Increasing influence of heat stress on French maize yields from the 1960s to the 2030s

Ed Hawkins[*1], Thomas E. Fricker[2], Andrew J. Challinor[3], Christopher A. T. Ferro[2,4], Chun Kit Ho[1], and Tom M. Osborne[1]

[1]NCAS-Climate, Department of Meteorology, University of Reading, Reading, UK
[2]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK
[3]Institute for Climate and Atmospheric Science, University of Leeds, Leeds, UK
[4]NCAS-Climate, University of Exeter, Exeter, UK

Abstract

Improved crop yield forecasts could enable more effective adaptation to climate variability and change. Here we explore how to combine historical observations of crop yields and weather with climate model simulations to produce crop yield projections for decision relevant timescales. Firstly, the effects on historical crop yields of improved technology, precipitation and daily maximum temperatures are modelled empirically, accounting for a non-linear technology trend and interactions between temperature and precipitation, and applied specifically for a case study of maize in France. The relative importance of precipitation variability for maize yields in France has decreased significantly since the 1960s, likely due to increased irrigation. In addition, heat stress is found to be as important for yield as precipitation since around 2000. A significant reduction in maize yield is found for each day with a maximum temperature above 32°C, in broad agreement with previous estimates. The recent increase in such hot days has likely contributed to the observed yield stagnation. Furthermore, a general method for producing near-term crop yield projections, based on climate model simulations, is developed and utilised. We use projections of future daily maximum temperatures to assess the likely change in yields due to variations in climate. Importantly, we calibrate the climate model projections using observed data to ensure both

[*]e.hawkins@reading.ac.uk

reliable temperature mean and daily variability characteristics, and demonstrate that these methods work using retrospective predictions. We conclude that, to offset the projected increased daily maximum temperatures over France, improved technology will need to increase base level yields by 12% to be confident about maintaining current levels of yield for the period 2016-2035; the current rate of yield technology increase is not sufficient to meet this target.

1 Introduction

The yield of most crops has increased over the past several decades. However, in the most recent decade, yields have stagnated for many crops in several regions, whilst temperatures have generally increased. The reasons for this stagnation are debated, and could include agricultural policy (Finger, 2010), fundamental genetic limits (Calderini and Slafer, 1998), climate (Lobell and Asner, 2003; Brisson et al., 2010), agronomic practice and crop management (Brisson et al., 2010). Here we explore the relative importance of different climatic factors.

Crops are known to be sensitive to various aspects of climate. Persistently elevated temperatures have long been known to accelerate progress towards maturity, and more recently have been shown to have a significant impact on leaf aging (or senescence) (Asseng et al., 2011; Lobell et al., 2012). Crop responses to shorter periods of high temperature, particularly when coincident with flowering, show yields falling dramatically beyond a threshold temperature (Luo, 2011). This mechanism is observed in both controlled environments and field studies (Ferris et al., 1998; Wheeler et al., 2000). Similar responses to hot days are beginning to be found at the regional scale: maize yields in the U.S. have been found to decrease sharply when exposed to temperatures over around 29-30°C, and this effect outweighs any yield increase due to higher temperatures more generally (Schlenker and Roberts, 2009).

Crop yields are also sensitive to precipitation. Quantifying the relative effect of temperature and precipitation variability is important for understanding impacts and developing adaptation options for future climatic changes. Whilst this relative importance will vary regionally (e.g. Sakurai et al., 2011), some generalisations may be possible through an analysis of mechanisms. For regions where irrigation is increasing, for example, it seems likely that the sensitivity of yield to rainfall will be decreasing. More detailed analyses also indicate that in particular environments (Thornton et al., 2010) or at the regional scale (Lobell and Burke, 2008), temperature may be a more significant driver of future yields than precipitation. Since temperatures are projected to significantly increase over the next few decades due to continuing anthropogenic emissions of greenhouse gases, whereas precipitation changes are far less certain (Meehl et al., 2007; Hawkins and Sutton, 2011), this suggests predictability in future crop yields.

In order to effectively guide adaptation to future changes, perhaps with different crop growing strategies (Rosenzweig and Tubiello, 2007) or selective crop breeding (Cattivelli et al., 2008), there are several key questions to consider. Firstly, can the relative effects of improved technology, precipitation variability and increasing temperatures be quantified? If so, what is the relative size of the effects of rainfall and hot temperatures on yields? And, what level of technology development may be required to overcome any impact of future climatic changes on

yield?

In this analysis we develop a methodology to address these questions, focussing on one particular crop (maize) and one country (France) as a case study to better understand the technology trend and the influence of climate on crops. France is chosen specifically for this case study because it has experienced recent extremes of climate. In particular, the heatwave in summer 2003 (Schär et al., 2004) has previously been linked to a drop in crop yields across Europe (Easterling et al., 2007; Battisti and Naylor, 2009; van der Velde et al., 2012).

2   Materials and Methods

The overall approach is to fit an empirical model to historical observations of climate and crop yield to determine the relative importance of technology, heat stress and precipitation. Climate model simulations are used to make calibrated projections of future heat stress, which are then used to produce yield forecasts assuming no technological development and that the present relationships between climate and yield variability apply in the future. Equivalently, this provides an estimate of how much technological development may be required to maintain yields at present levels. Unless otherwise stated, all uncertainties are given as a $5-95\%$ confidence range.

Observed climate and crop yield data

The relationships between yield and climate are examined using historical daily precipitation and maximum temperatures from the E-OBS dataset (Haylock et al., 2008), which is available on a $0.5°\times0.5°$ grid since 1950, and annual maize yield data from FAOSTAT (http://faostat.fao.org/). We choose to focus on national level yield data (1961-2010) to provide longer time series to examine trends. Regional yield data (for NUTS2 regions) is only available from 1980-2007 which does not allow such a long timescale view. However, we briefly compare the analysis on national scales with the regional data in the Supp. Info.

We consider two alternatives for measuring heat stress – a simple count of the number of days above a certain critical threshold, and an integrated measure of the degree days above a threshold. Both measures are defined using daily maximum temperature ($T_{\max}$) during the growing season (June, July and August – JJA), averaged over the whole of France, but weighted by the area of maize harvested in each region (Monfreda et al., 2008, Fig. 1). A precipitation index is defined as the mean JJA rainfall, similarly averaged over France. Although the locations of maize growth may have changed over time, similar conclusions are reached if no weighting is applied. In addition, the average planting day may have changed over time (Kucharik, 2006), but given that we are using seasonal averages of climate the effect on our analysis is likely to be

small. Finally, we have not considered the details of the timing of the weather events, although this may be extremely important for certain phenological stages of crop growth.

An empirical model for maize yield

A simple physical understanding for the causes of yield changes suggest that an empirical model for maize yield in France, considering a non-linear 'technology' trend and both the effects of temperature and precipitation, should effectively describe the yield variability. However, the variability in hot days and precipitation is not independent and it is also possible that the effect on yield of an increase in hot days will depend on the precipitation, suggesting that an interaction term may be required (e.g. Runge, 1968; Schlenker and Roberts, 2009). Hence, a generalised additive model (Rigby and Stasinopoulos, 2005) for maize yield (Y) is proposed,

$$Y(t) = g(t) + \beta_1 X(t) + \beta_2(t)(P(t) - \bar{P}) + \beta_3 X(t)P(t) + e(t) \tag{1}$$

where X and P are the temperature and precipitation indices respectively, $\bar{P}$ is the mean precipitation index over 1961-2010, the β parameters represent the size of the effects of the various terms, $g(t)$ is the expected yield in year t if there were no hot days and average precipitation, and e is a stochastic error term. We let g be a cubic regression spline to represent the increase in expected yield due to improving technology, which avoids the arbitrary, but often used, assumption that the technology trend is linear with time. The errors are assumed to be normally distributed and temporally independent, but we allow their variance to vary with time to allow for changes in the influence of weather (e.g. precipitation) on yield variability due to technological improvements such as irrigation. To facilitate this we let $e(t) = h(t)\epsilon(t)$ where $h(t)$ is a cubic regression spline and the $\epsilon(t)$ are independent standard normal random variables. The unknown β parameters and the spline functions g and h are all estimated by maximising a penalised likelihood function (see Supp. Info. for more details). Note that β_2 is time dependent - we assume a similar spline function for its variation. The justifications for the choice of this empirical yield model, as well as tests of simpler and more complex versions, are given below, in Section 3 and in the Supplementary Information.

Empirical yield model selection

Many different empirical models for crop yield have been proposed. A key benefit of choosing a generalised additive model such as Eqn. 1 is that all the empirical model parameters, including the non-linear trend component, are fitted *simultaneously* (e.g. Lobell et al., 2011), so as to reduce the chances of overfitting on certain parameters, in contrast to other studies (e.g. Sakurai et al., 2011). In addition, the choice of technology trend has been much discussed, with many

5

arbitrary assumptions used. For example, technology trends have been assumed to be linear (e.g. Lobell and Asner, 2003), or quadratic with time (e.g. Schlenker and Roberts, 2009; Lobell et al., 2011), or removed using local linear regression (e.g. Sakurai et al., 2011) or first-differences (e.g. Nicholls, 1997). In some cases the technology trend has not been considered at all (e.g. Tao et al., 2006; Knox et al., 2012).

Our choice of a cubic spline covers many of these other possibilities as a special case, but is far more flexible. However, to examine the sensitivity to the choice of technology trend in our analysis we considered a version of Eqn. 1 with a linear trend for $g(t)$, rather than a cubic spline. This version of the model produced a significantly poorer cross-validation (see Supp. Info.), and we argue that a non-linear trend is more robust.

In addition, we advocate 'appropriate complexity' for an empirical yield model but additional complexity needs to be considered. For example, the validation statistics of the model were found to be significantly improved if the direct influence of precipitation (β_2) varies with time (also see e.g. Sakurai et al., 2011), and so this factor was included. However, we also tested versions of the empirical model with higher order terms (such as quadratic in X and P) and also considered time-varying β_1 and β_3, but found that these changes did not improve the empirical relationship significantly (see Supp. Info.). This yield model also overcomes criticisms of simpler empirical models (e.g. Gregory and Marshall, 2012; Semenov et al., 2012) by including an interaction between temperature and precipitation, and basing the choice of possible heat stress indices on the known physical links between hot days and crop growth. Eqn. 1 is the simplest version which is found to produce yield estimates which are consistent with the assumptions made in the empirical model, i.e. the residuals are consistent with being independent and random.

Climate simulations and calibration

Our set of climate model simulations is a QUMP (Quantifying Uncertainty in Model Predictions) ensemble, which consists of 16 variants of the HadCM3 global climate model (GCM) (Gordon et al., 2000; Collins et al., 2011). This GCM has an atmospheric resolution of $2.5° \times 3.75°$. Each member of the ensemble differs only in values of particular atmospheric parameters which govern physical processes which are not fully resolved in the model. This ensemble is particularly appropriate for this analysis because it was designed specifically to sample a wide range of climate sensitivities (Collins et al., 2011). We use the daily maximum temperature data in JJA during 1960-2035. Historical radiative forcings were used before the year 2000, and the SRES A1B emissions scenario (Nakicenovic et al. 2000) was followed after 2000.

Here we utilise two approaches for the calibration, both of which are fairly standard in crop modelling, namely 'bias correction' (BC) and 'change factor' (CF). Both of these methods use historical observations and simulations to derive corrections which can be applied to the future projections, but using different assumptions. Additionally, we extend previous methods by also accounting for differences in daily temperature *variability* between the climate simulations and observations, as well as differences in mean climate (Ho, 2010; Ho et al., 2012; Hawkins et al., 2012), which is particularly important when considering the hot day counts over a threshold (see Supp. Info.).

To perform a calibration we require daily $T_{\max}$ timeseries from a GCM simulation and observations for the same reference period, which we denote by $T_{\mathrm{REF}}(t)$ and $O_{\mathrm{REF}}(t)$ respectively. We also need output from the GCM for some future period of the same length as the reference period, $T_{\mathrm{RAW}}(t)$. The question remains about how to best combine these three sources of information into the most robust projections of the unknown future observations $(\widehat{O}_{\mathrm{FUT}})$ to use as input for crop models. We consider both bias correction and change factor methods, including corrections for the variability as well as the mean climate, to sample this source of uncertainty.

