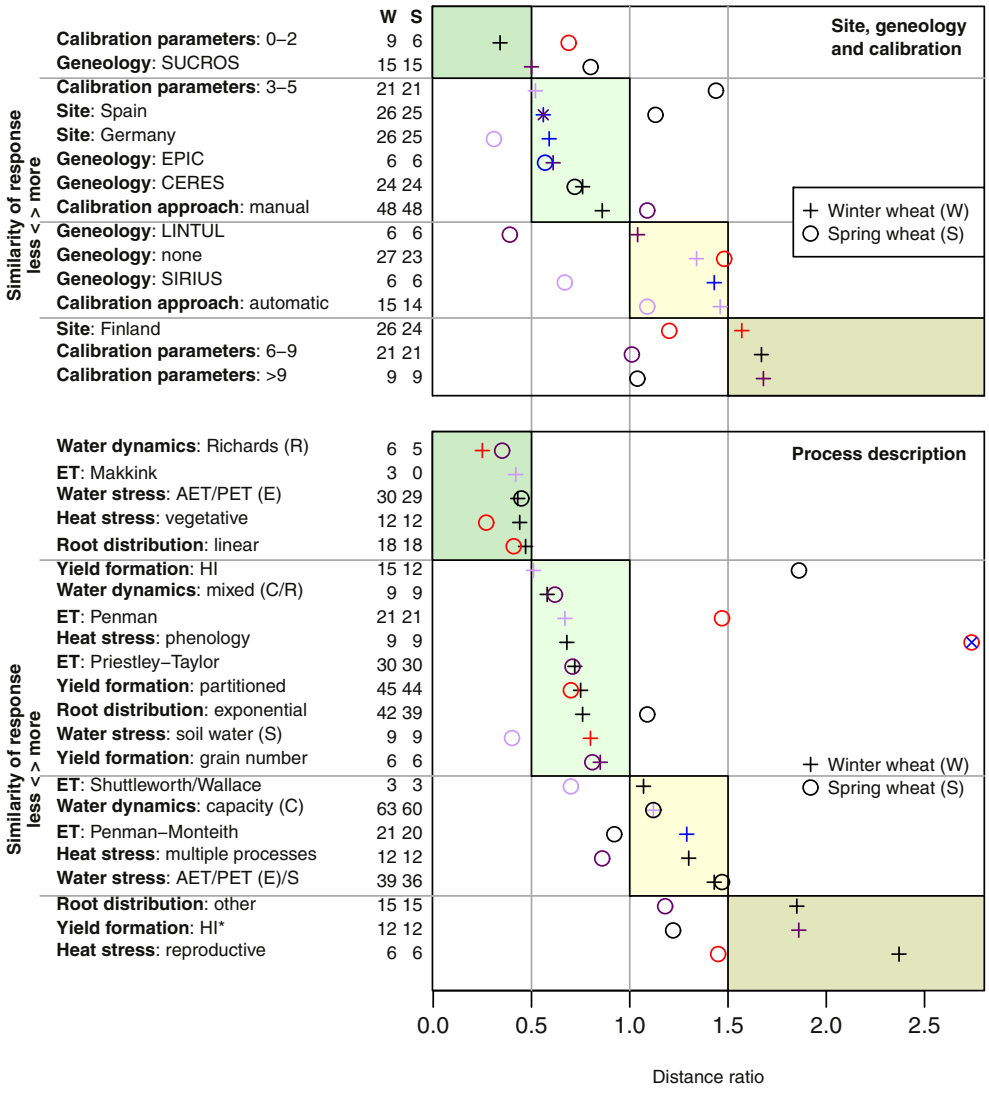


**Fig. 8.** Ratios of the mean distance metric of the statistical diagnostic approach (SDA – see text) comparing impact response surface patterns of all pairs found in groups of models having the same property type (cf. Table 1), to the metric computed for pairs randomly selected from all IRSs into 100 groups of the same size. Ratios are ordered from more similar to less similar groups for winter wheat (+) for groups defined by site, genealogy and calibration (top) and by process descriptions (bottom). Ratios are also shown for the same groups for spring wheat (circles). Group sizes are given as W for winter and S for spring wheat. The symbol colour indicates the most frequent class(es) of strength of response with frequency above 33.3% from the expert diagnostic approach (EDA – see Fig. 1 for explanation); a cross means that two classes are more frequent than 33.3%. Lines and shaded rectangles delineate model groups with IRSs that are very similar (distance ratio ≤0.5, green shading), more similar than random (0.5–1.0, light green), more dissimilar than random (1.0–1.5, light yellow) and very dissimilar (>1.5, brown). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cases, there was relative consistency in the dominant strength of response: at Nossen both temperature and precipitation responses were weak for spring wheat, whereas for winter wheat at Dikopshof and Lleida, precipitation responses were strong. On the other hand, patterns were more different than if selected at random for the Finnish site (both crops) and spring wheat for Spain (Fig. 8). The Finnish site showed strong yield sensitivity to temperature (T″), but high distance ratios. This was because two different patterns were important, one reflecting strong responses to increased temperature (i.e. C1, the dominant cluster for winter wheat – Fig. 7 and cf. Fig. 5b) and the other strong responses to cooling including frequent crop failure (i.e. C3, the dominant cluster for spring wheat, cf. Fig. 5a), with some other patterns also contributing. A higher distance ratio across spring wheat IRSs for the Spanish site (Fig. 8) can be explained by the even mix of clusters C1 and C2 (Fig. 7), which depicted patterns of strong yield response to temperature and to precipitation, respectively (cf. Fig. 5a).

