Utility of dynamical seasonal forecasts in predicting crop yield

Mikhail A. Semenov[1,*], Francisco J. Doblas-Reyes[2]

[1]Biomathematics & Bioinformatics, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK
[2]ECMWF (European Centre for Medium-range Weather Forecasting), Shinfield Park, Reading RG2 9AX, UK

ABSTRACT: Advance predictions of crop yield using crop simulation models require daily weather input for the whole growing season. Seasonal forecasts, based on coupled ocean–atmosphere climate models, are now available up to 6 mo in advance from a number of operational meteorological centres around the world. Seasonal forecasts are not directly suitable for crop simulations, because of model biases and mismatch of spatial and temporal scales. However, it is possible to utilise seasonal forecasts for yield predictions by constructing site-specific daily weather using a stochastic weather generator linked to seasonal forecasts. In our study, we use the LARS-WG weather generator and a subset of predictions by DEMETER (Development of a European Multimodel Ensemble system for seasonal to inTERannual climate prediction), i.e. seasonal ensemble hindcasts from the general circulation model (GCM) of ECMWF (European Centre for Medium-range Weather Forecasting) for 1980–2001. To assess the value of seasonal forecasts, 2 sets of scenarios were created, one based on seasonal forecasts and the other on historical climatology. The Sirius wheat simulation model was used to compute distributions of wheat yield at 2 locations in Europe and New Zealand. The main conclusion is that the use of dynamical seasonal forecasts at selected sites has not improved yield predictions compared with the approach based on historical climatology. The likely reason is that for dynamic seasonal forecasts, the skill score for temperature and precipitation is generally low for latitudes higher than 30° for northern and southern hemispheres, and our test locations are at 47.6° N and 43.6° S.

KEY WORDS: ECMWF GCM · General circulation model · Stochastic weather generator · LARS-WG · Wheat simulation model · Sirius · Probabilistic ensemble

1. INTRODUCTION

A methodology to predict grain yield with incomplete weather data, based on the use of the Sirius crop simulation model (Jamieson et al. 1998b, Jamieson & Semenov 2000, Lawless et al. 2005) and the LARS-WG stochastic weather generator (Semenov & Barrow 1997, Semenov et al. 1998, Semenov & Brooks 1999), was described in Lawless & Semenov (2005). Prediction of grain yield using a crop simulation model requires weather input for the growing season. When observed weather is available only for the beginning of the growing season, then 'synthetic weather' should be generated for the remaining part of the season. The approach described in Lawless & Semenov (2005) sug-

gested use of a stochastic weather generator for this purpose. A stochastic weather generator, calibrated for a selected site using historical weather, is capable of generating synthetic daily weather that is statistically similar to observed weather. This means that a range of statistical tests will show no significant differences between historical and synthetic weather. Using the generated ensemble as an input to a crop simulation model, the site-specific distribution of the final grain yield and other crop characteristics can be obtained.

Weather generator time series are based exclusively on the analysis of historical data and therefore incorporate substantial uncertainty associated with historical climatology. However, availability of specific information about future weather for the current growing

*Email: mikhail.semenov@bbsrc.ac.uk

season is likely to bring improvement in yield predictions. One such possibility explored in this paper is the use of seasonal forecasts. Seasonal forecasts, based on coupled ocean–atmosphere climate models, are now available at a number of operational meteorological centres around the world (Palmer et al. 2004a, Saha et al. 2006). Because signals arising from atmosphere–ocean coupling processes are intrinsically chaotic and sensitive to initial conditions (Palmer 1993), an ensemble of forecasts with individual members differing by small perturbations to the starting conditions of the atmosphere and oceans is used (Palmer 2001, Palmer et al. 2005). Seasonal forecasts are not suitable for direct use in crop simulation models because of model biases and mismatch of spatial and temporal scales (Challinor et al. 2005, Doblas-Reyes et al. 2006, Ines & Hansen 2006). In order to be used in yield predictions, seasonal forecasts require downscaling to a resolution suitable for crop simulation models. Downscaling can be achieved by linking a seasonal forecast with a weather generator in a fashion similar to that described by Barrow & Semenov (1995), Wilks (2002), Hansen & Indeje (2004), and Feddersen & Andersen (2005).

The aim of our study was to investigate whether seasonal forecasts can improve yield predictions in areas where wheat is commonly grown. We selected 2 sites for the study: Debrecen, Hungary and Lincoln, New Zealand. The choice of sites was dictated by several requirements: (1) sites should represent areas where wheat is a major crop; (2) both models, Sirius and LARS-WG, must be well validated for these sites; (3) seasonal forecasts should demonstrate reasonable performance at these locations; and (4) daily observed weather should be available for the period 1979–2001. All these requirements were satisfied for the selected sites. The Sirius wheat simulation model and the LARS-WG weather generator have been validated for both sites (Jamieson et al. 1998b, Semenov et al. 1998, Jamieson et al. 2000, Jamieson & Semenov 2000, Lawless et al. 2005). The European Centre for Medium-range Weather Forecasting General Circulation Model (ECMWF GCM) was used for seasonal dynamical forecast in our study and its performance at the 2 selected sites was comparable with performance in neighbouring regions. Mean Square Skill Score (MSSS) of the ECMWF GCM model has values for 2 m temperature between –0.1 and 0.1 in the Debrecen region, which are similar or higher than values in most regions of Western Europe. (www.ecmwf.int/research/demeter/d/charts/verification). The MSSS values for 2 m temperature in Lincoln are higher than in Debrecen, between 0.1 and 0.2. MSSS values for precipitation are low for Western Europe and New Zealand, ranging between –0.8 and –0.6.

2. PERFORMANCE OF SEASONAL FORECASTS

We used a subset of seasonal predictions from DEMETER, i.e. a series of seasonal ensemble predictions from the ECMWF GCM model (Palmer et al. 2004b, Palmer et al. 2005). Each forecast ensemble consisted of 9 members, each 6 mo long. Ensembles were produced 4 times per year starting on 1 February, 1 May, 1 August and 1 November retrospectively for 1980–2001. The spatial resolution of the ECMWF atmospheric model was $1.8 \times 1.8°$ (latitude $\times$ longitude). Predictions consist of monthly anomalies of climatic variables with respect to 1980–2001 modelled averages.