Bias correction: The usual bias correction (BC) methodology corrects the projected raw daily GCM output using the differences only in the mean between observations and the GCM in a particular reference period (e.g. Huntingford et al., 2005; Ines and Hansen, 2006). However, a more general case when correcting the variability also (Ho et al., 2012) is:

$$\widehat{O}_{\mathrm{FUT,BC}}(t) = \overline{O_{\mathrm{REF}}} + \frac{\sigma_{O,\mathrm{REF}}}{\sigma_{T,\mathrm{REF}}} \left(T_{\mathrm{RAW}}(t) - \overline{T_{\mathrm{REF}}} \right), \tag{2}$$

and $\sigma_{T,\mathrm{REF}}$ and $\sigma_{O,\mathrm{REF}}$ represent the standard deviation of the daily GCM output and observations in the reference period respectively.

Change Factor: The change factor (CF) methodology instead utilises the *observed* daily variability and changes the mean as simulated by the GCM (e.g. Arnell et al., 2003; Gosling et al., 2009). The general form when correcting the daily variance also (Ho et al., 2012) is:

$$\widehat{O}_{\mathrm{FUT,CF}}(t) = \overline{T_{\mathrm{RAW}}} + \frac{\sigma_{T,\mathrm{RAW}}}{\sigma_{T,\mathrm{REF}}} \left(O_{\mathrm{REF}}(t) - \overline{T_{\mathrm{REF}}} \right), \tag{3}$$

and $\sigma_{T,\mathrm{RAW}}$ represents the standard deviation of the daily raw model output for the future period.

The grid point of the climate model which includes the position of the observations is used in Eqns. 2, 3, so these methodologies also effectively downscale the simulated temperature data to the spatial scale of the available observations. Where the observations are in a location where

the climate model has an ocean grid point (grey areas in left column of Fig. 5 later), the nearest land point is selected from the climate model. We use each of the 16 QUMP simulations as independent projections and calibrate $T_{\max}$ separately for each simulation as above.

The assumptions in the choice of bias correction or change factor are slightly different. If considering future mean climate and no changes in variability, then the two methods produce identical results. However, the more general case above can produce differences in future calibrated climates which are as large as differences between emission scenarios (Ho et al., 2012; Hawkins et al., 2012). Both methods essentially assume that the change in climate is independent of the mean state, but CF starts from the observations and BC starts from the model output. These methodologies do not consider changing the shape of the distribution of climate data, but this does not matter for a hot days metric in our analysis (see Supp. Info.), but may be more important in other situations. Some limited idealised experiments suggested that change factor methods may outperform bias correction methods, because they utilise the spatial and temporal variability of the observations, but they may also under-estimate the uncertainty because of the limited sampling of the observed variability (Hawkins et al., 2012). In the absence of more concrete results, we assume that both methods are equally plausible.

3 Results

Observed changes to yield and climate

Maize is a widely grown crop in France (Fig. 1) and yields have gradually increased from $0.25\,\mathrm{kg}\ \mathrm{m}^{-2}$ to a peak of $0.97\,\mathrm{kg}\ \mathrm{m}^{-2}$ over the past 40 years[1] (Fig. 2c). This increase has been attributed to a combination of improved technology (such as fertilisers, pesticides and machinery), more robust and productive crop varieties, as well as CO_2 fertilisation effects (e.g. Gervois et al., 2008).

In addition, the number of hot days has increased in France since the 1960s (Fig. 2b, using a $32°C$ threshold). Assuming a linear relationship with global mean temperatures suggests a significant increase of 4.5 (0.7 to 8.3) hot days per $1°C$ global temperature rise. Particularly hot years, when compared to nearby years, occurred in 1964, 1976, 1990, 2003 and 2006 (also see Fig. S1) and the corresponding maize yield also shows depressed yields in the same years (Fig. 2c). There is no significant trend in precipitation since the 1960s, but variations in maize yields in the 1960s and 1970s seem to be strongly related to precipitation variability (Fig. 2a).

During the heatwave of 2003, the maize yield in France fell to $0.71\,\mathrm{kg}\ \mathrm{m}^{-2}$ – a 20% drop on the previous year (also see e.g. van der Velde et al., 2010). It seems likely that this yield decrease

---

[1] We use SI units for yield, but $1\,\mathrm{kg}\ \mathrm{m}^{-2}$ is equivalent to $10\ \mathrm{T}\ \mathrm{ha}^{-1}$ or $10^4\ \mathrm{kg}\ \mathrm{ha}^{-1}$

was related to the hot temperatures that summer – but is this true of less extreme years? And, what is the role of precipitation variability?

Considering temperature only

To explore these suggestive qualitative links we first utilise a simple form of Eqn. 1, considering the effects of temperature alone, i.e. $\beta_2 = \beta_3 = 0$. We consider two choices for the temperature index, X: firstly, a simple count of the number of days over a critical temperature threshold, and secondly, the integrated temperature-days above a critical threshold. By fitting the suggested yield model (simplified from Eqn. 1) to the observed data, it is found that a threshold of $32°C$ is the optimal choice for a simple hot day count, and $26.5°C$ is optimal for the integrated temperature-days (Fig. S3a). For the analysis which follows, we utilise a simple hot day count, which produces a superior fit to the observations than an integrated measure, and also better accounts for differences between the observed and climate model simulated temperature variability (Fig. S4).

In addition, Fig. 3a (red line) shows the residuals from the expected yield,

$$\text{residuals} = \text{observed yield} - g(t) - \beta_1 X(t) \tag{4}$$

when considering the count of days above $32°C$ as X. The running standard deviation of the residuals (red line in Fig. 3b) shows a noticeable decline with time, suggesting increased yield stability recently. In addition, the residuals are well correlated with the mean precipitation anomaly for France ($r = 0.57$, Fig. 3a), especially for the earlier years when there was less irrigation (FAOSTAT; Fig. 3b). This finding demonstrates the need to include precipitation in the empirical model and is consistent with an increase in irrigation, and/or the development of maize varieties which are more robust to drought, reducing the impact of precipitation anomalies. However, it should be noted that irrigation may also be increased by farmers during periods of high temperature stress, such as the 2003 heatwave, as an adaptation strategy (van der Velde et al., 2010).

Considering both temperature and precipitation

The findings above suggest improving the yield model by adding the effects of precipitation, including an interaction term. However, the influence of precipitation should decrease over time (Fig. 3b), suggesting that β_2 should be a smooth function, rather than a constant. Note also that the hot day index and precipitation are not independent - the correlation, $r = -0.46$.

Fitting the full model (Eqn. 1) to the data retains the finding that a $32°C$ threshold is optimal (Fig. S3a). When precipitation is included the residuals are consistent with having

9

a constant variance and there is no significant improvement in the yield model by allowing a time-dependent effect for $h(t)$. For this full yield model, the $h(t)$ term is therefore assumed to be a constant.

The predicted yield (Y; Eqn. 1) with associated uncertainties reliably encompasses the observed yields (Fig. 2c). The red shading indicates the uncertainty in expected yield (without the ϵ term) and the red lines indicate the total uncertainty in actual annual yields.

The derived technology trend (g) for this yield model increases non-linearly since 1961 with a noticeable plateau in the 1970s (Fig. 2c, grey shading). Although the absolute rate of increase has also slowed again in the most recent decade, the technology trend is still increasing more rapidly than the actual yield. This supports the hypothesis that the recent increase in the number of hot days has caused the actual yield to stagnate (Brisson et al., 2010), and is inconsistent with suggestions that the observed plateau in yields is evidence of a fundamental genetic limit on potential yields (Calderini and Slafer, 1998). We now only consider this full empirical yield model.

Relative importance of temperature and precipitation

A key aspect of this analysis is the ability to determine the relative importance of temperature and precipitation, and how this importance has changed over recent decades. Examining the relative size of the different β parameters suggests that precipitation variability was the dominant contributor to yield variability until around 2000 (Fig. 4). For the most recent decade, the effects of heat stress variability are now as important as precipitation variability, perhaps due to increased irrigation of maize in France (Fig. 3b). Although the interaction term slightly complicates this simple interpretation, it is clear that the *relative* importance of temperature has increased over time. However, it is worth noting that the presence and sign of the interaction term means that hot days become less damaging for yields as precipitation increases.

For the near-term (2016-2035), the mean IPCC AR4 projection for summer over Europe is an increase of around 1°C in mean temperature and a 5% decrease in precipitation from 1980-1999 levels (Meehl et al., 2007). However, the uncertainty in precipitation projections is far larger than for temperature (Hawkins and Sutton, 2011), and confidence in the sign of the precipitation change is much lower (Meehl et al., 2007), partly because present day simulations of both mean precipitation and its variability are worse than for temperature (Randall et al., 2007). In addition, it is likely that temperature will have the largest impact as the projected changes are far further outside the range of natural variability than for precipitation changes (Lobell and Burke, 2008), and because of the seasonal timing of changes in climate (Semenov

and Shewry, 2011).

So, for making future projections of crop yields we use the full empirical model considering temperature and precipitation, but focus purely on the effects of changes in temperature, and make the (slightly optimistic) assumption that the climatological distribution of precipitation (from 1961-2010) will not change.

Retrospective calibrated projections of climate

The construction of the empirical model suggests that yields can be forecast if the number of hot days is known. In principle, climate model simulations can be used to make this projection. However, a key issue in using climate model simulations to study impacts is that the models are biased and do not perfectly reproduce the current climate. For instance, the QUMP ensemble of simulations used here is generally too warm over Europe and produces too many hot days when compared to observations (Fig. 5). Other climate models are less or more biased in this metric (Hawkins et al., 2012). Therefore, some calibration is needed before the simulations can be used. To increase confidence in the ability of the calibrated climate model simulations to make forecasts for the number of hot days, we test the predictions retrospectively by comparing with historical observations.

Using the observational data from 1966-1985 only and climate model data from 1966-1985 and 1991-2010, it is possible to make an out-of-sample calibrated projection for the number of hot days observed in the 1991-2010 period using the two different calibration methods (Fig. 5). The calibrations correct much of the warm bias in the raw simulations and produce robust projections of the number of hot days. Note that the variability corrections introduced in Eqns. 2, 3 are essential to producing reliable predictions (compare Fig. 5 with Fig. S5).

Averaged over the maize growing regions of France, the raw simulations would produce a hot day index of more than 30 for the 1991-2010 period, but the calibrated projection for the hot day index is 6.3 (3.9-9.4) days (CF) and 4.1 (0.8-7.8) days (BC). The observed hot day index for the 1991-2010 period was 6.4 days (or 5.5 days without the extreme of 2003), an increase on 3.2 days from the 1966-1985 period. The observations are therefore within the uncertainties predicted by the calibrated climate model simulations. It is worth reiterating that we have not used the observations for 1991-2010 to train this climate model prediction – it is made out-of-sample.

When considering a particular location where the largest fraction of the maize is grown (black dots in Fig. 5), the calibrations have narrowed the QUMP spread, reduced the projected number of hot days, and now encompass the observations for both calibration methodologies, unlike the raw simulations (Fig. 6). Histograms are shown for the projections of the QUMP ensemble using

11

raw model output (left column) and calibrated output (right column). The remaining spread in the projected number of hot days represents differences between the climate models used, and also different realisations of climate variability.

When considering projections on annual timescales (Fig. S6) it is shown that the projected calibrated probability for a 2003-type summer would have been less than 0.6% for the 1991-2010 period. This suggests that the 2003 summer was extreme, even considering the climatic changes, and consistent with other studies examining this heatwave which suggested it was a roughly a 1-in-200 year event (Stott et al., 2004).

Overall, these retrospective tests, along with previous studies in idealised situations (Hawkins et al., 2012), provide evidence that relatively short lead time (a decade or two) calibrated projections of the number of hot days can be made.

Projections of future hot days over France

Having demonstrated that the methodology works retrospectively, it can be applied to make a projection for the future period 2016-2035, using 1991-2010 (without 2003) as our training data. There is a projected increase in the number of hot days for many regions (Fig. 5), particularly in regions where maize is grown. Projections for the mean number of hot days per year in the 2016-2035 period for an individual region in south-west France are shown in Fig. 6. The calibrations have narrowed the QUMP range and reduced the projected number of hot days when compared to the raw ensemble. However, the projections still indicate an increase in the number of hot days in 2016-2035 from present, to around 15-20 days per summer for this location. Note particularly that the observations from 2003 are deliberately excluded from the calibration as it was such an extreme year, and could bias the projections to produce too many hot days.

Finally, Fig. 2b shows a calibrated probabilistic near-term projection for the period 2016-2035 for the average number of hot days per year, averaged over France, and weighted for maize growing regions. The projected ranges for 2016-2035 show a likely increase in the number of hot days to around 10 hot days per year, compared to the present day (1991-2010, without 2003) of around 5.5 hot days per year. The two calibration methods do not produce significantly different estimates – 6.8-16.9 (CF) and 4.4-14.4 (BC).