Another way of comparing models is to examine the response patterns of individual models across the three sites (Fig. S5). Distance metrics between three IRSs varied between models and crop variety from nearly identical IRSs for some models to large differences between the three sites for others. 17 of the 26 winter wheat models gave IRSs for the three sites that were more similar to each other than IRS triplets that were randomly selected from the full ensemble. For spring wheat, this was the case for 14 of 25 models. 11 models showed closer IRSs than random triplets for both spring and winter wheat, whereas for six models IRSs differed for both spring and winter wheat.

### 3.3. Responses in anomalous years

1995 and 2003 were selected as examples of wet and dry years at the German spring wheat site, Nossen. During the wet year of 1995, growing season (April–September) precipitation was 68% above the 1981–2010 mean. In contrast, during the very dry year of 2003, monthly precipitation totals were below the long-term mean in all months except January and November, with the deviation from the long-term mean during summer (June–July) the greatest in the 30-year period. Additionally, regional yield statistics showed close to average yields in 1995 but anomalously low yields in 2003 (cf. Fig. S1 in Pirttioja et al., 2015). IRSs for spring wheat from all model simulations for both years are shown in Fig. S6. The IRSs of several models have denser contour lines for 2003 with a wider range of yield changes across the IRS space compared to 1995. This can also be seen in the counts of the strength of yield responses using the EDA classification, which are presented for each year alongside the 30-year mean in Fig. 9. Compared to the long-term average, this shows a clear shift towards an increased number of model simulations showing weak responses to both temperature and precipitation for the wet year 1995, and conversely an increased number showing a strong response to both for the dry year 2003.

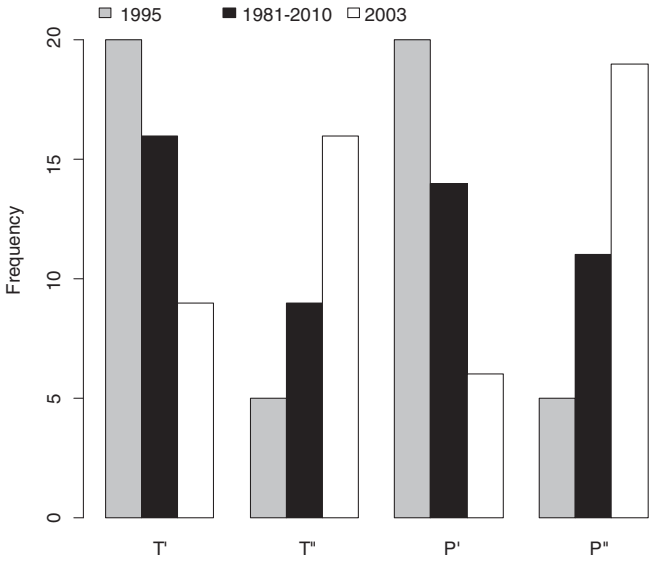


**Fig. 9.** Number of crop models showing weak (T' and P') and strong (T" and P") yield responses to respective perturbations of temperature and precipitation for spring wheat in Nossen (Germany) for the 1981–2010 period mean and for wet (1995) and dry (2003) years relative to their respective baseline yields.

## 4. Discussion and conclusions

The preceding results demonstrate the application of two approaches that offer alternative, complementary methods for classifying IRS patterns of modelled yield response to climate. In addition, an attempt has been made to relate classes of yield response to characteristics of the model simulations. Here we discuss possible explanations for differences in patterns revealed by the analysis and offer additional observations on the potential utility of the classification approaches.