We used 3 statistics to compare the observed meteorological variables with the seasonal forecasts: anomaly ranges, temporal correlation and root mean squared error (RMSE). For a seasonal forecast, an anomaly range $AR_{m,y}$ was defined for a selected month m and a selected year y, as the difference between maximum and minimum values within an ensemble

$$AR_{m,y} = (\max - \min)\{A^k_{m,y}, k \leq N\} \qquad (1)$$

where $A^k_{m,y}$ is the kth ensemble member for the month m and the year y and N is the size of ensemble (in our case $N = 9$). For observed weather only one 'realization' of weather is available for each year and the observed range OR_m is defined for a selected month m as a difference between monthly maximum and minimum values over the whole range 1980–2001:

$$OR_m = (\max - \min)\{O_{m,y}, 1980 \leq y \leq 2001\} \qquad (2)$$

where $O_{m,y}$ is a monthly anomaly for the month m and the year y. We should emphasize that model anomalies were computed with respect to the model ensemble mean and observed anomalies with respect to the mean observed weather for 1980–2001. Observed ranges represent uncertainty in the historical climate; model anomaly ranges represent uncertainty in seasonal forecasts associated with imprecise knowledge of initial boundary conditions and model parameters. We should expect that a skilful seasonal forecast for a given year will reduce the uncertainty embedded in historical climatology, i.e. $AR_{m,y} \leq OR_m$.

The temporal correlation was defined as a correlation between the ensemble mean of model anomalies and observed monthly anomalies for the 6 mo forecast period. In a similar way, RMSE was calculated between the ensemble mean of model anomalies and observed monthly anomalies for the 6 mo forecast period.

We calculated the above statistics for monthly Tmax and monthly total precipitation at 2 locations, Debrecen and Lincoln. Observed ranges over the period of 1980–2001 in total monthly precipitation (Fig. 1a,c) and ranges in mean monthly Tmax (Fig. 2a,c) are presented for the 6 mo from February to July. We calculated model
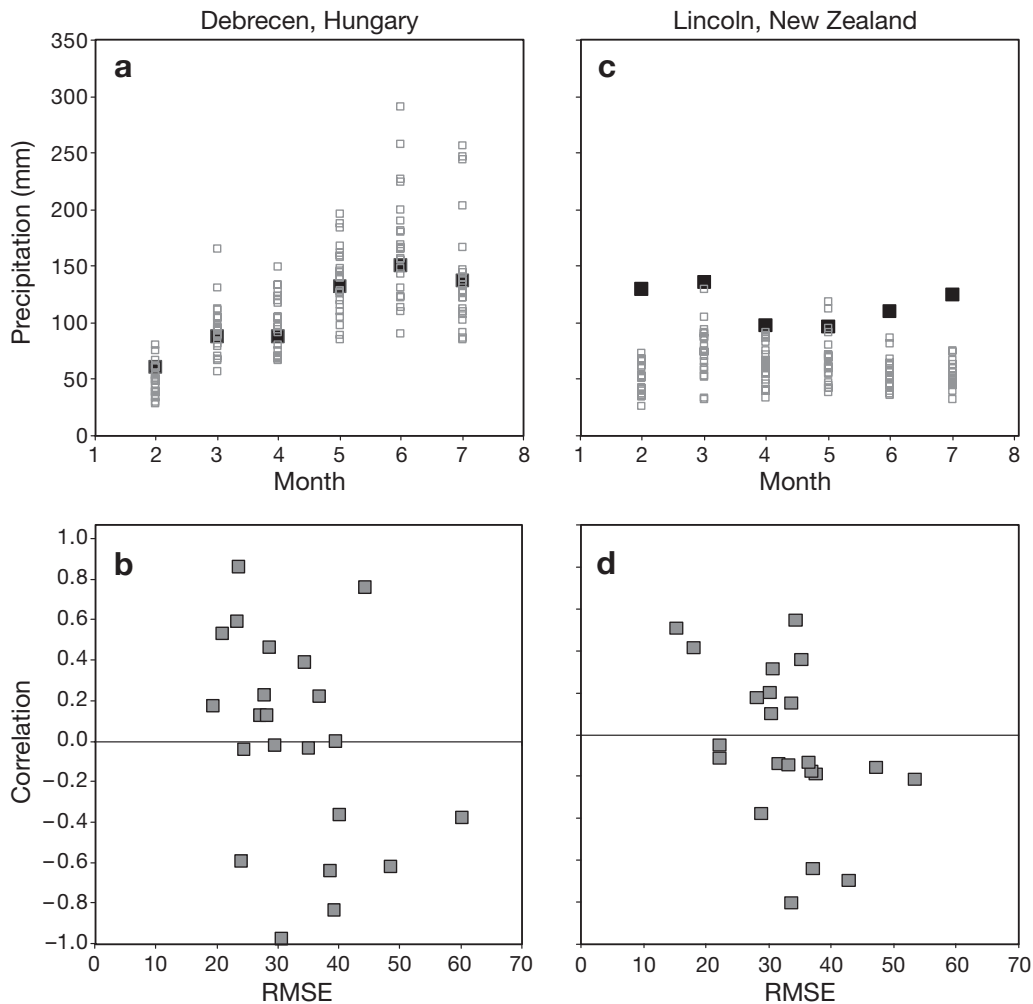
Fig. 1. (a,c) Ranges for total monthly precipitation (1980–2001) for observed (■) and forecasted (□) weather. Each open square represents the anomaly range, calculated for 9-member ensembles for each of years 1980–2001. (b,d) Correlation coefficient and root mean squared error (RMSE) between observed and forecasted anomalies for precipitation. Statistics were calculated for 6 mo observed and forecasted values for years between 1980 and 2001

anomaly ranges for the forecast, which started on the 1st of February for each year and plotted them in Figs. 1 & 2. For precipitation in Lincoln, observed ranges are larger than anomaly ranges for the seasonal forecast. As a result, we can expect a reduction in uncertainty when generating precipitation using the seasonal forecast data (Fig. 1c). In Debrecen, the situation is different, i.e. anomaly ranges derived from the seasonal forecast dataset are in some cases larger then those from the observed dataset (Fig. 1a). RMSE at both sites ranges between 15 and 60 mm and the mean of correlation coefficients is close to 0 (Fig. 1b,d). We also computed RMSE and correlation coefficients between an individual ensemble member and observed data (not presented here). In this case the correlation coefficient did not improve and RMSE was even higher. The relatively high values for RMSE could be explained by the lack of correlation between forecasted and observed data.