Using annual projections, the chance of a 2003-type summer in the 2016-2035 period is projected to be around 3% per year (Fig. S6), equivalent to an increase of risk of about an order of magnitude from the historical period. This suggests that the probability of at least one summer like 2003 is around 0.5 in this near-term period, assuming independence between years.

12

Consequences for future maize yield

At the time of writing, yield data for 2011 has not been published by FAOSTAT. However, the observed climate variability data is available from E-OBS, suggesting a summer close to the long-term mean in terms of precipitation and hot days (Fig. 2). Applying our full empirical model, the yield forecast for 2011 is 0.90-1.00 kg m^{-2}, assuming no change in yield due to technology since 2010. Over the past decade yield has increased at roughly 0.005 kg m^{-2} per year due to the technological trend (g).

We also define the base level yield (0.92 kg m^{-2}) as the mean present day yield (1991-2010, without 2003), corrected for the technology trend increases over the same period. For the future, we do not know the technology trend, and can only make projections for the yield assuming the technology remains constant.

Fig. 2c shows probabilistic projections of mean maize yield for France for 2016-2035 using both calibration methodologies (colours) and for two different assumptions on the links between future temperature and precipitation (Fig. S7). Assuming future precipitation is independent of temperature then the projected yield for 2016-2035 is 0.93 (0.89-0.96) kg m^{-2} (BC) and 0.92 (0.88-0.96) kg m^{-2} (CF). However, if the historical correlation between precipitation and temperature is maintained, which we consider more likely, then the predicted yield decreases to 0.91 (0.86-0.96) kg m^{-2} (BC) and 0.88 (0.81-0.92) kg m^{-2} (CF), demonstrating the need to consider correlations between temperature and precipitation in yield projections. We see no reason why a correlation of the same sign would not remain in this near term period, although its magnitude may change.

Equivalently, according to these climate model simulations and calibration techniques, technology developments must increase yield by 0.11 kg m^{-2} (or around 12% of the current base level yield) to be confident of maintaining yield at present levels. The current rate of yield increase due to technology is not sufficient to meet this target, but would be sufficient to meet the median projection of a required 0.04 kg m^{-2}, or a 4% increase in base level yield.

4   Discussion

We have quantified the relative importance of temperature and precipitation for historical and future maize yield on France. In addition, we have outlined a methodology for producing calibrated projections of future climate and crop yields, and tested the methods retrospectively. Our main findings are as follows:

1. Our modelled historical technology trend for yield is non-linear, and suggests a recent slowing in potential yield increases.

2. Maize yield stability in France has increased markedly since the 1960s, likely due to irrigation and technology improvements.

3. The relative importance of precipitation variability for maize yields in France has decreased since the 1960s and the effect of heat stress variability is now as important as precipitation.

4. The number of hot days (above $32°C$), averaged over France, has increased since the 1960s and is projected to increase further to around 10 per summer in the period 2016-2035. For some large maize producing regions, around 15 days per summer are expected.

5. Improved technology will need to increase base level yields by 12% above current levels to be confident about maintaining current maize yields. The current rate of yield increase due to technology is not sufficient to meet this target.

6. Appropriate use of climate model simulations by taking account of differences in both the mean *and* variability of climate is essential, and a rigorous assessment of the characteristics of GCM output is required before its use.

Uncertainty in the projected yields comes from various sources. The component of uncertainty due to the choice of calibration method is not negligible, although 'change factor' (CF) performs slightly better in retrospective forecasts (see Supp. Info.) and idealised modelling studies (Hawkins et al., 2012). Each QUMP simulation produces a different calibrated projection, and we have assumed that the QUMP ensemble spans the full range of climate response uncertainty and climate variability for European temperatures. In addition, there are other potential sources of uncertainty in our projections that we have not considered. For example, we have only used a single (SRES A1B) future emissions scenario, but the relative importance of emissions uncertainty is likely to be small for the near-term for temperature and precipitation (Hawkins and Sutton, 2009, 2011). The effect of these caveats could be reduced by utilising the forthcoming CMIP5 climate model simulations which will produce daily data at a higher spatial resolution for more climate models than QUMP (Taylor et al., 2012). Finally, we have not considered the effects of changes in ozone, which could be significant for maize yields (e.g. Heagle et al., 1972; Hollaway et al., 2012). This will be explored in further work.

There has been recent, and we believe correct, criticism of the use of simple empirical relationships between climate and crop yields to infer future yields (e.g. Gregory and Marshall, 2012; Semenov et al., 2012). We suggest that the careful consideration of non-linear technology

trends and an interaction between temperature and precipitation is essential in any such empirical model. In addition, the empirical yield model parameters, including the trend component, should be fitted simultaneously.

The availability of smaller spatial scale crop yield data may also allow improvements in the empirical relationships between hot days, precipitation and yield, although the regional yield time series are not currently long enough to make robust conclusions about long term trends in the temperature and precipitation effects (see Supp. Info.).

Although this is a case study aimed at providing decision relevant information for a single crop for a single country, future work will aim to provide a wider scale view of future crop yields, based on appropriate use of climate model simulations.

References

Arnell, N. W., D. A. Hudson, and R. G. Jones, 2003: Climate change scenarios from a regional climate model: Estimating change in runoff in southern Africa. *J. Geophys. Res.*, **108**, 4519–4535.

Asseng, S., I. Foster, and N. C. Turner, 2011: The impact of temperature variability on wheat yields. *Glob. Change Biol.*, **17**, 997–1012, doi:10.1111/j.1365-2486.2010.02262.x.

Battisti, D. S. and R. L. Naylor, 2009: Historical warnings of future food insecurity with unprecedented seasonal heat. *Science*, **323**, 240–244, doi:10.1126/science.1164363.

Brisson, N., P. Gate, D. Gouache, G. Charmet, F.-X. Oury, and F. Huard, 2010: Why are wheat yields stagnating in europe? a comprehensive data analysis for france. *Field Crops Research*, **119**, 201 – 212, doi:10.1016/j.fcr.2010.07.012.

Calderini, D. and G. Slafer, 1998: Changes in yield and yield stability in wheat during the 20th century. *Field Crops Research*, **57**, 335–347.

Cattivelli, L., F. Rizza, F.-W. Badeck, E. Mazzucotelli, A. Mastrangelo, E. Francia, C. Mar, A. Tondelli, and A. M. Stanca, 2008: Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. *Field Crops Res.*, **105**, 1–14, doi: 10.1016/j.fcr.2007.07.004.

Collins, M., B. B. B. Booth, B. Bhaskaran, G. R. Harris, J. M. Murphy, D. M. H. Sexton, and M. J. Webb, 2011: Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. *Clim. Dyn.*, **36**, 1737–1766, doi:10.1007/s00382-010-0808-0.

Easterling, W., P. Aggarwal, P. Batima, K. Brander, L. Erda, S. Howden, A. Kirilenko, J. Morton, J.-F. Soussana, J. Schmidhuber, and F. Tubiello, 2007: *Food, fibre and forest products. In: Climate Change 2007: Impacts, Adaptation and Vulnerability.* Cambridge University Press, Cambridge, UK.

Ferris, R., R. Ellis, T. Wheeler, and P. Hadley, 1998: Effect of high temperature stress at anthesis on grain yield and biomass of field-grown crops of wheat. *Ann. Bot.*, **82**, 631–639.

Finger, R., 2010: Evidence of slowing yield growth - the example of Swiss cereal yields. *Food Policy*, **35**, 175–182, doi:10.1016/j.foodpol.2009.11.004.

Gervois, S., P. Ciais, N. de Noblet-Ducoudr, N. Brisson, N. Vuichard, and N. Viovy, 2008: Carbon and water balance of european croplands throughout the 20th century. *Glob. Biogeochem. Cycles*, **22**, GB2022, doi:10.1029/2007GB003018.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.*, **16**, 147–168.

Gosling, S., G. McGregor, and J. Lowe, 2009: Climate change and heat-related mortality in six cities part 2: climate model evaluation and projected impacts from changes in the mean and variability of temperature with climate change. *Int. J. of Biomet.*, **53**, 31–51, doi: 10.1007/s00484-008-0189-9.

Gregory, P. J. and B. Marshall, 2012: Attribution of climate change: a methodology to estimate the potential contribution to increases in potato yield in scotland since 1960. *Global Change Biology*, **18**, 1372–1388, doi:10.1111/j.1365-2486.2011.02601.x.

Hawkins, E., C. K. Ho, T. Osborne, and A. J. Challinor, 2012: Calibration and bias correction of climate projections for crop modelling: an idealised case study over europe. *Ag. For. Met.*, **in press**, doi:10.1016/j.agrformet.2012.04.007.

Hawkins, E. and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Met. Soc.*, **90**, 1095–1107, doi:10.1175/2009BAMS2607.1.

———, 2011: The potential to narrow uncertainty in projections of regional precipitation change. *Clim. Dyn.*, **37**, 407–418, doi:10.1007/s00382-010-0810-6.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New, 2008: A European daily high-resolution gridded dataset of surface temperature and precipitation. *J. Geophys. Res.*, **113**, D20 119, doi:10.1029/2008JD10201.

Heagle, A., D. Body, and E. Pounds, 1972: Effect of ozone on yield of sweet corn. *Phytopathology*, **62**, 683–687.

Ho, C. K., 2010: Projecting extreme heat-related mortality in Europe under climate change. Ph.D. thesis, University of Exeter, UK.

Ho, C. K., D. B. Stephenson, M. Collins, C. A. T. Ferro, and S. J. Brown, 2012: Calibration strategies: a source of additional uncertainty in climate change projections. *Bull. Amer. Met. Soc.*, **93**, 21–26, doi:10.1175/2011BAMS3110.1.

Hollaway, M. J., S. R. Arnold, A. J. Challinor, and L. D. Emberson, 2012: Intercontinental trans-boundary contributions to ozone-induced crop yield losses in the northern hemisphere. *Biogeosciences*, **9**, 271–292, doi:10.5194/bg-9-271-2012.

Huntingford, C., H. F. Lambert, J. H. C. Gash, C. M. Taylor, and A. J. Challinor, 2005: Aspects of climate change prediction relevant to crop productivity. *Phil. Trans. B.*, **360 (1463)**, 1999–2009, doi:10.1098/rstb.2005.1748.

Ines, A. V. M. and J. W. Hansen, 2006: Bias correction of daily GCM rainfall for crop simulation studies. *Agr. For. Met.*, **138**, 44 – 53, doi:10.1016/j.agrformet.2006.03.009.

Knox, J. W., A. Hurford, L. Hargreaves, and E. Wall, 2012: Climate change risk assessment for the agriculture sector. Tech. rep., Department for Environment, Food and Rural Affairs, UK.

Kucharik, C. J., 2006: A multidecadal trend of earlier corn planting in the central USA. *Agron. J.*, **98**, 1544–1550, doi:10.2134/agronj2006.0156.

Lobell, D. B. and G. P. Asner, 2003: Climate and management contributions to recent trends in u.s. agricultural yields. *Science*, **299**, 1032, doi:10.1126/science.1078475.

Lobell, D. B. and M. B. Burke, 2008: Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation. *Environ. Res. Lett*, **3**, 034 007, doi:10.1088/1748-9326/3/3/034007.

Lobell, D. B., W. Schlenker, and J. Costa-Roberts, 2011: Climate trends and global crop production since 1980. *Science*, **333**, 616–620, doi:10.1126/science.1204531.

Lobell, D. B., A. Sibley, and J. I. Ortiz-Monasterio, 2012: Extreme heat effects on wheat senescence in india. *Nature Clim. Change*, **2**, 186–189, doi:10.1038/nclimate1356.

Luo, Q., 2011: Temperature thresholds and crop production: a review. *Climatic Change*, **109**, 583–598, doi:10.1007/s10584-011-0028-6.

Meehl, G. A., T. F. Stocker, W. Collins, P. Friedlingstein, A. T. Gaye, J. M. Gregory, A. Kitoh, R. Knutti, J. M. Murphy, A. Noda, S. C. B. Raper, I. G. Watterson, A. J. Weaver, and Z.-C. Zhao, 2007: *Global climate projections. In: Climate Change 2007: The Physical Science Basis.* Cambridge University Press, Cambridge, UK.

Monfreda, C., N. Ramankutty, and J. A. Foley, 2008: Farming the planet: 2. geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Glob. Biogeochem. Cyc.*, **22**, GB1022, doi:10.1029/2007GB002947.

Nakicenovic, N. et al., 2000: *Special Report on Emissions Scenarios.* Cambridge University Press, Cambridge, UK.