### *4.1. Explaining IRS patterns of wheat yield response to climate*

The majority of IRS patterns for ensemble mean yields reported in Pirttioja et al. (2015) could be found in only a few of the 36 classes defined by the expert diagnostic approach (EDA). Contour lines were roughly U-shaped with increases in yield towards the maximum located at increased precipitation; the maximum yield appeared to the cooler side of the baseline for the German and Spanish sites and close to or to the warmer side of the baseline for the Finnish site. The differences for these main classes of IRSs were in the strength of response to changes in temperature and precipitation.

The exceptions to these main response types were IRSs with their maximum close to or below the baseline precipitation, which were in all cases also classified to have a weak precipitation response. These included IRSs with o-shaped contours with a yield decrease for increases in precipitation (T0P− in Fig. 3), found only for winter wheat at the Finnish site. For one model, CARAIB, in which cloud cover is inferred from precipitation, circumstances can arise whereby increased precipitation restricts crop photosynthesis through enhanced cloud cover and reduced radiation receipt, even where water may otherwise not be limiting for growth. Similar model behaviour could also be caused by the effect of water-logging or by leaching of nutrients when too much water is available (Geerts and Raes, 2009), processes not simulated by most crop models. One exception is the DNDC model, which has been used to simulate soil carbon processes that are strongly affected by water table dynamics (Giltrap et al., 2010; Kröbel et al., 2010). Other exceptional IRSs showed very weak response to changes in precipitation, under which the exact location of the maximum yield along the precipitation axis might be based on very small differences in yield. Here, the representation of soil and crop growth processes and their parameter calibrations might be an explanation for a model's relative insensitivity to changes in precipitation. Indeed, for cases in which model simulations show a very weak precipitation sensitivity at sites where water limitation is known to be a dominant constraint on yield, there would be strong grounds for discarding model results as implausible, as demonstrated at Lleida by Castañeda-Vera et al. (2015) and Ruiz-Ramos et al. (2017) in a related IRS modelling study for winter wheat.

The similarity of response patterns among members of model groups (cf. Fig. 8) gave an indication of which model properties are influential in determining responses. Membership of the same model family (i.e. models with a shared genealogy) tended to produce similar response patterns. These models share some model principles and components, but also differ in others. Documenting how models of the same family differ from one another is a tedious task, as information is scattered or not published as well as difficult to compare between different models. One of the precepts for applying multi-model ensembles to represent uncertainties in model estimates is that an ensemble should encompass a diversity of model behaviour, and the “degree of relatedness” of models has been advocated as one possible criterion for determining model choice (Wallach et al., 2016). In this context, it would seem all the more important to improve and further develop existing crop model genealogies that were the basis for our analysis (Asseng et al., 2013; Rosenzweig et al., 2014) so as to encompass a larger set of models and include descriptions of how the models differ.

The method used to describe evapotranspiration in a crop model also helped to explain differences in yield response patterns. These approaches differ in their weather input data requirements as well as in their complexity, accuracy and intended applicability to different environmental conditions. Of the three most common equations used in the ensemble of wheat models, Priestley-Taylor has the simplest description and only requires information on temperature and solar radiation, whereas Penman-Monteith requires wind speed and relative humidity in addition, and Penman wind speed and air pressure. Further investigation would be needed to discriminate between the effects of different evapotranspiration schemes on a model-by-model basis, see for example a comparison for maize using a single model (Webber et al.,

**Table 3**
Some features of the expert and statistical diagnostic approaches presented in this study.

| | Classification approach | |
|---|---|---|
| | Expert diagnostic (EDA) | Statistical diagnostic (SDA) |
| Diagnostics | • Location of maximum response (1)<br>• Rate of change of response (2)<br>• Combined (1 & 2) | • Pattern correlation (1)<br>• Euclidean distance (2)<br>• Distance metric (combines 1 & 2) |
| Typical features | • Related to system-relevant criteria<br>• Applicable independently to other studies and ensembles<br>• Class members can be averaged<br>• Subjective criteria for classification<br>• No metric readily available to compare patterns | • Related to IRS pattern alone<br>• Constrained to available ensemble members<br>• Class members can be averaged<br>• Largely statistical criteria for classification<br>• Distance ratio to compare patterns |

2016). However, other work suggests that a focus on evapotranspiration schemes alone may not offer a sufficient explanation for simulated model behaviour. For instance, Cammarano et al. (2016) analysed the temperature sensitivity of crop water use simulated by 26 wheat models and found that the main source of uncertainty was due to model differences in the partitioning between soil evaporation and crop transpiration, rather than the choice of the evapotranspiration formula itself.

The analysis of models using the same root distribution function and water stress processes provided examples where models with simpler model descriptions exhibited more similar IRS patterns than those with more complex process descriptions. While this can be seen as a tendency that more complex model descriptions can result in more complex results, in order to generalize these two examples one would ideally expand the analysis with further descriptions of model properties, such as the number of parameters used in a model or a classification of the ensemble members into different categories of model complexity.