The mean of the observed range of maximum monthly temperatures for 1980–2001 in Debrecen was 9.0°C with the maximum range of 13.7°C in February. Ranges for model anomalies of Tmax are substantially lower than ranges calculated for observed temperature only for the 2 first mo, February and March. For the remaining months, model anomaly ranges are similar to observed ranges (Fig. 2a). For Lincoln, Tmax ranges for model anomalies are consistently lower than ranges for observed anomalies. Correlation coefficients and RMSE for Tmax anomalies are presented in Fig. 2b,d for Debrecen and Lincoln, respectively. In Lincoln the mean RMSE is almost half of the mean RMSE in Debrecen, i.e. 1.3 and 2.2, respectively, and the mean correlation coefficient at both sites is 0.2. Based on these statistics, the ECMWF GCM demonstrated better skill in predicting Tmax in Lincoln than in Debrecen.
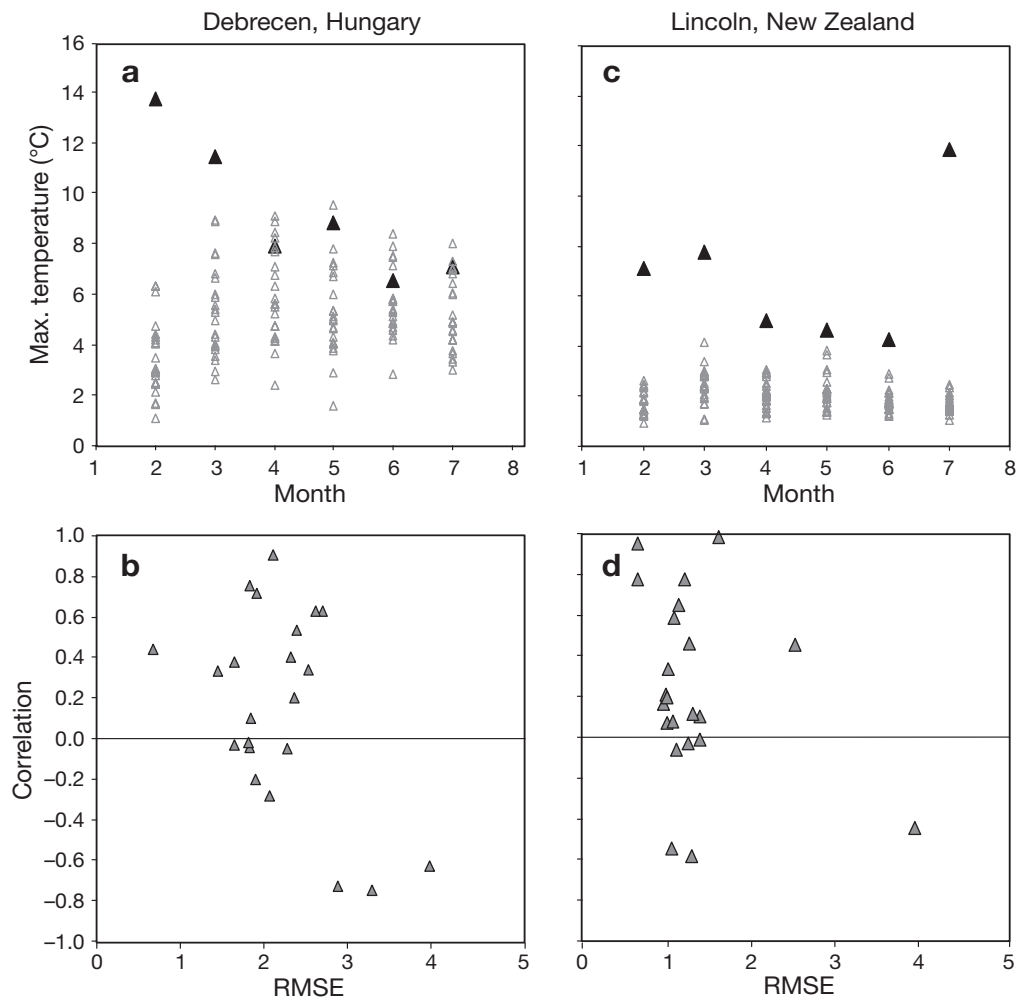
Fig. 2. The same as Fig. 1 for Tmax. For (a,c): observed (▲); forecasted (△)

In conclusion, the ECMWF GCM skill in modelling precipitation is limited at both sites. Although in Lincoln the ranges of model anomalies for precipitation are lower than for observed anomalies, the prediction accuracy is relatively poor (average RMSE ~33 mm). The skill in predicting Tmax is better than in predicting precipitation.

3. LINKING SEASONAL FORECAST WITH A WEATHER GENERATOR

The ECMWF seasonal forecast was provided as an ensemble of monthly anomalies of meteorological variables at a coarse spatial resolution. To be useful for a variety of applications, including agricultural and hydrological models, monthly anomalies need to be downscaled to site-specific daily weather. One approach for achieving this makes use of a stochastic weather generator. Weather generator parameters for a site can be altered using anomalies from the seasonal forecast and

a new set of parameters can be used to generate weather time series with statistical properties similar to those of the seasonal forecast (Wilks 2002, Meza & Wilks 2004, Feddersen & Andersen 2005).

In the present study, we use the LARS-WG stochastic weather generator (Semenov et al. 1998, Semenov & Brooks 1999). LARS-WG produces synthetic daily time series of maximum and minimum temperature, precipitation and solar radiation. Precipitation occurrence is modeled using the series approach (Racsko et al. 1991) rather than the Markov chain model used in the WGEN weather generator (Richardson 1981). Appendix 1 gives a more technical description of LARS-WG. Given several years of observed weather at a site s, LARS-WG computes a set of site-specific parameters WG_s. These parameters are used by LARS-WG to generate synthetic weather time series of a required length with statistical properties similar to the observed weather at a site. Information about future weather, e.g. a seasonal forecast, can be used to alter WG_s parameters to allow generation of weather

series with new statistical properties similar to that of the future climate (Wilks 1992, Barrow et al. 1996, Semenov & Barrow 1997, Wilks 2002). For a seasonal forecast, we assume that every member $SF_{y,k}$ from the ensemble $E_y = \{SF_{y,k}, k = 1,...,N\}$ is equally probable (E_y is the seasonal forecast ensemble and y is a year). We use each individual member of the seasonal forecast ensemble $SF_{y,k}$ to adjust a set of parameters WG_s, producing a new set $WG_{s,y,k}$ specific to the site s and the member $SF_{y,k}$:

$$WG_{s,y,k} = \text{adjust}(WG_s, SF_{y,k}) \tag{3}$$

The adjust (WG_s, $SF_{y,k}$) procedure is described in Appendix 1. For site s and year y we calculate a megaset $WG_{s,y}^{SF} = \{WG_{s,y,k}, k = 1,...,N\}$; each $WG_{s,y,k}$ corresponds to the ensemble member $SF_{y,k}$. To generate daily weather series of a length of Y years for the site s and the year y, we select a random sequence of indices $\{k_i \in \{1,...,N\}, i = 1,...,Y\}$, where $k_i = \text{rand}_i(N)$ are independent and uniformly distributed values from $\{1,...,N\}$. Then for each year i ($i = 1,...,Y$), we pick a set of parameters WG_{s,y,k_i} from the megaset $WG_{s,y}^{SF}$ and use WG_{s,y,k_i} to generate one year of daily weather. The procedure is repeated Y times until we generate Y years of daily weather specific to the site s and the seasonal forecast ensemble E_y.