Nicholls, N., 1997: Increased Australian wheat yield due to recent climate trends. *Nature*, **387**, 484–485, doi:10.1038/387484a0.

Randall, D. A., R. A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R. J. Stouffer, A. Sumi, and K. E. Taylor, 2007: *Climate Models and Their Evaluation. In: Climate Change 2007: The Physical Science Basis.* Cambridge University Press, Cambridge, UK.

Rigby, R. and D. Stasinopoulos, 2005: Generalized additive models for location, scale and shape. *J. Royal Stat. Soc.: Series C*, **54**, 507–554.

Rosenzweig, C. and F. Tubiello, 2007: Adaptation and mitigation strategies in agriculture: an analysis of potential synergies. *Mitigation and Adaptation Strategies for Global Change*, **12**, 855–873, doi:10.1007/s11027-007-9103-8.

Runge, E. C. A., 1968: Effects of Rainfall and Temperature Interactions During the Growing Season on Corn Yield. *Agron J.*, **60**, 503–507, doi:10.2134/agronj1968.00021962006000050018x.

Sakurai, G., T. Iizumi, and M. Yokozawa, 2011: Varying temporal and spatial effects of climate on maize and soybean affect yield prediction. *Clim. Res.*, **49**, 143–154, doi:10.3354/cr01027.

Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, 2004: The role of increasing temperature variability in European summer heatwaves. *Nature*, **427**, 332–336, doi:10.1038/nature02300.

Schlenker, W. and M. J. Roberts, 2009: Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *PNAS*, **106**, 15 594–15 598, doi: 10.1073/pnas.0906865106.

Semenov, M. A., R. A. C. Mitchell, A. P. Whitmore, M. J. Hawkesford, M. A. J. Parry, and P. R. Shewry, 2012: Shortcomings in wheat yield predictions. *Nature Clim. Change*, **2**, 380–382, doi:10.1038/nclimate1511.

Semenov, M. A. and P. R. Shewry, 2011: Modelling predicts that heat stress, not drought, will increase vulnerability of wheat in europe. *Sci. Rep.*, **1: 66**, doi:10.1038/srep00066.

Stott, P. A., D. A. Stone, and M. R. Allen, 2004: Human contribution to the European heatwave of 2003. *Nature*, **432**, 610–614, doi:10.1038/nature03089.

Tao, F., M. Yokozawa, Y. Xu, Y. Hayashi, and Z. Zhang, 2006: Climate changes and trends in phenology and yields of field crops in china, 1981-2000. *Ag. For. Met.*, **138**, 82 – 92, doi: 10.1016/j.agrformet.2006.03.014.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Thornton, P., P. Jones, G. Alagarswamy, J. Andresen, and M. Herrero, 2010: Adapting to climate change: agricultural system and household impacts in East Africa. *Agricultural Systems*, **103**, 73–82.

van der Velde, M., F. Tubiello, A. Vrieling, and F. Bouraoui, 2012: Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data. *Clim. Change*, **113**, 751–765, doi:10.1007/s10584-011-0368-2.

van der Velde, M., G. Wriedt, and F. Bouraoui, 2010: Estimating irrigation use and effects on maize yield during the 2003 heatwave in france. *Ag. Eco. Environ.*, **135**, 90–97, doi: 10.1016/j.agee.2009.08.017.

Wheeler, T., P. Craufurd, R. Ellis, J. Porter, and P. Prasad, 2000: Temperature variability and the annual yield of crops. *Agric. Ecosyst. Environ.*, **82**, 159–167.
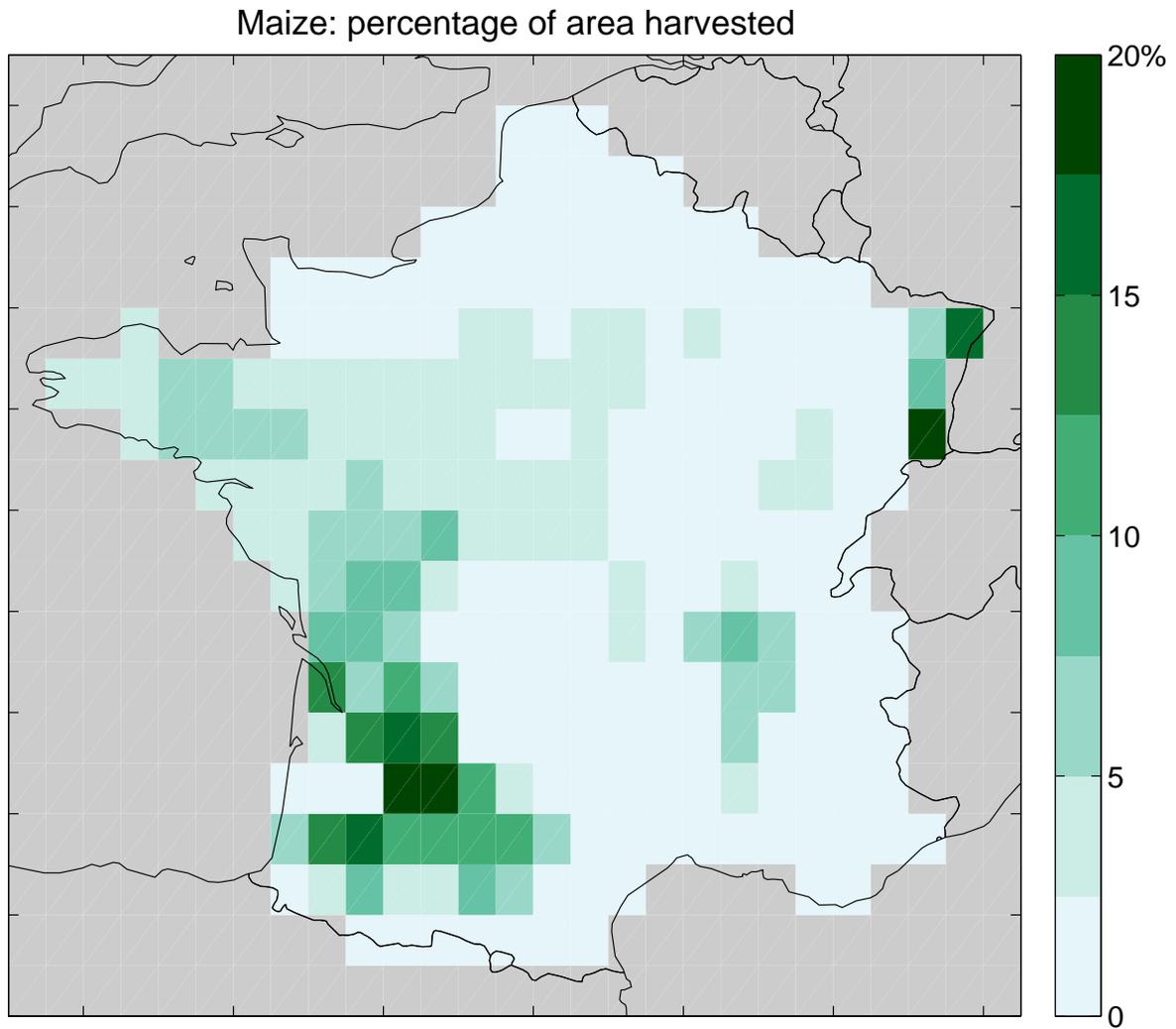
Figure 1: The percentage of land harvested for maize in France in the year 2000, using data from Monfreda et al. (2008). The France averages of hot days and precipitation shown throughout the paper are weighted using this distribution of maize growth.
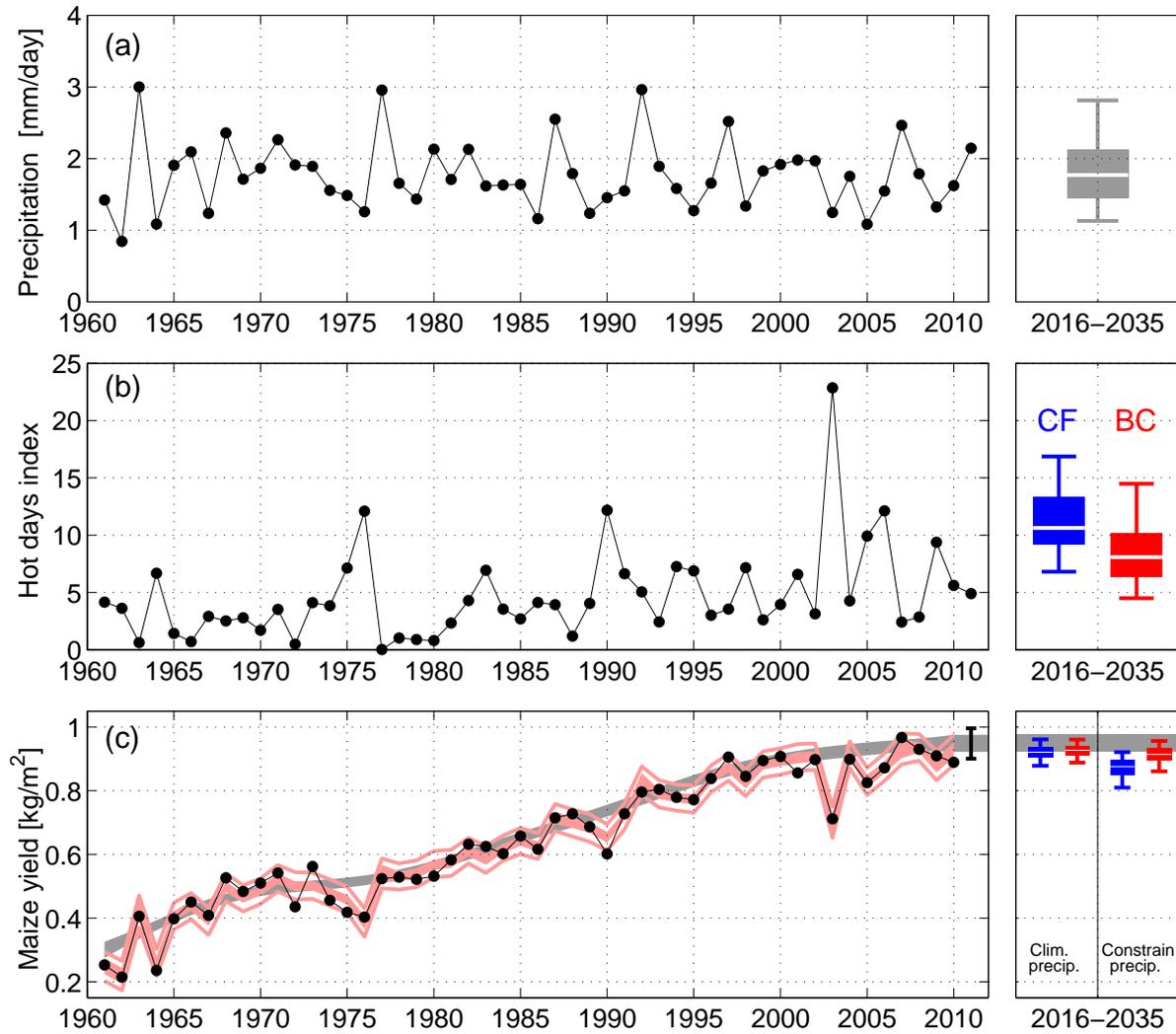
Figure 2: Historical observations and future projections of climate and maize yield for France. (a) Mean JJA precipitation, averaged over all grid-cells in France and weighted by the fraction of maize grown. (b) Number of JJA days with $T_{\max}$ over 32°C from the E-OBS dataset (Haylock et al., 2008), averaged over all grid-cells in France and weighted by the fraction of maize grown. (c) French maize yields from FAOSTAT (*black points*; http://faostat.fao.org/) and empirical model predictions for the technology trend (*grey shading*) and expected yield (*red shading*) with total uncertainties (*red lines*), using Eqn. 1 and considering both temperature and precipitation. The black errorbar indicates the forecast for 2011, assuming a flat technology trend since 2010. For the 2016-2035 periods, the boxes show the 25th-75th percentiles and the whiskers indicate the 5th and 95th percentiles. The climatological distribution for JJA precipitation is shown (top), along with the projected and calibrated number of hot days using bias correction (BC) and change factor (CF) methods (middle). The yield projections (bottom) assume a flat technology trend and are shown for both climatological precipitation and precipitation constrained by historical correlations between temperature and precipitation.