The automatic and manual approaches to calibration were not ideal descriptors of the detailed procedures actually followed, but the information provided by modellers on calibration was patchy and incomplete and grouping was problematic. It is perhaps to be expected that those models for which calibration focused on only a few parameters (commonly relating to phenology) were those showing greatest similarity of response (cf. section 3.2.2). The effect of parameter selection on yield outcomes can be large (Angulo et al., 2013), which also becomes evident from the large distances between IRSs of the same model calibrated for different sites (e.g. model 1 in Fig. S5). Hence, the descriptions of the calibration approach that we were able to collect may be insufficient to address the full effects of calibration strategies on the yield response.

### 4.2. Assessment of the classification approaches

This paper has introduced and applied two diagnostic approaches to the classification of IRS patterns generated in a multi-model sensitivity analysis of simulated wheat yields to temperature and precipitation perturbations. Characteristics of the expert (EDA) and statistical (SDA) diagnostic approaches are summarised in Table 3 and outlined below.

#### 4.2.1. Utility of the expert diagnostic approach (EDA)

The expert diagnostic approach (EDA) classifies patterns of response using knowledge about the circumstances of the model simulations (e.g. types of models and processes simulated, calibration, site conditions, input data and other assumptions). Here, IRS patterns are placed in classes defined using criteria pre-specified by experts.

One advantage of this approach is that any IRS pattern can be readily described using criteria that relate directly to features of the system being simulated. In this paper, the pattern of wheat yield sensitivity is described in terms of the position on the IRS of the maximum yield and the rate of yield decline from this maximum with respect to temperature and precipitation change. The same criteria can be used for different sites, different crops and patterns for period-averages as well as individual years (cf. Sub-section 3.3). The EDA could also be applied to other climate change sensitivity studies applying IRS for different sectors or systems (e.g. hydrology, fisheries, forestry, human health), though different classification criteria would need to be identified for a given system. One such example is a classification of IRSs of water availability for 18 European river catchments simulated with a hydrological model (Weiß and Alcamo, 2011).

The EDA classification proved useful for describing model responses in two anomalous years for the German spring wheat site (Sub-section 3.3). It demonstrated how the sensitivity of modelled responses can be strongly influenced by the nature of the baseline climate – yields displayed stronger relative responses under anomalously dry than anomalously wet baseline conditions. Other sites and years (not reported) showed different deviations. For example, responses to cool years in Finland were sometimes characterised by crop failure, which can complicate the interpretation. A systematic analysis of IRS patterns for individual years using the EDA classification might be an alternative approach for describing aspects of inter-annual yield variability, complementing measures such as the coefficient of variation that are typically used in crop model studies. This could be a topic for future research.

A potential disadvantage of the EDA is that the classification criteria are chosen subjectively, so different studies might adopt different criteria, making inter-comparison of results problematic. Another is that very similar patterns may fall in different classes by virtue of falling on either side of a threshold defined for a given selection criterion, without their similarity being readily quantifiable.

#### 4.2.2. Utility of the statistical diagnostic approach (SDA)

The statistical diagnostic approach (SDA) groups patterns of response into classes based on statistical measures of their similarity. Here, the similarity between IRS patterns and magnitudes is evaluated without attempting to interpret these features. IRSs can then be classified using a clustering algorithm. One advantage of this approach is its ease of application, requiring only the values describing each IRS pattern and statistical software for computing the relevant metrics (i.e. pattern correlation, Euclidean distance and the clustering procedure for this study), though subjective choices are required for clustering.

While we have plotted average IRS patterns across members of individual clusters in Fig. 5, such an analysis is not a pre-requisite for classification. However, if this is the only approach used to classify responses, then it may become necessary to provide labels to describe patterns represented in each of the clusters. For example, hierarchical clustering of Euclidean distances was used by Prudhomme et al. (2013a) to classify IRSs of river floods simulated with a hydrological model. They plotted average patterns of response for each cluster, labelling and relating each one statistically to different characteristics of the river catchments (analogous to the crop model properties examined in the present study). In further work, they used these statistical relationships to infer likely flood responses to climate change in other catchments that were not part of the original analysis (Prudhomme et al., 2013b).

The distance metric that is used for clustering the IRS patterns can also be applied separately to indicate relative similarity in the patterns presented by different IRSs (noted as a weakness of the EDA approach). A disadvantage of this approach is that it applies solely to the ensemble of IRS patterns analysed. Thus evaluations of different ensembles, perhaps from other studies, would result in clusters that are not equivalent (as demonstrated for spring and winter wheat in Fig. 5). The choice of ensemble size to be analysed depends on what aspects of the model simulations are of interest. For instance, in this study it was found that IRS patterns of yield response differed between sites, so for a strict inter-comparison of model behaviour, clusters should have been estimated separately for each site. However, this would have resulted in smaller sample sizes of the IRS clusters. Finally, it would only be possible to include a new IRS pattern by computing the distance metric between the new member and all others in the ensemble and/or repeating the clustering procedure.