4. YIELD PREDICTIONS

A brief description of the Sirius wheat simulation model, including Sirius parameterization for soils and cultivars used in the present study, is given in Appendix 2. Initially we ran Sirius with the fully observed weather and computed the yield, which we called 'the model yield'. It is important to note that in our prediction with incomplete weather, we compare predicted yields with the model yield, not with the measured yield. We present some results on how well Sirius can simulate measured yields in Appendix 2.

Where observed weather was not available for the whole duration of the growing season, we generated scenarios which consisted of a mixture of observed weather from the beginning of growing season until a specific date D and synthetic weather from that date D until the end of the growing season generated by LARS-WG. For our analysis, 2 dates—the early D_1 and late D_2 forecast dates—were selected for each site, coinciding with the beginning of 6 mo ECMWF forecasts (i.e. 1st February and 1st May at Debrecen; and 1st August and 1st November at Lincoln, to account for the austral seasons). For each site s and each date D we created 2 scenarios, WG_s and $WG_{s,y}^{SF}$: (1) in the WG_s scenario, observed weather was used up to the D date, and synthetic weather thereafter generated by LARS-

WG using WG_s parameters for the site, and (2) in the $WG_{s,y}^{SF}$ scenario, observed weather was used up to the D date, and synthetic weather generated by LARS-WG, using parameters adjusted by the seasonal forecast $WG_{s,y}^{SF}$ (as described in Section 3). Each generated scenario consisted of 100 yr of daily weather, which was used as an input to Sirius to calculate the distribution of yield. Results of simulations are presented in Fig. 3a for the D_1 date and 3b for the D_2 date for Debrecen and in Fig. 4a for D_1 and 4b for D_2 for Lincoln. Following Lawless & Semenov (2005), we used the mean yield, calculated for 100 yr of weather, as a predictor of model yield, i.e. the yield calculated by Sirius using fully observed weather. In Figs. 3 & 4 we included 99% confidence intervals (CIs) for the mean values. It has been shown for Sirius that even small random perturbations of the input and/or model parameters, e.g. 0.5% variation of the parameter value, can lead to yield variations up to 0.25 t ha^{-1} (Brooks et al. 2001). Therefore, if the predicted yield falls within ± 0.25 t ha^{-1} of the model yield, then we consider this prediction successful.

CIs for the predicted mean yields could be considered as an indicator of prediction accuracy, and we expect them to be smaller for the D_2 date. At Lincoln, CIs for the D_2 date were about 50% smaller than for the D_1 date (Fig. 4); in comparison, at Debrecen, CIs for the D_2 date were only 15 to 20% smaller (Fig. 3). As expected, correlation coefficients between yields simulated with observed weather and mean yields simulated with WG_S and $WG_{s,y}^{SF}$ scenarios were higher for the D_2 date than for the D_1 date (Table 1). RMSE decreased for the D_2 date at Debrecen for both the WG_s (12%) and $WG_{s,y}^{SF}$ (21%) scenarios and at Lincoln only for the WG_s (12%) scenario. Correlation coefficients between yields were noticeably higher at Lincoln and tended to increase for the D_2 date. Comparing correla-

Table 1. Correlations with significance levels and RMSE between yields, simulated with observed weather, and mean yields, simulated with WG_s and $WG_{s,y}^{SF}$ scenarios, for 1980–2001 at Debrecen and Lincoln for early and late forecast periods

Forecast period	Ensemble type	Correlation with significance level	RMSE
Debrecen			
Feb–Jul	WG	−0.35 (0.100)	1.28
	SF	−0.05 (0.384)	1.40
May–Oct	WG	0.48 (0.025)	1.13
	SF	0.45 (0.037)	1.10
Lincoln			
Aug–Jan	WG	0.37 (0.084)	1.12
	SF	0.46 (0.032)	1.17
Nov–Apr	WG	0.62 (0.002)	0.98
	SF	0.46 (0.032)	1.19

Fig. 3. Wheat yield simulated by Sirius using observed weather 1980–2001 (□), mean yields with 99% confidence intervals simulated for 100 yr of the $WG^{SF}_{s,y}$ scenario (♦), and 100 yr of the WG_s scenario (✕) at Debr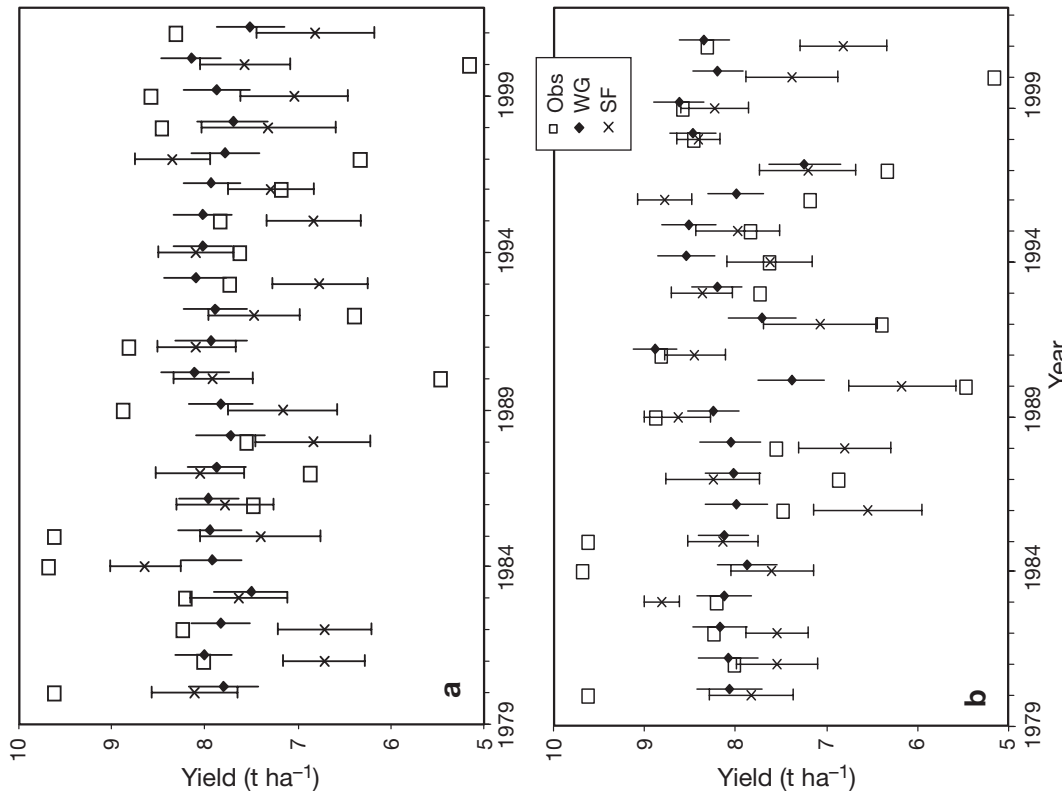ecen. For WG_s and $WG^{SF}_{s,y}$ scenarios, observed weather was used (a) from 1 October to 31 January, and (b) from 1 October to 30 April