22
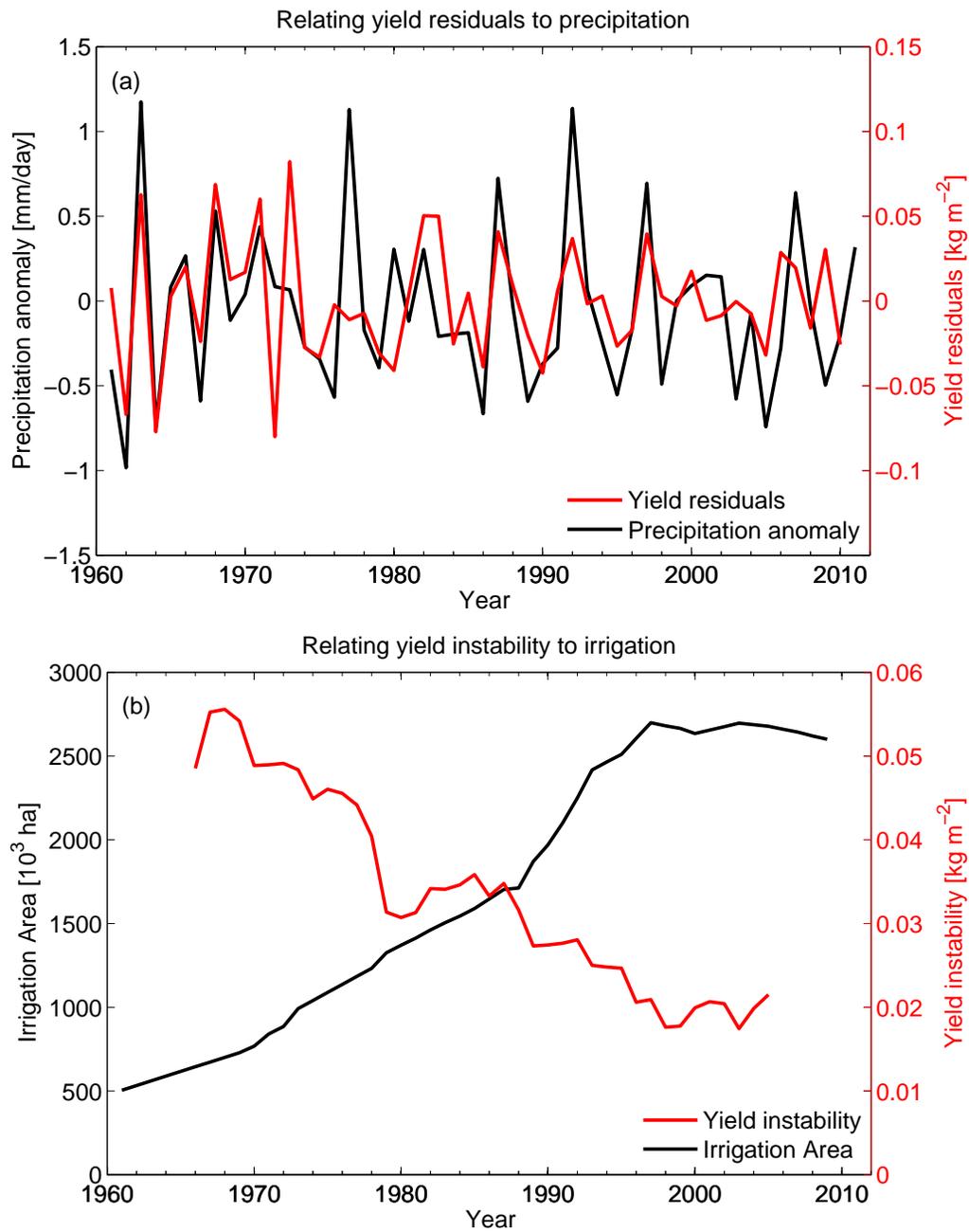
Figure 3: The temperature only empirical model. (a) The residuals from the expected yield (Eqn. 4) as a function of time (red) and precipitation anomaly (black). (b) The running 11-year standard deviation of yield residuals (red), showing a decrease over time, indicating higher yield stability, and the area of irrigation in France from FAOSTAT (black).
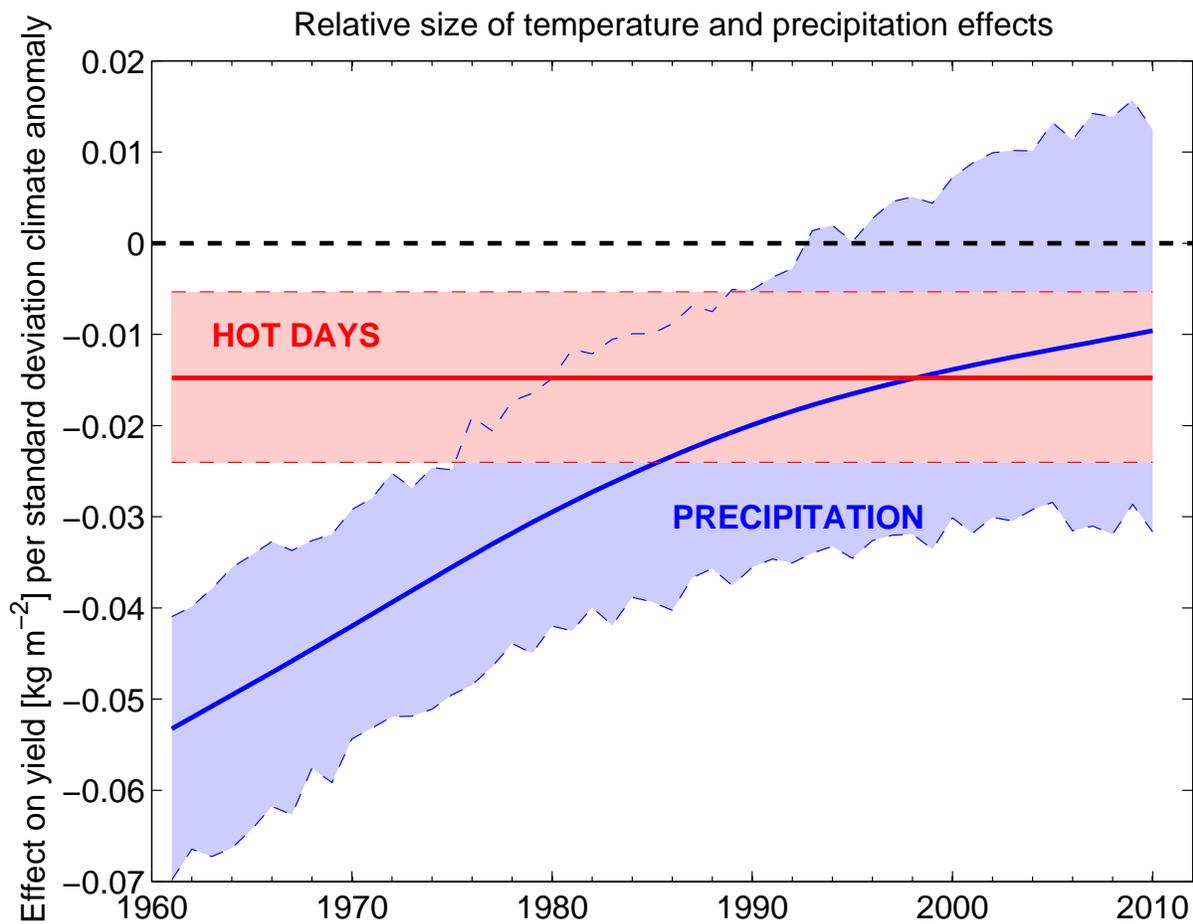
Figure 4: The relative importance of temperature and precipitation on yield, expressed per standard deviation anomaly, assuming median values for the other climate variable, with associated uncertainties. Both effects are shown producing a reduction in yield, i.e. increasing hot days and decreasing precipitation.
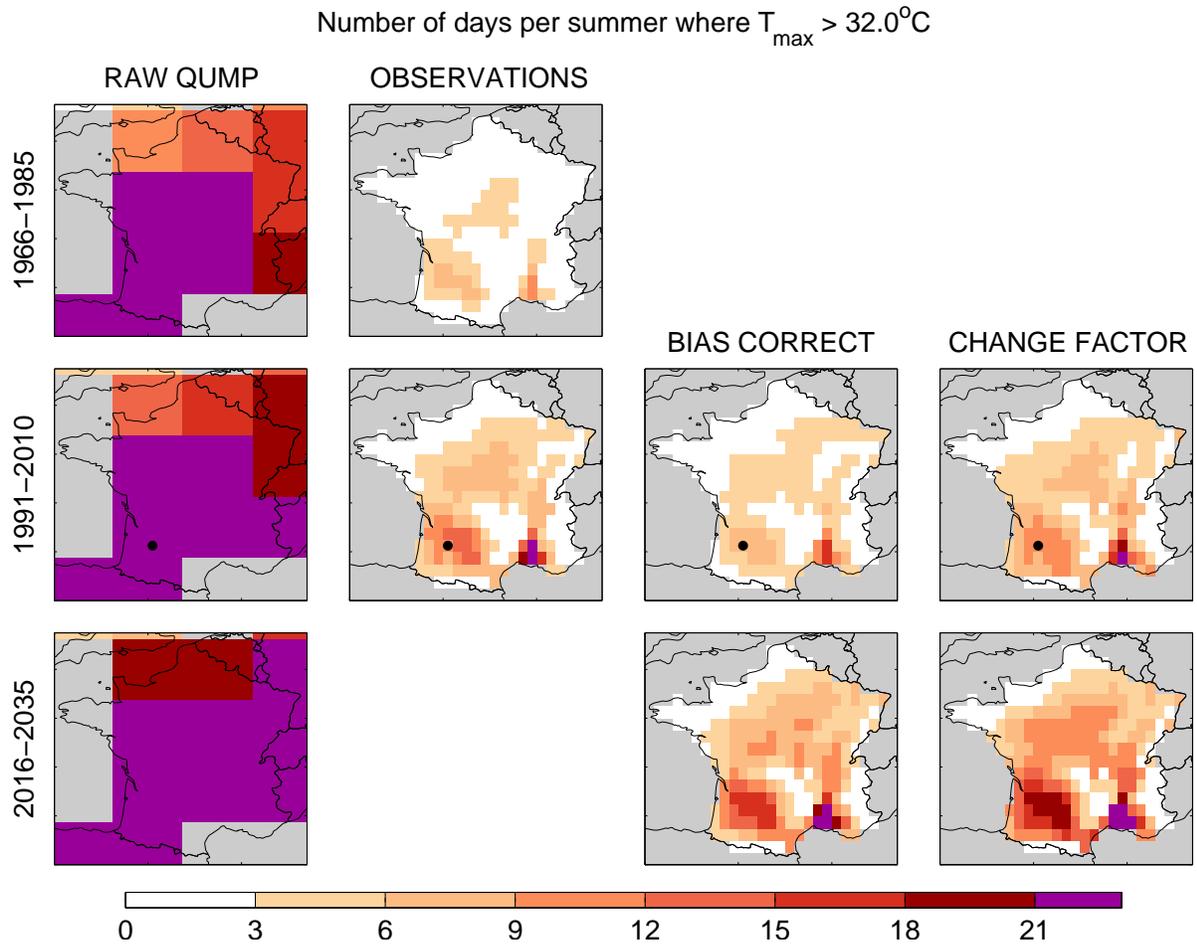
Figure 5: Mean number of hot days in France from the raw QUMP ensemble (left) and E-OBS observations (second column), for various time periods. The mean number of hot days are shown for the out-of-sample prediction of 1991-2010 (second row) and of the future 2016-2035 period (third row) after applying bias correction calibration (third column) and change factor calibration (right column), including corrections to daily temperature variability, to each QUMP ensemble member separately.
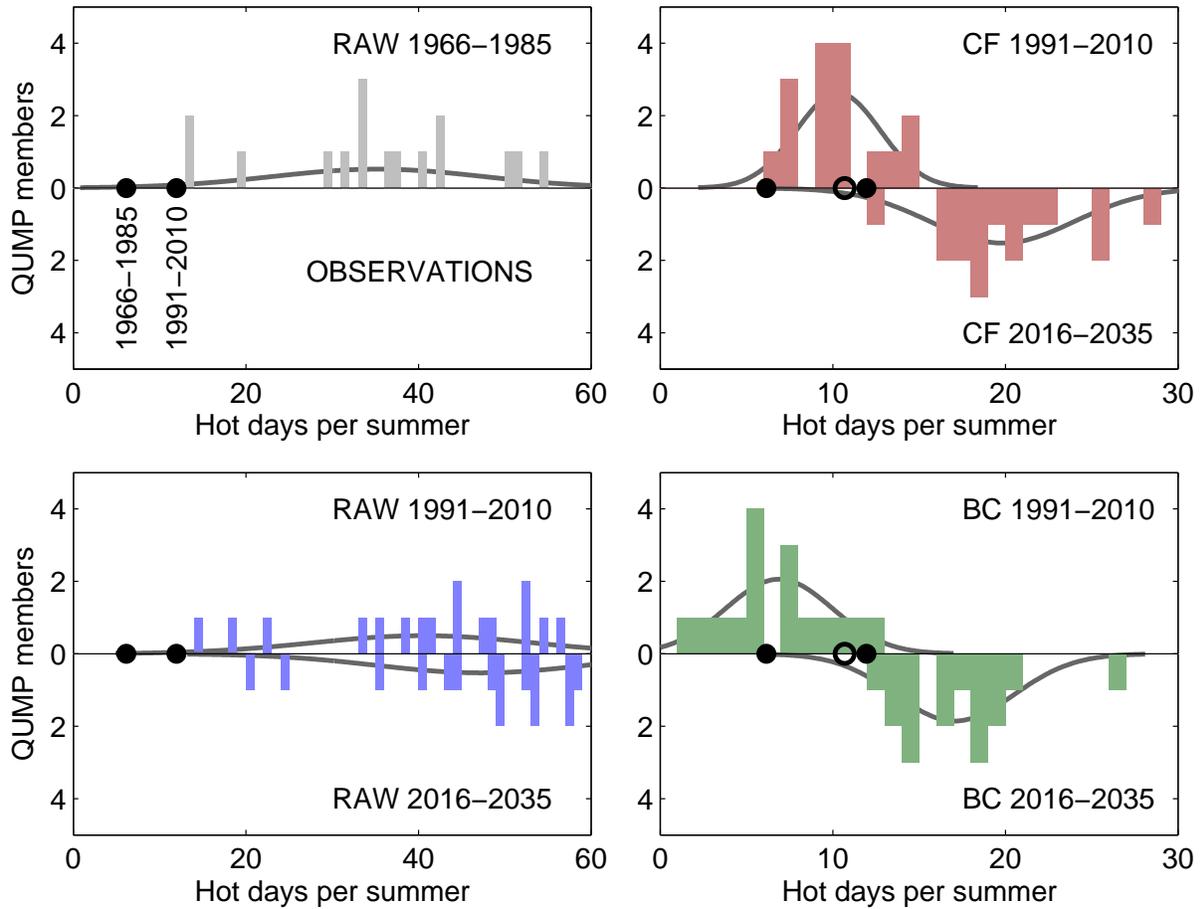
Figure 6: Histograms of 20-year mean projections for the number of hot days for a particular grid point in south-west France (black dot in Fig. 5). Left column: raw QUMP output for reference period (top) and calibrated periods (bottom). Observations are shown in each panel for both reference periods. Right column: Calibrated projections using bias correction (BC, bottom) and change factor (CF, top) for two time periods. The dark grey lines indicate an estimate for a normal distribution from which the QUMP members are drawn. The open black circle indicates the observations for 1991-2010 without 2003. Note the different range on the x-axis for the raw and calibrated projections (columns).