#### 4.2.3. Relating EDA to SDA patterns and their sensitivity to modelling assumptions

There are clear overlaps between dominant patterns described using the statistical and expert diagnostic approaches, even if the approaches themselves differ. A strong correlation between two IRSs (Supplement 1, Eq. (1)) indicates that these have a similar pattern, a finding analogous to identifying the location of maximum yield for the EDA, which is a specific aspect of the IRS pattern. Likewise, the Euclidian distance bears similarities to the strength of response criterion used in the EDA. Thus, it is not surprising to find correspondences between dominant SDA clusters and EDA classes, although no exact matches were

identified (cf. Fig. 6).

Model results were obtained following a protocol that necessarily included a number of simplifying assumptions. For instance, a synthetic soil was assumed with parameters common for all sites. Simulated crop yields can be sensitive to soil properties and status in most models (e.g. Varella et al., 2010), so had the analysis assumed a soil with a lower water holding capacity, simulated responses to changes in precipitation could be expected to have been stronger, thus affecting the outcome of the EDA and SDA classifications. Another example is the management assumption of a constant sowing date for baseline and perturbed climates. Had sowing dates been shifted to match the altered growing conditions, modelled yield responses are likely to have been weaker than those reported (Ruiz-Ramos et al., 2017).

### 4.3. Conclusions

In this paper two new approaches have been presented for classifying impact response surfaces (IRSs) used to depict the modelled sensitivity of crop yields to perturbed climate. We hypothesized that such classifications might offer insights into differences in model behaviour under changed climate and on the possible reasons for these differences. To examine this, we have applied the classification approaches to IRSs generated in a multi-model sensitivity analysis of simulated wheat yields to temperature and precipitation perturbations.

Based on the analysis presented in the paper, we conclude that:

1. The expert diagnostic approach (EDA) and statistical diagnostic approach (SDA) offer alternative, complementary and in some cases overlapping methods for classifying and discriminating between patterns of modelled response to temperature and precipitation.
2. EDA provides a useful means of describing different patterns of response in terms directly relevant to a given impact (e.g. the strength of yield response to changing temperature and precipitation) and is directly applicable in other similar studies.
3. SDA offers straightforward procedures for comparing patterns of response based on statistical measures, including a distance metric that quantifies the similarity between two patterns, though its application is constrained to those ensemble members under study.
4. There are multiple options for presenting and inter-comparing IRS classes, including comprehensive tables with frequency counts of all classes, cross tabulation of different classification methods, counts of dominant classes, distance metrics, and use of coloured icons for rapid visual comparison.
5. Complex models represent multiple processes in different ways, thus no single model property across a large model ensemble was found (or could realistically be expected) to explain the integrated yield response to temperature and precipitation perturbations. However, application of the EDA and SDA approaches to classify IRS patterns in a large wheat model ensemble for sites in Europe and attempts to relate these classes to characteristics of the model simulations revealed some capability to distinguish:

- a stronger response to precipitation perturbations for winter wheat than spring wheat;
- differing strengths of response to both temperature and precipitation changes in anomalous weather-years compared to period-average conditions;
- the effect of site conditions on IRS yield patterns, which showed clear differences between sites in the location of the maximum yield and strength of decline from that maximum with climate change;
- similarities in IRS patterns among models with related genealogy;
- greater similarity in IRS patterns for models with simpler process descriptions for the root distribution and water stress than for models with more complex descriptions;
- a closer correspondence of IRS patterns in models using partitioning schemes to represent yield formation than in those using a fixed or modified harvest index

Considering the many challenges reported in undertaking a consistent wheat model ensemble IRS analysis, as well as the myriad options available for classifying IRS patterns, these conclusions support our original hypothesis concerning the utility of classifying IRS patterns for agricultural crops. By means of comparison, it would be interesting to apply similar methods to IRS patterns obtained from other regions and for other crops, as well as for impacts in other sectors, building on the few existing examples in the literature. This could also be complemented by a comparison to observed yields, also plotted as an IRS, but for annual or period-averaged climate anomalies relative to the long-term mean. Such results could be used to inform model selection in future multi-model crop simulation studies by identifying ensemble members that span a wide range of model responses, those that are closely related and outliers that exhibit clearly implausible behaviour.