Fig. 4. The same as Fig. 3 for Lincoln. In WG_s and $WG^{SF}_{s,y}$ scenarios, observed weather was used (a) from 1 April to 31 July, and (b) from 1 April to 31 October

tions and RMSE from Table 1, we conclude that the use of $WG_{s,y}^{SF}$ scenarios did not improve yield predictions compared with predictions based on historical climatology, i.e. WG_s scenarios, at the 2 sites selected for the study.

Detailed analysis at Debrecen showed that there are several years which produced model yield outliers (high model yields for the years 1980, 1984, 1985 and low model yields for 1990 and 2000). Yield predictions for these years were poor for both WG_s and $WG_{s,y}^{SF}$ scenarios, and for both starting dates, D_1 and D_2 (Fig. 3a,b). These years are likely to be meteorologically unusual, making it difficult for the ECMWF GCM model to make an accurate prediction.

Distributions of simulated yields for these years (see above paragraph) for the D_2 date are presented in Fig. 5. In most cases model yields are situated near the end of whiskers for both WG_s or $WG_{s,y}^{SF}$ yield distributions, or even outside box-and-whisker plots. For all 5 yr the $WG_{s,y}^{SF}$ distributions have larger variances (on average 40% more) than variances for the corresponding WG_s distributions. For example, for 1990 the standard deviation of $WG_{s,y}^{SF}$ yield distribution is 1.69 t ha^{-1} compared with 0.93 t ha^{-1} for the WG_s yield distribution.

For the year 2000 the model yield was especially low at 5.15 t ha^{-1}. Wheat growth was severely limited by water availability during May–August. Fig. 6a presents cumulative monthly precipitation anomalies for 2000 for observed weather (solid line) and for 9 members of the ensemble as predicted by the ECMFW GCM with the start date 1 May 2000. In 2000 during May–August
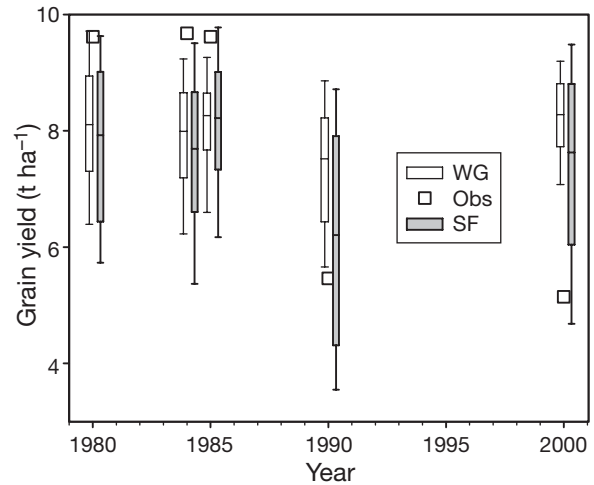

Fig. 5. Model yields simulated by Sirius using observed weather (□), box–plots of yield distributions simulated for 100 yr of the WG_s scenario (▯) and for 100 yr of the $WG_{s,y}^{SF}$ scenario (▬) for Debrecen for selected years identified as outliers. For WG_s and $WG_{s,y}^{SF}$ scenarios, observed weather was used from 1 October to 30 April. Probability—bottom whiskers: 0.1, top whiskers: 0.9; bottom of box: 0.25, top of box: 0.75

the crop received substantially less precipitation than was expected on average. However, the majority of the ensemble members predicted relatively close-to-normal cumulative anomalies of precipitation for these months, and only 2 members showed consistently low amounts of precipitation. As a result, the mean $WG_{s,y}^{SF}$ yield was 7.38 t ha^{-1} with the large standard deviation 1.69 t ha^{-1}.
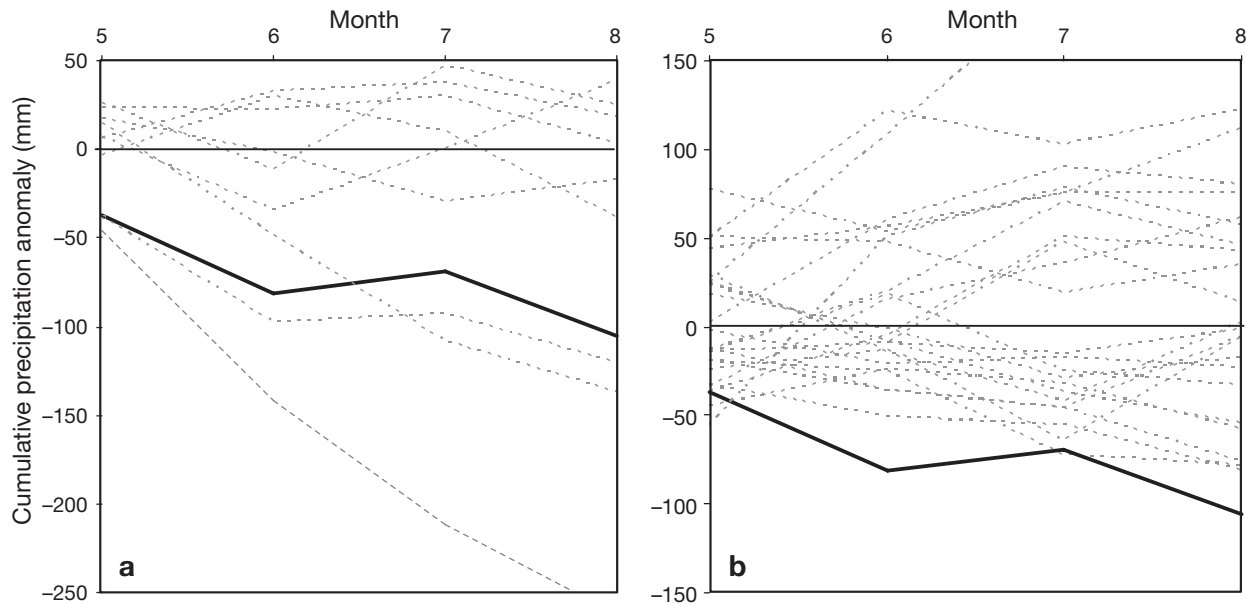