Increasing influence of heat stress on French maize yields from the 1960s to the 2030s

Ed Hawkins[*1], Thomas E. Fricker[2], Andrew J. Challinor[3], Christopher A. T. Ferro[2,4], Chun Kit Ho[1], and Tom M. Osborne[1]

[1]NCAS-Climate, Department of Meteorology, University of Reading, Reading, UK
[2]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK
[3]Institute for Climate and Atmospheric Science, University of Leeds, Leeds, UK
[4]NCAS-Climate, University of Exeter, Exeter, UK

SUPPLEMENTARY INFORMATION

This supplementary information augments the analysis presented in the main paper published in *Global Change Biology*.

Observed climate

Fig. S1 shows the number of days with a maximum daily temperature ($T_{\max}$) exceeding 32°C over France for each year from 1950 to 2011, according to the E-OBS dataset (Haylock et al., 2008). There is considerable spatial and temporal variability in this quantity, with 1977 and 2003 being two opposite extremes. Fig. S2 shows similar features for mean JJA precipitation, with different extreme years (1962, 1976, 2005).

Validation of the empirical model

Eqn. 1 describes the statistical model for maize yield, using a cubic regression spline for the smooth functions g, h and β_2. The parameters of the model (the β parameters, and the smoothing parameters in g and h) are all estimated by maximising a penalised likelihood function, performed using the gamlss() package (Rigby and Stasinopoulos, 2005) in R (R Development Core Team, 2010). To estimate out-of-sample predictive power we use leave-one-out cross validation (LOOCV) in which we sequentially omit a year from the data, fit the model using the remaining data, and predict the omitted year. The optimal threshold temperature for the variable X (either hot day count or integrated temperature-days) is found by minimising the mean squared error of the LOOCV predictions.

Yield model for temperature alone

Fig. S3a shows the value of the LOOCV RMSE when considering temperature only ($\beta_2 = \beta_3 = 0$), and demonstrates that 32°C is the optimal choice using the hot days count, and 26.5°C is optimal using integrated-temperature days. We choose to use the hot day count for simplicity, and also

---

[*]e.hawkins@reading.ac.uk

because we have more confidence that the QUMP simulations can represent this quantity more robustly (see later).

Tables S1, S2 give the best fit parameters, associated confidence ranges and validation statistics for the empirical model. The effect of temperature on yield is summarised by β_1, which suggests a 1% drop in yield from 2010 levels per hot day over 32°C.

We further examine the predictive power of the empirical yield model using a one-step-ahead cross validation method, designed to test the model's ability to predict outside of the range of its fitting data. Here, the response parameter (β_1) is fitted only using years which have *fewer* hot days than the year under test. The fitted parameter is then used to predict the yield response to the hotter year. Fig. S3b demonstrates that the predictions and the estimated uncertainties are robust for predicting the response to out-of-sample years which always have more hot days. Note especially that the yield anomaly in the extreme summer of 2003 is well predicted.

Yield model including precipitation

When considering temperature and precipitation, the optimal choice for the hot days count is again 32°C, and for integrated crop degree days it is 28°C (Fig. S3a). Tables S1, S2 give the best fit parameters, associated confidence ranges and validation statistics for this version of the empirical model. Again, we choose to use the hot day count as it produces a smaller LOOCV RMSE, and also because we have more confidence that the QUMP simulations can represent this quantity more robustly (see later).

As for the yield model for temperature alone, we examine the predictive power of the empirical yield model using a one-step-ahead cross validation method, designed to test the model's ability to predict outside of the range of its fitting data. Here, the response parameters are fitted only using years which have *fewer* hot days (or *less* precipitation) than the year under test. The fitted parameter is then used to predict the yield response to the hotter (or wetter) year. Figs. S3c,d demonstrate that the predictions and the estimated uncertainties are robust for predicting the response to out-of-sample years. Note especially that the yield anomaly in the extreme summer of 2003 is well predicted (Fig. S3c).

Yield model validation - further checks

Note that the version of the empirical model assuming a linear trend has significantly poorer cross-validation statistics (Table S2) than the non-linear version. Finally, the standard version of the empirical model (Eqn. 1) is superior to other versions with constant or time-varying β parameters, according to the Akaike Information Criterion (AIC, Akaike, 1974) (Table S3).

Comparing climate model variability with observations

Fig. S4 shows quantile-quantile plots for the observations and QUMP simulations for JJA $T_{\max}$ in 1991-2010 for a particular grid point in south-west France (black dots in Fig. 5). The thick black lines represent the linear calibration between the QUMP and observational data. The non-linearity seen in the tails of the distributions on both sides indicates that the shape of the temperature distributions are different between the QUMP simulations and the observations. This difference could make calibrated projections of $T_{\max}$ biased.

However, the QUMP simulations are close to the correct distribution up to the 32°C threshold (dashed lines) in all members after linear calibration. As the hot day index simply counts numbers of days over the threshold, and ignores how far above the threshold the temperature is, the

errors in shape of the variability at higher temperatures does not affect this particular metric. Although both an integrated temperature metric and hot day index might be justified physically and empirically (Fig. S3a), the integrated temperature metric would be biased by these differences in shape, so should not be used in this case.

In addition, Fig. S5 is similar to Fig. 5, but does not include a correction for the daily temperature variability in Eqns. 2 & 3, i.e. it uses the standard variants of BC and CF. It can be seen that the BC method in particular does not produce robust estimates of the number of hot days if the variability is not corrected.

Finally, Fig. S6 is an annual version of Fig. 6, showing each individual year in each of the QUMP ensemble members. For bias correction, the shape of the annual distribution does not match the observed distribution particularly well, giving us less confidence in its use for annual values. However, for the 20 year means considered in the paper, the method performs adequately.

These findings demonstrate the need to carefully examine the characteristics of climate variability in any model simulations before using them.

Future yield projections

When producing future projections of maize yield we assume that the present climatological distribution of precipitation (estimated using data from 1961-2010) will not change (Fig. 2). We make two different assumptions about the relationship between temperature and precipitation: first that temperature and precipitation are independent, and second that the observed correlation between precipitation and temperature remains constant in the future period. These cases are illustrated in Fig. S7, which shows the relationship between temperature and precipitation in the observations (white circles), and Monte-Carlo realisations of the future distribution under the first assumption (green) and under the second assumption (blue). The coloured contours represent the yield deficit (which is defined as the difference between the actual yield and the base level of yield due to technology). We see that, under both calibration methods, assuming that the current temperature-precipitation correlation remains constant in the future results in a stronger shift towards large deficits than does assuming independence of temperature and precipitation.

Regional yield data

To examine the potential for using smaller scale yield data we fit our empirical yield model (Eqn. 1) to the yield and climate data from two NUTS2 (Nomenclature of Territorial Units for Statistics) regions in France (Fig. S8) where large fractions of the land is harvested for maize (Fig. 1). The maize yield data is taken from EUROSTAT (http://ec.europa.eu/eurostat/) and is only available from 1980-2007.

The E-OBS weather data (shown in Figs. S1, S2) is averaged over the same NUTS2 regions. Note that the weather can be very different in the two regions. For example, 1990 produced a large number of hot days in the southern region, but not the northern region, and 1992 was very wet in the southern region, but average in the northern region. In 1987 the rainfall anomalies were reversed.

The same empirical model is fit to both sets of regional data (Fig. S8), and produces residuals which are consistent with the model assumptions. However, the shorter time series means that the uncertainties are much larger and robust statements about trends in the effects sizes are not possible. In this instance, the longer temporal information is more important than the smaller spatial information.

References

Akaike, H., 1974: A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723, doi:10.1109/TAC.1974.1100705.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New, 2008: A European daily high-resolution gridded dataset of surface temperature and precipitation. *J. Geophys. Res.*, **113**, D20 119, doi:10.1029/2008JD10201.

R Development Core Team, 2010: *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org`, ISBN 3-900051-07-0.

Rigby, R. and D. Stasinopoulos, 2005: Generalized additive models for location, scale and shape. *J. Royal Stat. Soc.: Series C*, **54**, 507–554.

Table S1: Best-fit and uncertainty ranges of parameters (Eqn. 1) for the two different versions of the empirical model, with and without including precipitation. The $5-95\%$ confidence levels are given in brackets.

	Empirical model version	
Parameter	Temperature only	Temperature and Precipitation
β_1	-0.0101 (-0.0120 to -0.0087)	-0.0277 (-0.0216 to -0.0379)
β_3	0	0.0143 (0.0100 to 0.0210)
h	varying	0.0235 (0.0191 to 0.0280)

Table S2: Validation statistics for maize yield (Eqn. 1) for the different versions of the empirical model, with and without including precipitation, and the full model with a linear $g(t)$.

	Empirical model version		
Statistic	Temperature only	Temperature and Precipitation	Linear $g(t)$
Correlation with trend	0.984	0.993	0.983
Correlation without trend	0.762	0.899	0.819
R^2 with trend	0.969	0.986	0.966
R^2 without trend	0.581	0.841	0.695
LOOCV RMSE [$\mathrm{kg\,m^{-2}}$]	0.050	0.034	0.043

Table S3: Akaike Information Criterion (AIC, Akaike, 1974) values for different versions of the maize yield empirical model (Eqn. 1) with constant or time-varying parameters for β_1 and β_2. More negative values of AIC indicate a superior model. The standard version of the model (Eqn. 1) is shown in bold, and this has the lowest AIC.