## 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) (www.macsur.eu).

The authors would like to acknowledge financial support from the following sources: the Academy of Finland for the PLUMES (decisions: 277276, 277403 and 292836) project, the Finnish Ministry of Agriculture and Forestry (FACCE-MACSUR), the European Commission Seventh Framework Programme IMPRESSIONS project (grant agreement no. 603416), the Polish National Centre for Research and Development for the LCAgri (BIOSTRATEG1/271322/3/NCBR/2015) and GyroScan (BIOSTRATEG2 298782) projects and the German Federal Ministry of Food and Agriculture through the Federal Office for Agriculture and Food (grant number 2851ERA01J), FACCE MACSUR (2812ERA 147), the German Federal Ministry of Education and Research via the 'Limpopo Living Landscapes' project within the SPACES programme (grant number 01LL1304A) and the MACMIT project (01LN1317A); and the Italian Ministry of Agricultural Food and Forest Policies (AGROSCENARI Project) and the Italian Ministry of Education, University and Research (FIRB 2012, RBFR12B2K4_004).

An anonymous referee provided constructive comments on an earlier draft of this manuscript.

## Appendix A. Supplementary material

Supplementary material to this article can be found online at http://dx.doi.org/10.1016/j.agsy.2017.08.004.

## References

Abeledo, L.G., Savin, R., Slafer, G.A., 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. http://dx.doi.org/10.1016/j.eja.2007.12.001.

Acutis, M., Confalonieri, R., 2006. Optimization algorithms for calibrating cropping systems simulation models. A case study with simplex-derived methods integrated in the WARM simulation environment. Ital. J. Agrometeorol. 3, 26–34.

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. http://dx.doi.org/10.1016/j.agrformet.2012.11.017.

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rötter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P.K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainty in simulating wheat yields under climate change. Nat. Clim. Chang. 3, 827–832. http://dx.doi.org/10.1038/nclimate1916.

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V.V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.-K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., Zhu, Y., 2014. Rising temperatures reduce global wheat production. Nat. Clim. Chang. 5, 143–147. http://dx.doi.org/10.1038/nclimate2470.

Børgesen, C.D., Olesen, J.E., 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. http://dx.doi.org/10.5194/nhess-11-2541-2011.

Bouman, B.A.M., van Keulen, H., van Laar, H.H., Rabbinge, R., 1996. The "School of de Wit" crop growth simulation models: a pedigree and historical overview. Agric. Syst. 52, 171–198. http://dx.doi.org/10.1016/0308-521X(96)00011-X.

Cammarano, D., Rötter, R.P., Asseng, S., Ewert, F., Wallach, D., Martre, P., Hatfield, J.L., Jones, J.W., Rosenzweig, C., Ruane, A.C., Boote, K.J., Thorburn, P.J., Kersebaum, K.C., Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Müller, C., Kumar, S.N., Nendel, C., O'Leary, G.J., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2016. Uncertainty of wheat water use: Simulated patterns and sensitivity to temperature and $CO_2$. Field Crop Res. 198, 80–92. http://dx.doi.org/10.1016/j.fcr.2016.08.015.

Cartelle, J., Pedró, A., Savin, R., Slafer, G.A., 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. http://dx.doi.org/10.1016/j.eja.2006.07.004.

Castañeda-Vera, A., Leffelaar, P.A., Álvaro-Fuentes, J., Cantero-Martíneze, C., Mínguez, M.I., 2015. Selecting crop models for decision making in wheat insurance. Eur. J. Agron. 68, 97–116. http://dx.doi.org/10.1016/j.eja.2015.04.008.

Challinor, A.J., Ewert, F., Arnold, S., Simelton, E., Fraser, E., 2009. Crops and climate change: progress, trends, and challenges in simulating impacts and informing adaptation. J. Exp. Bot. 60, 2775–2789. http://dx.doi.org/10.1093/jxb/erp062.

Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R., Chhetri, N., 2014. A meta-analysis of crop yield under climate change and adaptation. Nat. Clim. Chang. 4, 287–291. http://dx.doi.org/10.1038/nclimate2153.

Confalonieri, R., Orlando, F., Paleari, L., Stella, T., Gilardelli, C., Movedi, E., Pagani, V., Cappelli, G., Vertemara, A., Alberti, L., Alberti, P., Atanassiu, S., Bonaiti, M., Cappelletti, G., Ceruti, M., Confalonieri, A., Corgatelli, G., Corti, P., Dell'Oro, M., Ghidoni, A., Lamarta, A., Maghini, A., Mambretti, M., Manchia, A., Massoni, G., Mutti, P., Pariani, S., Pasini, D., Pesenti, A., Pizzamiglio, G., Ravasio, A., Rea, A., Santorsola, D., Serafini, G., Slavazza, M., Acutis, M., 2016. Uncertainty in crop model predictions: what is the role of users? Environ. Model. Softw. 81, 165–173. http://dx.doi.org/10.1016/j.envsoft.2016.04.009.