Fig. 6. Cumulative monthly precipitation anomalies for May–August at Debrecen. Thick solid line is for cumulative anomalies of observed precipitation in 2000. Dashed lines are for (a) cumulative precipitation anomalies for each of 9 members of the ensemble for 2000, and (b) cumulative anomalies of observed precipitation for 1979–2001

The prediction based on the WG_s scenario was even worse for 2000. The observed amount of precipitation in May–August 2000 was very low and extremely unlikely. Fig. 6b shows observed cumulative monthly precipitation anomalies starting from 1 May, for 1979–2001. The cumulative anomaly for 2000 was the lowest of the whole set. The approach based on WG_s scenarios is, in principle, not effective in making accurate predictions for such extreme situations. The only way to improve the yield prediction for this quite unusual year using the WG_s approach would be to increase the amount of observed weather available (shifting the D_2 date to the later date in the season).

5. CONCLUDING REMARKS

Dynamic seasonal forecast at the 2 selected locations has not improved yield predictions compared with an approach based on historical climatology. This conclusion is likely to be valid at other locations where the seasonal forecast has similar skill. It is known that seasonal forecast skill, expressed as the ranked probability score (RPS) (Epstein 1969), varies significantly from area to area (Wilks & Godfrey 2002). For example, it was shown for the probabilistic seasonal forecast of the International Research Institute for Climate Prediction (IRI), that the RPS skill score is significantly higher for latitudes less than 30° compared with latitudes more than 30° for temperature and precipitation for Northern and Southern Hemispheres (Wilks & Godfrey 2002). A similar conclusion is valid for other seasonal forecast models (Saha et al. 2006). We have tested our approach only at 2 locations with latitudes 47.6° N and 43.6° S, where the skill of seasonal forecast is generally low.

We selected 2 sites, Debrecen and Lincoln, based on our previous study, which investigated the lead time for predicting yield using WG_s scenarios at various locations (Lawless & Semenov 2005). It was shown, that for these sites, advanced prediction of grain yield is difficult and the lead time for predicting yield with probability of 0.8 using the WG_s approach is about 45 d before maturity for both sites (in comparison, at a different site—Rothamsted, UK—the lead time was 75 d, Fig. 7). At Debrecen average maturity date for our simulations with WG_s and $WG_{s,y}^{SF}$ scenarios was at the beginning of August and at Lincoln average maturity date was at the beginning of January. For the D_2 scenarios, we attempted to predict yield approximately 90 d before maturity at Debrecen and 60 d before maturity at Lincoln. The probability of successful yield prediction for these dates—based on WG_s scenarios— is low: 0.3 for Debrecen and 0.45 for Lincoln (see
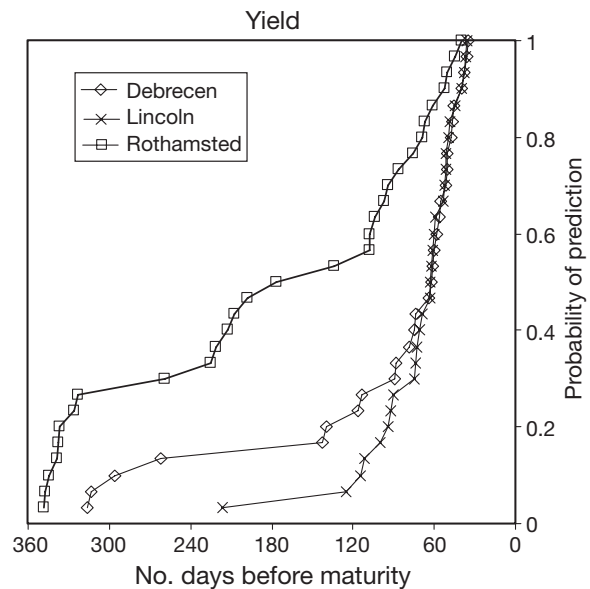

Fig. 7. Cumulative distribution functions of lead time for Sirius predictions of grain yield at 3 sites: Rothamsted, UK, Debrecen, Hungary and Lincoln, NZ. For each observed year the lead time for yield prediction is computed as the maximum number of days before maturity of unobserved weather which has no affect on the accuracy of predictions (the 95% confidence intervals of the mean yield estimate stays within the tolerance range of the yield simulated with full observed weather). For details, see Lawless & Semenov (2005)

Fig. A1 in Appendix 2). Unfortunately, the use of $WG_{s,y}^{SF}$ scenarios, based on the seasonal forecast, did not improve the accuracy of yield predictions (Figs. 5b, 6b and Table 1).

We used a subset of the DEMETER seasonal forecasts. An alternative would be to use the full DEMETER multi-model seasonal forecast, which accounts for modelling uncertainties that arise from fundamental choices made when building the GCM, as well as from the parameterization of processes unresolved at the grid scale (Hagedorn et al. 2005). In a study on the prediction of malaria epidemics, for example, a seasonal forecast based on multi-model ensembles was used (Thomson et al. 2006). A system was developed that helped predict malaria risk in Botswana, adding up to 4 mo lead time over malaria warnings issued with observed data.

Some other methods for linking yield predictions with dynamical seasonal forecasts—including selection of probability-weighted historic analogs, and statistical prediction by non–linear regression and stochastic disaggregation—were examined for prediction of field-scale maize yields in semi-arid Kenya as simulated by the CERES–maize crop model with observed daily weather (Hansen & Indeje 2004). Kenya is situated on the equator between latitudes 5° N and 5° S, where skills of seasonal forecasts are much better, and

all methods showed potential for translating seasonal climate forecasts into predictions of crop responses. Positive examples utilising seasonal forecasts for regional yield predictions for wheat in Australia and groundnut in India have also been reported (Hansen et al. 2004, Challinor et al. 2005)

Another example of an application of the seasonal forecast was reported in Cantelaube & Terres (2005), where authors used the DEMETER seasonal forecast to improve predictions of the Joint Research Centre crop growth modelling system at the level of administrative regions in Europe. One of their conclusions was that 'reliable crop yield predictions can be obtained using an ensemble multi–model approach'. However, this conclusion was based on the analysis of only 4 growing seasons, 1995–1998. Our analysis suggests that they might have just been lucky selecting these years. In our study, $WG_{s,y}^{SF}$ yield predictions outperformed WG_s predictions at Debrecen for 1995–1998 (Fig. 5b), but not for the whole period 1980–2001.