β_1	β_2	AIC
Constant	Constant	-196.5
Constant	**Time-varying**	**-202.2**
Time-varying	Constant	-195.9
Time-varying	Time-varying	-201.9
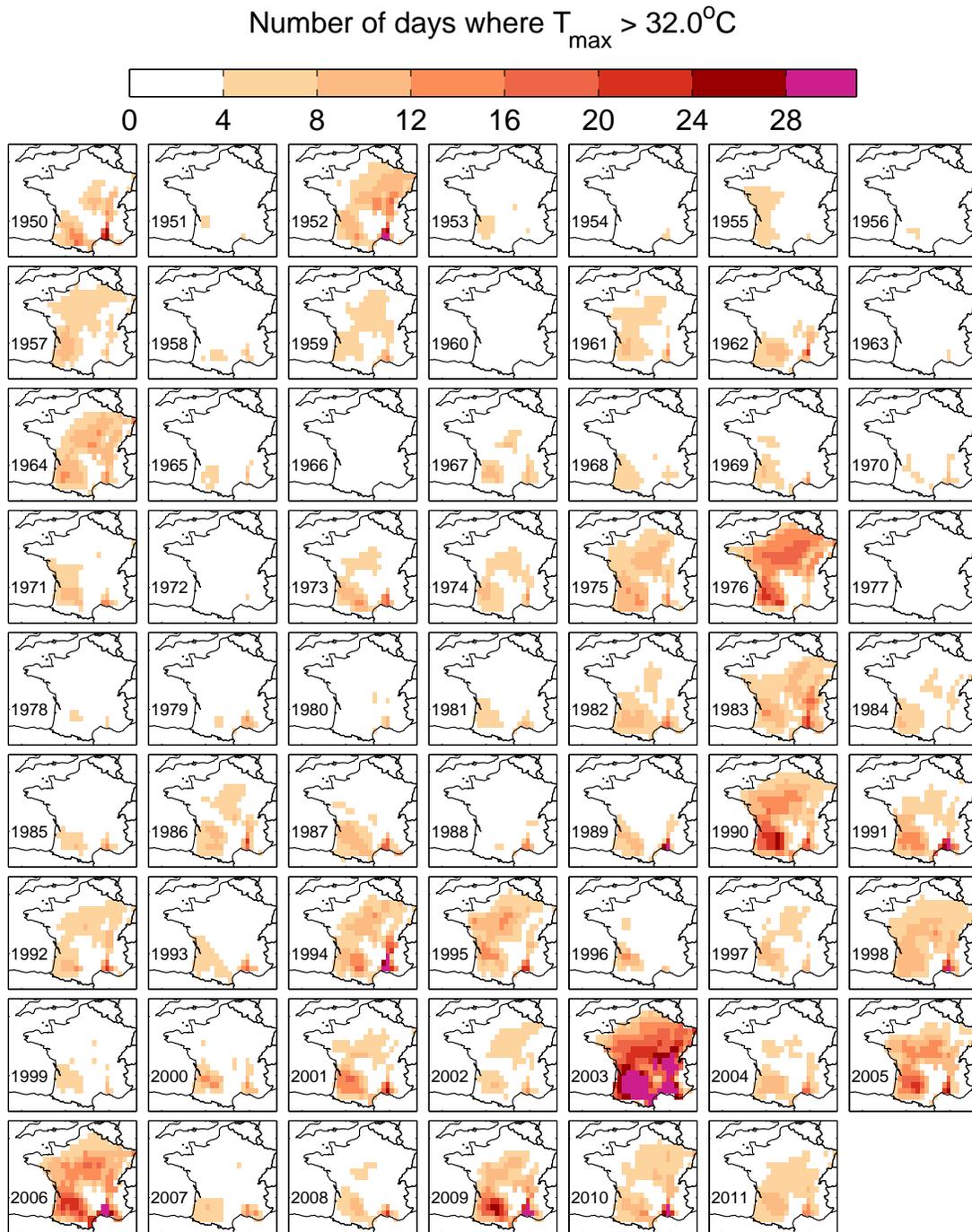
Figure S1: The number of days in each summer exceeding 32°C over France, from the E-OBS v5.0 dataset (Haylock et al., 2008).
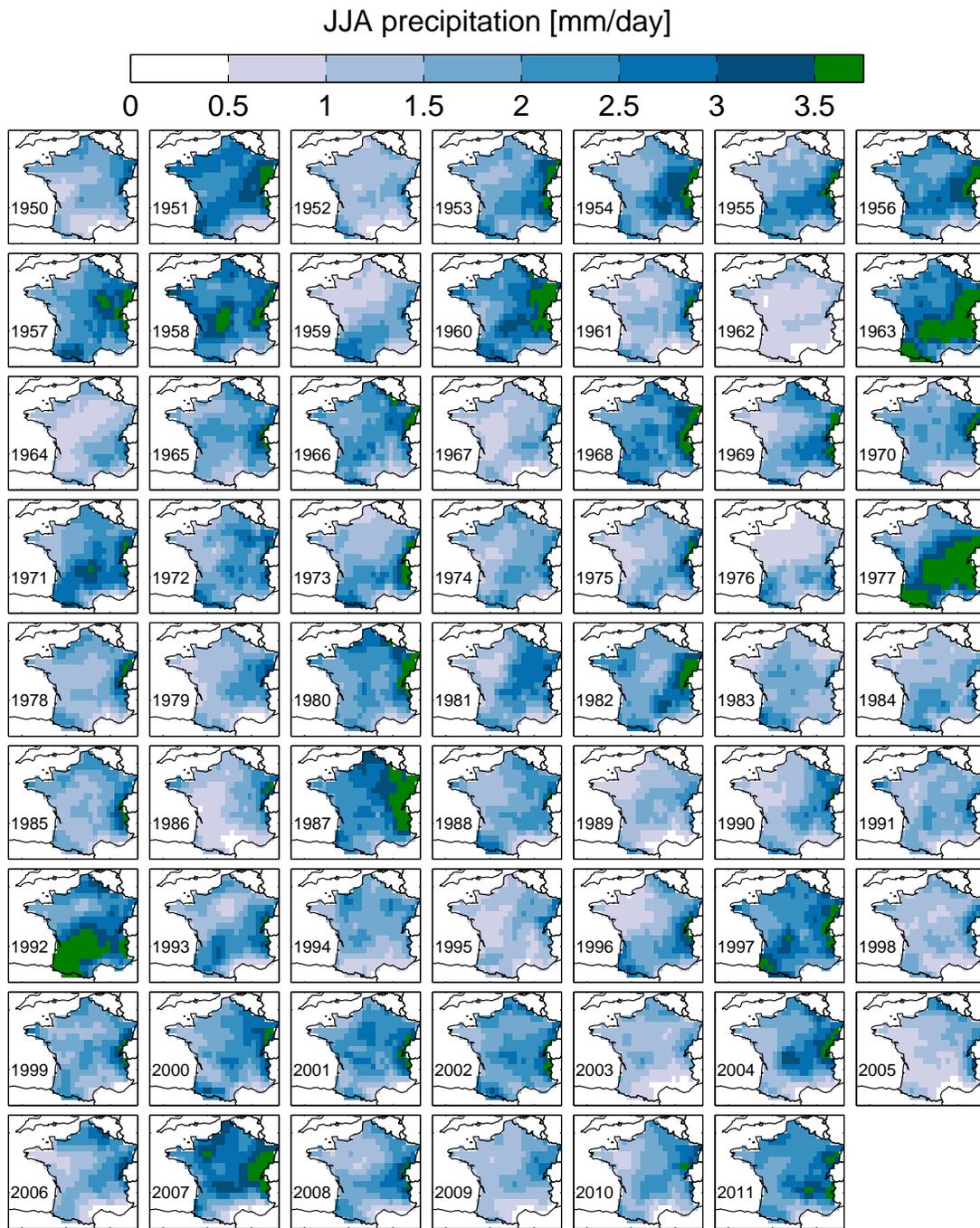
Figure S2: The mean summer (JJA) precipitation over France, from the E-OBS v5.0 dataset (Haylock et al., 2008).
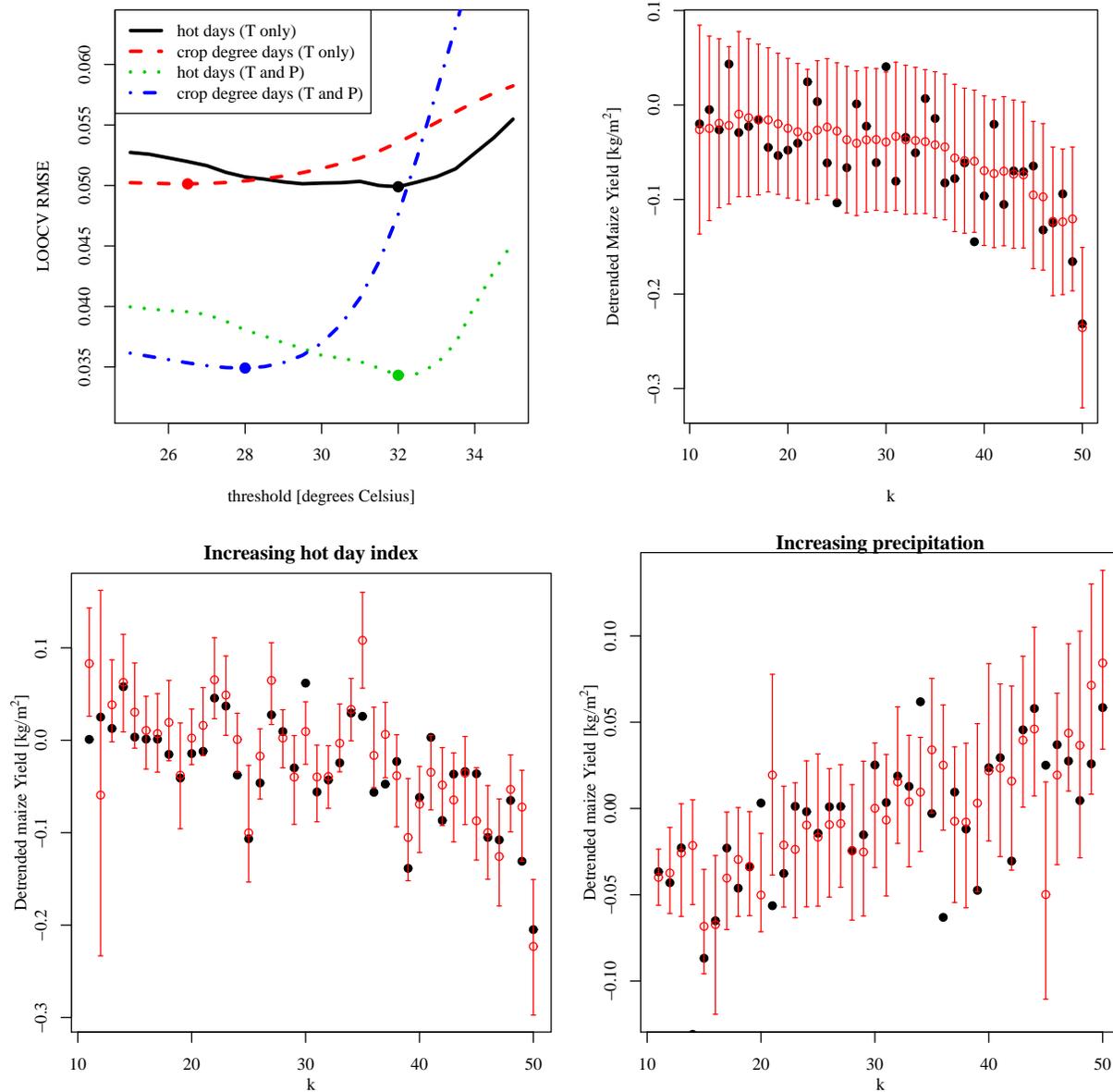
Figure S3: **Testing the empirical model.** (top left) Validation statistics when fitting the empirical crop yield models using leave-one-out cross-validation. (top right) Testing the temperature only empirical model in predicting the response to higher temperatures (one-step-ahead). (bottom) Testing the full empirical model in predicting the response to higher temperatures (left) and precipitation (right). k is an index of years of increasing temperature or precipitation. The black dots are the actual yield, and the red error bars are the out-of-sample predictions of the empirical model.

Figure S4: Quantile-quantile diagnostics for daily $T_{\max}$ for JJA in 1991-2010, for the E-OBS dataset and each QUMP member as labelled, for a particular location (the black dots in Fig. 5). The solid black line in all panels represents the linear calibration, and the dashed black line is at 32°C, the threshold for hot days.