Ekström, M., Grose, M.R., Whetton, P.H., 2015. An appraisal of downscaling methods used in climate change research. Wiley Interdiscip. Rev. Clim. Chang. 6, 301–319. http://dx.doi.org/10.1002/wcc.339.

Ewert, F., Rötter, R.P., Bindi, M., Webber, H., Trnka, M., Kersebaum, K.C., Olesen, J.E., van Ittersum, M.K., Janssen, S., Rivington, M., Semenov, M.A., Wallach, D., Porter, J.R., Stewart, D., Verhagen, J., Gaiser, T., Palosuo, T., Tao, F., Nendel, C., Roggero, P.P., Bartošová, L., Asseng, S., 2015. Crop modelling for integrated assessment of risk to food production from climate change. Environ. Model. Softw. 72, 287–303. http://dx.doi.org/10.1016/j.envsoft.2014.12.003.

Fronzek, S., Carter, T.R., Räisänen, 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. Chang. 99, 515–534. http://dx.doi.org/10.1007/s10584-009-9679-y.

Fronzek, S., Carter, T.R., 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. http://dx.doi.org/10.5194/nhess-11-2981-2011.

Geerts, S., Raes, D., 2009. Deficit irrigation as an on-farm strategy to maximize crop water productivity in dry areas. Agric. Water Manag. 96, 1275–1284. http://dx.doi.org/10.1016/j.agwat.2009.04.009.

Giltrap, D.L., Li, C., Saggar, S., 2010. DNDC: a process-based model of greenhouse gas fluxes from agricultural soils. Agric. Ecosyst. Environ. 136, 292–300. http://dx.doi.org/10.1016/j.agee.2009.06.014.

Holmberg, M., Futter, M.N., Kotamaki, N., Fronzek, S., Forsius, M., Kiuru, P., Pirttioja, N., Rasmus, K., Starr, M., Vuorenmaa, J., 2014. Effects of changing climate on the hydrology of a boreal catchment and lake DOC–probabilistic assessment of a dynamic model chain. Boreal Environ. Res. 19, 66–83.

van Ittersum, M., Leffelaar, P., van Keulen, H., Kropff, M., Bastiaans, L., Goudriaan, J., 2003. On approaches and applications of the Wageningen crop models. Eur. J. Agron. 18, 201–234. http://dx.doi.org/10.1016/S1161-0301(02)00106-5.

Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C.H., Rosenzweig, C., Wheeler, T.R., 2017a. Brief history of agricultural systems modeling. Agric. Syst. 155, 240–254. http://dx.doi.org/10.1016/j.agsy.2016.05.014.

Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C.H., Rosenzweig, C., Wheeler, T.R., 2017b. Toward a new generation of agricultural system data, models, and knowledge products: state of agricultural systems science. Agric. Syst. 155, 269–288. http://dx.doi.org/10.1016/j.agsy.2016.09.021.

Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data an Introduction to Cluster Analysis. Wiley, New York.

Kersebaum, K.C., Boote, K.J., Jorgenson, J.S., Nendel, C., Bindi, M., Frühauf, C., Gaiser, T., Hoogenboom, G., Kollas, C., Olesen, J.E., Rötter, R.P., Ruget, F., Thorburn, P.J., Trnka, M., Wegehenkel, M., 2015. Analysis and classification of data sets for calibration and validation of agro-ecosystem models. Environ. Model. Softw. 72, 402–417. http://dx.doi.org/10.1016/j.envsoft.2015.05.009.

Kim, D.-J., Roh, J.-H., Yun, J.I., 2013. Grain yield response of CERES-barley adjusted for domestic cultivars to the simultaneous changes in temperature, precipitation, and $CO_2$ concentration. Korean Agric. For. Meteorol. 15, 312–319. http://dx.doi.org/10.5532/KJAFM.2013.15.4.312.

Kröbel, R., Sun, Q., Ingwersen, J., Chen, X., Zhang, F., Müller, T., Römheld, V., 2010. Modelling water dynamics with DNDC and DAISY in a soil of the North China Plain: a comparative study. Environ. Model. Softw. 25, 583–601. http://dx.doi.org/10.1016/j.envsoft.2009.09.003.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2016. Cluster: Cluster Analysis Basics and Extensions, R Package Version 2.0.4.

Metzger, M.J., Bunce, R.G.H., Jongman, R.H., Mücher, C.A., Watkins, J.W., 2005. A climatic stratification of the environment of Europe. Glob. Ecol. Biogeogr. 14 (6), 549–563. http://dx.doi.org/10.1111/j.1466-822X.2005.00190.x.

Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.E., Patil, R.H., Ruget, F., Rumbaur, C., Takáč, J., Trnka, M., Bindi, M., Çaldağ, B., Ewert, F., Ferrise, R., Mirschel, W., Şaylan, L., Šiška, B., Rötter, R., 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. http://dx.doi.org/10.1016/j.eja.2011.05.001.

Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minet, J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M., Semenov, M., Slawinski, C., Stratonovitch, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., Rötter, R., 2015. Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces. Clim. Res. 65, 87–105. http://dx.doi.org/10.3354/cr01322.

Prudhomme, C., Crooks, S., Kay, A.L., Reynard, N., 2013a. Climate change and river flooding: part 1 classifying the sensitivity of British catchments. Clim. Chang. 119, 933–948. http://dx.doi.org/10.1007/s10584-013-0748-x.

Prudhomme, C., Kay, A.L., Crooks, S., Reynard, N., 2013b. Climate change and river flooding: part 2 sensitivity characterisation for British catchments and example vulnerability assessments. Clim. Chang. 119, 949–964. http://dx.doi.org/10.1007/s10584-013-0726-3.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T.A.M., Schmid, E., Stehfest, E., Yang, H., Jones, J.W., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proc. Natl. Acad. Sci. 111, 3268–3273. http://dx.doi.org/10.1073/pnas.1222463110.

Rötter, R.P., Palosuo, T., Kersebaum, K.C., Angulo, C., Bindi, M., Ewert, F., Ferrise, R., Hlavinka, P., Moriondo, M., Nendel, C., Olesen, J.E., Patil, R.H., Ruget, F., Takáč, J., Trnka, M., 2012. Simulation of spring barley yield in different climatic zones of Northern and Central Europe: a comparison of nine crop models. Field Crop Res. 133, 23–36. http://dx.doi.org/10.1016/j.fcr.2012.03.016.

Ruane, A.C., Cecil, L.D., Horton, R.M., Gordón, R., McCollum, R., Brown, D., Killough, B., Goldberg, R., Greeley, A.P., Rosenzweig, C., 2013. Climate change impact uncertainties for maize in Panama: farm information, climate projections, and yield sensitivities. Agric. For. Meteorol. 170, 132–145. http://dx.doi.org/10.1016/j.agrformet.2011.10.015.

Ruiz-Ramos, M., Ferrise, R., Rodríguez, A., Lorite, I., Bindi, M., Carter, T.R., Fronzek, S., Palosuo, T., Pirttioja, N., Baranowski, P., Buis, S., Cammarano, D., Chen, Y., Dumont, B., Ewert, F., Gaiser, T., Hlavinka, P., Hoffmann, H., Höhn, J.G., Jurecka, F., Kersebaum, K.C., Krzyszczak, J., Lana, M., Mechiche-Alami, A., Minet, J., Montesino, M., Nendel, C., Porter, J.R., Ruget, F., Semenov, M., Steinmetz, Z., Stratonovitch, P., Supit, I., Tao, F., Trnka, M., de Wit, A., Rötter, R.P., 2017. Adaptation response surfaces for local management of wheat under perturbed climate and $CO_2$ concentration in a Mediterranean environment. Agric. Syst. http://dx.doi.org/10.1016/j.agsy.2017.01.009.

Varella, H., Guérif, M., Buis, S., 2010. Global sensitivity analysis measures the quality of parameter estimation: the case of soil parameters and a crop model. Environ. Model. Softw. 25, 310–319. http://dx.doi.org/10.1016/j.envsoft.2009.09.012.

Wallach, D., Mearns, L.O., Ruane, A.C., Rötter, R.P., Asseng, S., 2016. Lessons from climate modeling on the design and use of ensembles for crop modeling. Clim. Chang. http://dx.doi.org/10.1007/s10584-016-1803-1.

Webber, H., Gaiser, T., Oomen, R., Teixeira, E., Zhao, G., Wallach, D., Zimmermann, A., Ewert, F., 2016. Uncertainty in future irrigation water demand and risk of crop failure for maize in Europe. Environ. Res. Lett. 11, 074007. http://dx.doi.org/10.1088/1748-9326/11/7/074007.

Weiß, M., Alcamo, J., 2011. A systematic approach to assessing the sensitivity and vulnerability of water availability to climate change in Europe. Water Resour. Res. 47. http://dx.doi.org/10.1029/2009WR008516.

White, J.W., Hoogenboom, G., Kimball, B.A., Wall, G.W., 2011. Methodologies for simulating impacts of climate change on crop production. Field Crop Res. 124, 357–368. http://dx.doi.org/10.1016/j.fcr.2011.07.001.