LITERATURE CITED

Barrow EM, Semenov MA (1995) Climate change scenarios with high spatial and temporal resolution for agricultural applications. Forestry 68:349–360

Barrow EM, Hulme M, Semenov MA (1996) Effect of using different methods in the construction of climate change scenarios: examples from Europe. Clim Res 7:195–211

Brooks RJ, Semenov MA, Jamieson PD (2001) Simplifying Sirius: sensitivity analysis and development of a meta-model for wheat yield prediction. Eur J Agron 14:43–60

Cantelaube P, Terres JM (2005) Seasonal weather forecasts for crop yield modelling in Europe. Tellus Ser A Dyn Meteorol Oceanogr 57:476–487

Challinor AJ, Slingo JM, Wheeler TR, Doblas-Reyes FJ (2005) Probabilistic simulations of crop yield over western India using the DEMETER seasonal hindcast ensembles. Tellus Ser A Dyn Meteorol Oceanogr 57:498–512

Doblas-Reyes FJ, Hagedorn R, Palmer TN (2006) Developments in dynamical seasonal forecasting relevat to agricultural management. Clim Res 33:19–26

Epstein ES (1969) A scoring system for probability forecasts of ranked categories. J Appl Meteorol 8:985–987

Ewert F, Rodriguez D, Jamieson P, Semenov MA and 10 others (2002) Effects of elevated CO_2 and drought on wheat: testing crop simulation models for different experimental and climatic conditions. Agric Ecosyst Environ 93:249–266

Feddersen H, Andersen U (2005) A method for statistical downscaling of seasonal ensemble predictions. Tellus Ser A Dyn Meteorol Oceanogr 57:398–408

Hagedorn R, Doblas-Reyes FJ, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting. I. Basic concept. Tellus Ser A Dyn Meteorol Oceanogr 57:219–233

Hansen JW, Indeje M (2004) Linking dynamic seasonal climate forecasts with crop simulation for maize yield prediction in semi-arid Kenya. Agric For Meteorol 125:143–157

Hansen JW, Potgieter A, Tippett MK (2004) Using a general circulation model to forecast regional wheat yields in northeast Australia. Agric For Meteorol 127:77–92

Ines AVM, Hansen JW (2006) Bias correction of daily GCM rainfall for crop simulation studies. Agric For Meteorol 138:44–53

Jamieson PD, Semenov MA (2000) Modelling nitrogen uptake and redistribution in wheat. Field Crops Res 68:21–29

Jamieson PD, Berntsen J, Ewert F, Kimball BA, Olesen JE, Pinter PJJ, Porter JR, Semenov MA (2000) Modelling CO_2 effects on wheat with varying nitrogen supplies. Agric Ecosyst Environ 82:27–37

Jamieson PD, Brooking IR, Semenov MA, Porter JR (1998a) Making sense of wheat development: a critique of methodology. Field Crops Res 55:117–127

Jamieson PD, Semenov MA, Brooking IR, Francis GS (1998b) Sirius: a mechanistic model of wheat response to environmental variation. Eur J Agron 8:161–179

Lawless C, Semenov MA (2005) Assessing lead-time for predicting wheat growth using a crop simulation model. Agric For Meteorol 135:302–313

Lawless C, Semenov MA, Jamieson PD (2005) A wheat canopy model linking leaf area and phenology. Eur J Agron 22:19–32

Meza FJ, Wilks DS (2004) Use of seasonal forecasts of sea surface temperature anomalies for potato fertilization management. Theoretical study considering EPIC model results at Valdivia, Chile. Agric Syst 82:161–180

Palmer TN (1993) Extended-range atmospheric prediction and the Lorenz model. Bull Am Meteorol Soc 74:49–65

Palmer TN (2001) A nonlinear dynamical perspective on model error: a proposal for non-local stochastic–dynamic parametrization in weather and climate prediction models. QJR Meteorol Soc 127:279–304

Palmer TN, Alessandri A, Andersen U, Cantelaube P and 21 others (2004a) Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). Bull Am Meteorol Soc 85:853–888

Palmer TN, Alessandri A, Andersen U, Cantelaube P and 21 others (2004b) Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). Bull Am Meteorol Soc 85:853–888

Palmer TN, Doblas-Reyes FJ, Hagedorn R, Weisheimer A (2005) Probabilistic prediction of climate using multi-model ensembles: from basics to applications. Philos Trans R Soc B Biol Sci 360:1991–1998

Racsko P, Szeidl L, Semenov M (1991) A serial approach to local stochastic weather models. Ecol Model 57:27–41

Richardson CW (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour Res 17:182–190

Saha S, Nadiga S, Thiaw C, Wang J and 13 others (2006) The NCEP climate forecast system. J Clim 19:3483–3517

Semenov MA, Barrow EM (1997) Use of a stochastic weather generator in the development of climate change scenarios. Clim Change 35:397–414

Semenov MA, Brooks RJ (1999) Spatial interpolation of the LARS-WG stochastic weather generator in Great Britain. Clim Res 11:137–148

Semenov MA, Brooks RJ, Barrow EM, Richardson CW (1998) Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. Clim Res 10:95–107

Thomson MC, Doblas-Reyes FJ, Mason SJ, Hagedorn R, Connor SJ, Phindela T, Morse AP, Palmer TN (2006) Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. Nature 439:576–579

Wilks DS (1992) Adapting stochastic weather generation algorithms for climate changes studies. Clim Change 22:67–84

Wilks DS (2002) Realizations of daily weather in forecast seasonal climate. J Hydrometeorol 3:195–207

Wilks DS, Godfrey CM (2002) Diagnostic verification of the IRI net assessment forecasts, 1997–2000. J Clim 15:1369–1377

Wolf J, Evans LG, Semenov MA, Eckersten H, Iglesias A (1996) Comparison of wheat simulation models under climate change. I. Model calibration and sensitivity analyses. Clim Res 7:253–270

Appendix 1. A description of the LARS-WG stochastic weather generator

LARS-WG is based on the series weather generator described in (Racsko et al. 1991) with a detailed description being given in Barrow & Semenov (1995), Semenov et al. (1998), and Semenov & Brooks (1999). LARS-WG produces synthetic daily time series of maximum and minimum temperature, precipitation and solar radiation. The weather generator distinguishes dry and wet days depending on whether or not the precipitation is >0. Precipitation is modelled using semi-empirical probability distributions for each month for the lengths of series of wet and dry days and for the amount of precipitation on a wet day. A semi-empirical distribution, $E = \{e_0, e_i; h_i, i = 1,....,10\}$, is a histogram with 10 intervals, $[e_{i-1},e_i)$, where $e_{i-1} < e_i$, and h_i denotes the number of events from the observed data in the ith interval. In the case of precipitation, $e_0 = 0$. Minimum temperature, Tmax and radiation are related to the amount of cloud cover, and so LARS-WG uses separate wet and dry day distributions for each of these variables. The normal distribution is used for temperature variables with the mean and standard deviation varying daily according to finite Fourier series of order 3. Time auto-correlations used for minimum and Tmax are constant through the year for the particular site and the cross-correlation of the normalized residuals from the daily mean is pre-set at 0.6. Semi-empirical distributions with equal interval size are used for solar radiation.