9
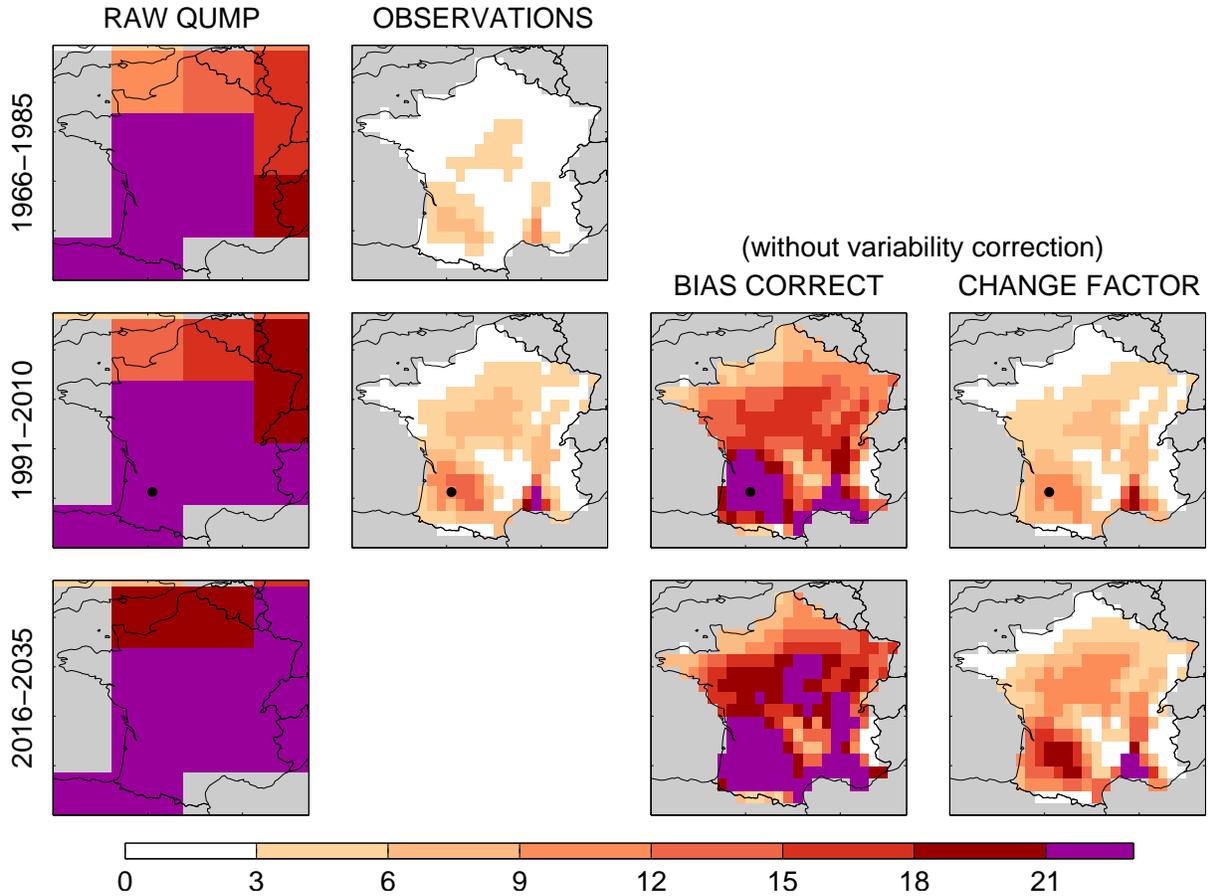
Figure S5: As Fig. 5, but without correcting the daily temperature variability.
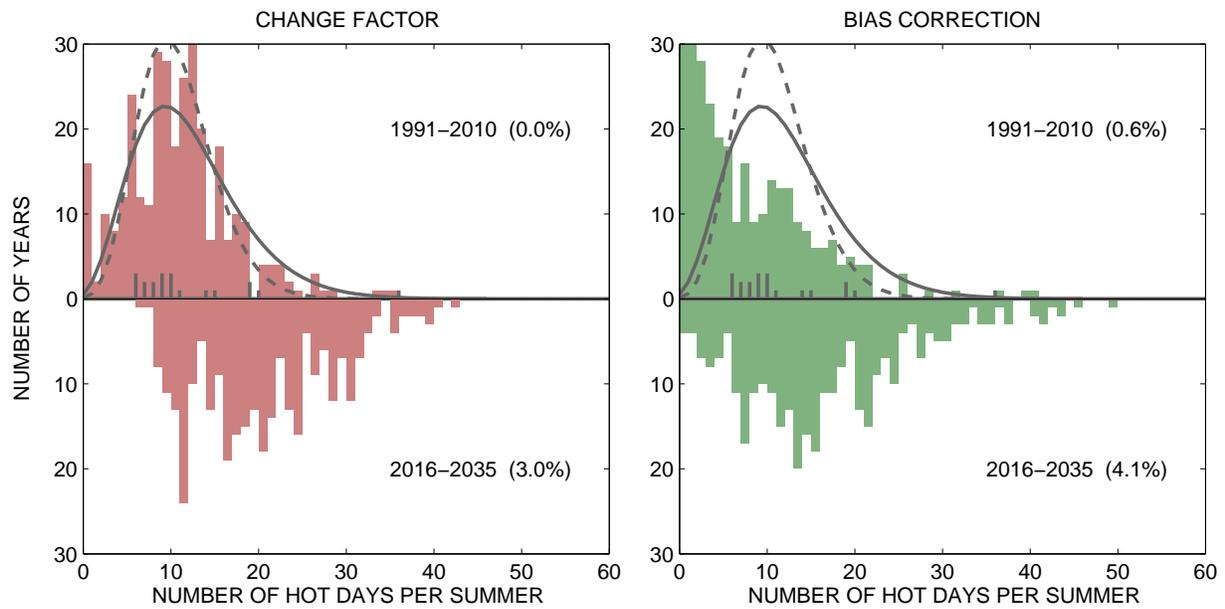
Figure S6: Histograms showing number of years across all QUMP members of annual calibrated projections in 20 year periods of the number of hot days for a particular grid point in south-west France (black dot in Fig. 5). The observations are shown with the grey histogram, and the grey lines show a negative binomial distribution fit to the observations with (solid) and without (dashed) 2003. The labelled percentages indicate the suggested probability of exceeding the number of hot days (44) seen in 2003.
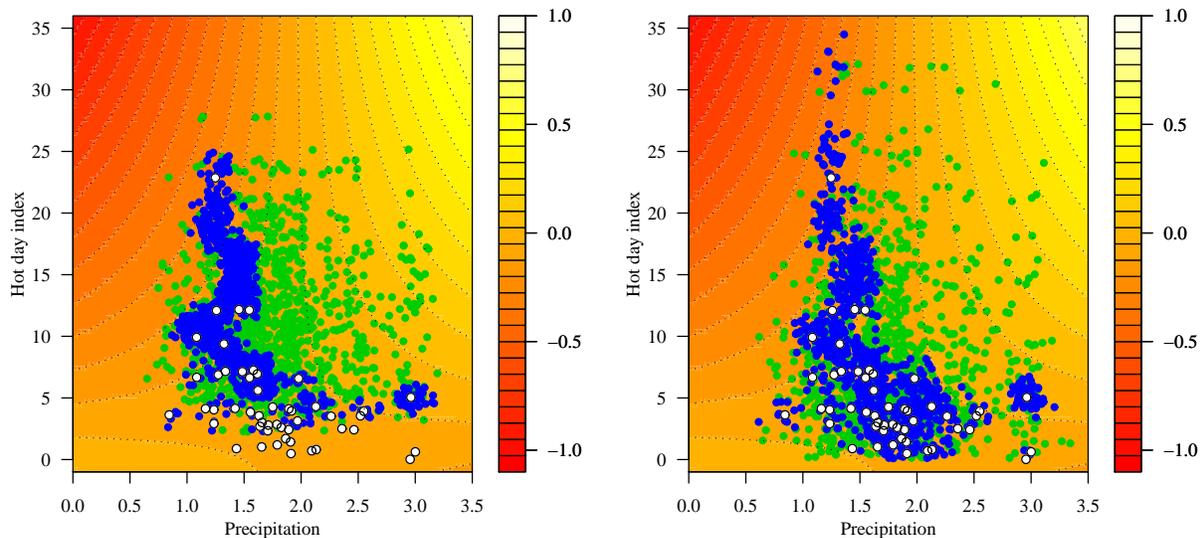
Figure S7: The relationship between temperature and precipitation and the effect on yield. The observations (white circles) are compared with predictions assuming no relationship (green) and constant correlation (blue) for CF (left) and BC (right) calibrations. The filled contours represent the yield deficit [kg m^{-2}].

11
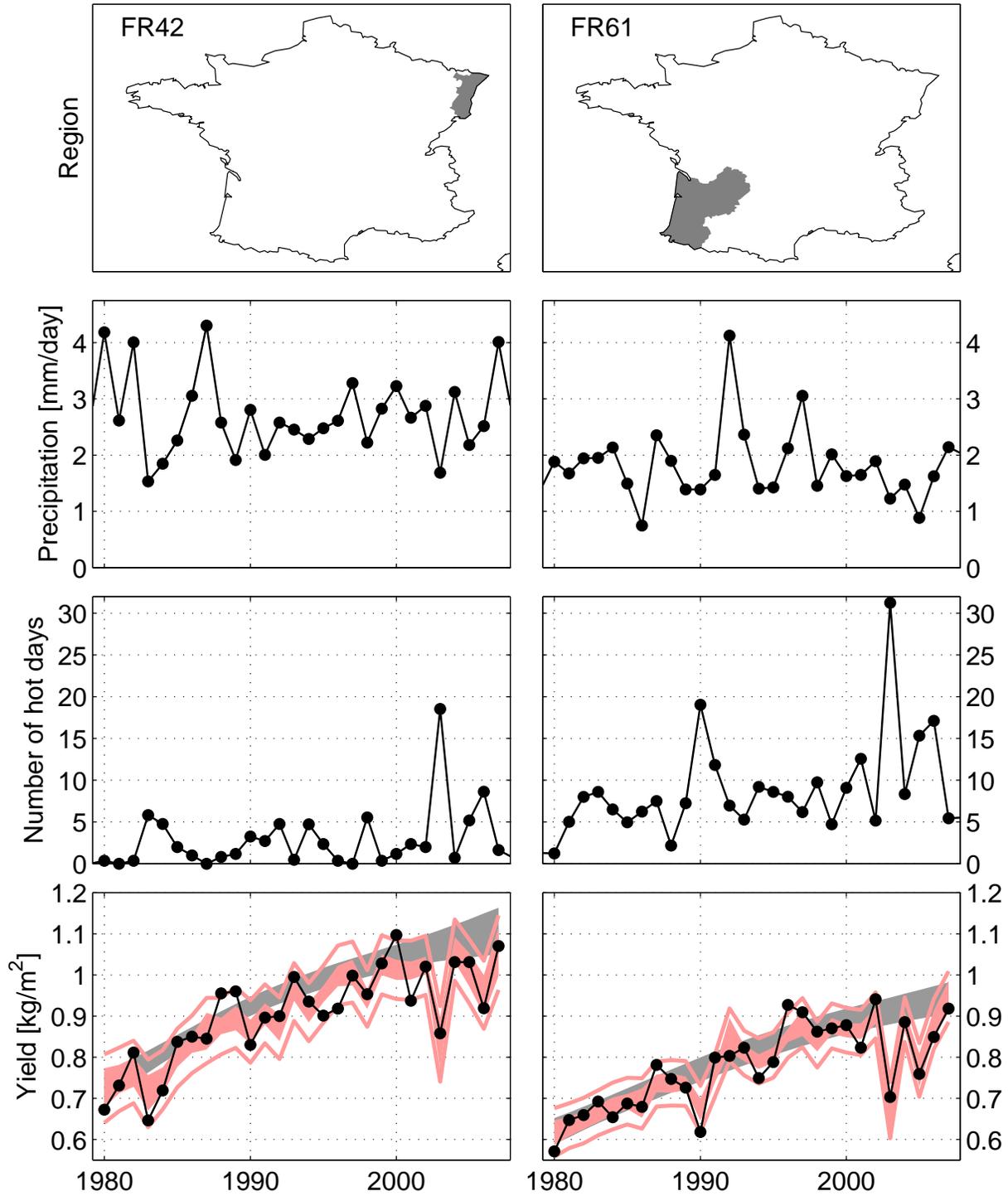
Figure S8: The empirical yield model fitted to two regions of France where a large fraction of the area is harvested for maize (Fig. 1). The maize yield data for these NUTS2 regions is only available from 1980-2007.