The absolute monthly anomaly for each individual members of the seasonal forecast ensemble $SF\{pr,tn,tx,rd\}$ (pr: precipitation; tn, tx: Tmin and Tmax; rd: radiation) is defined as the deviation of monthly values of the member from the average values over the ensemble and over all years for which seasonal forecast was available. The ECMWF GCM seasonal forecast was available for 22 yr (1980–2001) and an ensemble size for each year is 9. For our study we used anomalies for monthly total precipitation, minimum and Tmax and radiation.

To adjust the LARS-WG semi-empirical distributions for precipitation and radiation we stretch the distribution, changing the mean value according to the corresponding anomaly. The end points of the adjusted distribution are given by E_i and H_i respectively:

$$E_i = \rho e_i$$
$$H_i = h_i \qquad (i = 0, 1,...,10)$$

where ρ is a relative predicted change in mean value. Anomalies from the seasonal forecast ensemble $SF.pr$ and $SF.rd$ are given as absolute changes, therefore they were converted into relative anomalies by calculating monthly average precipitation and radiation from observed weather for a site. On a number of occasions, negative anomalies for precipitation have absolute values larger than the observed monthly mean; in these cases, they were limited to reduce the monthly mean value by 75%.

The shape of the daily temperature distributions for wet and dry days are fixed as the normal distribution with the values of mean and standard deviation changing daily and calculated by a Fourier series $f(t)$ of order 3, so that

$$f(t) = \frac{1}{2}a_0 + \sum_{j=1}^{3}(a_j\cos(j\omega t) + b_j\sin(j\omega t))$$

where t = Julian day, $\omega = 2\pi/365$, a_j and b_j = Fourier coefficients.

To adjust a Fourier series for temperature, we calculated new monthly mean temperature values by adding observed monthly means and corresponding monthly mean anomalies, $SK.tn$ or $SK.tx$, and recalculating Fourier coefficients for new mean values.

The auto correlation coefficients (for minimum and Tmax and radiation) were kept unchanged.

Appendix 2. A description of the Sirius wheat simulation model

The Sirius wheat simulation model was used for quantifying the utility of seasonal forecasts for predicting grain yields (Jamieson et al. 1998b, Brooks et al. 2001, Lawless et al. 2005). Sirius calculates biomass from intercepted photosynthetically active radiation (PAR) and grain growth from simple partitioning rules, on a daily basis. Leaf area index (LAI) is developed from a simple canopy model (Lawless et al. 2005). Phenological development is calculated from the mainstem leaf appearance rate and final leaf number, with the latter determined by responses to daylength and vernalisation (Jamieson et al. 1998a). Effects of water and N deficits are calculated through their influences on LAI development and radiation-use efficiency (Jamieson & Semenov 2000). Despite there being no calculation of tiller dynamics, the model accurately simulates the behaviour of crops exposed to a wide range of conditions. The model was calibrated and validated for several modern wheat cultivars and tested in many environments and climates, including Europe, NZ, USA and Australia, and under conditions of climate change (Fig. A1) (Wolf et al. 1996, Jamieson & Semenov 2000, Jamieson et al. 2000, Ewert et al. 2002).

Sirius requires certain data as input. It needs daily weather data (minimum and Tmax, radiation, and rainfall). It also requires a set of cultivar parameters, which includes phyllochron, maximum canopy area, vernalisation rate parameters, daylength sensitivity and grainfill kinetic parameters. Sirius requires a description of the soil, including moisture retention properties of the soil, since they directly affect both water and nitrogen availability. And finally Sirius needs a management file, which includes sowing date, N applications and irrigations, and initial inorganic N.

In the present study, we used cultivar parameters representing cultivar Mercia for Debrecen and cultivar Batten for Lincoln, which were calibrated against agronomic experimental data from the UK and New Zealand. The management description consisted of a sowing date of 10th November for Debrecen and 10 May for Lincoln with an initial amount of inorganic N in the soil of 100 kg ha^{-1}, a single mineral N application of 200 kg ha^{-1} 3 mo after sowing. The same soil description was used for both sites, corresponding to a Lincoln shallow soil. The soil has an available water capacity (AWC) of 135 mm with a percolation constant of 0.3 d^{-1} and saturated moisture content and drained upper and lower limits of 44, 30 and 12% respectively over the whole profile. The initial inorganic N was described as being split over a 0.75 m profile, 80% in the top 0.25m, 10% in the mid 0.25 m and the remainder in the bottom 0.25m. Organic N content was 10 t ha^{-1} in Debrecen and 5 t ha^{-1} in Lincoln with mineralization constants of 0.07 and 0.14 (kg mineral N) (t organic N)$^{-1}$ ha^{-1} d^{-1}, for Debrecen and Lincoln, respectively.
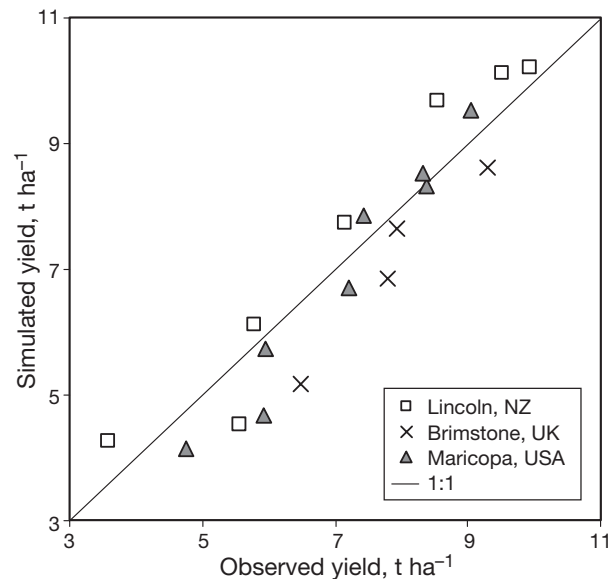

Fig. A1. Comparison between Sirius simulated and observed wheat yields in diverse environments: rainshelter experiment, Lincoln, NZ; nitrogen experiment, Brimstone, UK; and nitrogen and water experiment, Maricopa, USA