Journal of Agricultural Science (2016), **154**, 1218–1240. © Cambridge University Press 2015
doi:10.1017/S0021859615001124

CROPS AND SOILS RESEARCH PAPER

Comparing the performance of 11 crop simulation models in predicting yield response to nitrogen fertilization

T. J. SALO[1]*, T. PALOSUO[2], K. C. KERSEBAUM[3], C. NENDEL[3], C. ANGULO[4], F. EWERT[4], M. BINDI[5], P. CALANCA[6], T. KLEIN[6], M. MORIONDO[7], R. FERRISE[5], J. E. OLESEN[8], R. H. PATIL[8,9], F. RUGET[10], J. TAKÁČ[11], P. HLAVINKA[12,13], M. TRNKA[12,13] AND R. P. RÖTTER[2]

[1] *Natural Resources Institute Finland, 31600 Jokioinen, Finland*
[2] *Natural Resources Institute Finland, 50100 Mikkeli, Finland*
[3] *Leibniz Centre for Agricultural Landscape Research (ZALF), Institute of Landscape Systems Analysis, Eberswalder Straße 84, 15374 Müncheberg, Germany*
[4] *University of Bonn, Institute of Crop Science and Resource Conservation, Katzenburgweg 5, 53115 Bonn, Germany*
[5] *DISPAA, Department of Agri-food Production and Environmental Sciences, University of Florence, Piazzale delle Cascine 18, 50144 Florence, Italy*
[6] *Agroscope Reckenholz-Tänikon ART, 8046 Zurich, Switzerland*
[7] *National Research Council of Italy, IBIMET-CNR, Institute of Biometeorology, via Caproni 8, 50145 Florence, Italy*
[8] *Department of Agroecology, Aarhus University, Blichers Allé 20, DK-8830 Tjele, Denmark*
[9] *Department of Agronomy, University of Agricultural Sciences, 580005 Dharwad, India*
[10] *INRA, UMR 1114 EMMAH Environement et Agronomie, F-84000, Avignon, France*
[11] *National Agricultural and Food Centre – Soil Science and Conservation Research Institute, Gagarinova 10, 827 13 Bratislava, Slovak Republic*
[12] *Global Change Research Centre AS CR, v.v.i., Bělidla 986/4a, 603 00 Brno, Czech Republic*
[13] *Institute of Agrosystems and Bioclimatology, Mendel University in Brno, Zemedelska 1, Brno 613 00, Czech Republic*

(*Received 10 December 2014; revised 21 September 2015; accepted 27 October 2015; first published online 22 December 2015*)

SUMMARY

Eleven widely used crop simulation models (APSIM, CERES, CROPSYST, COUP, DAISY, EPIC, FASSET, HERMES, MONICA, STICS and WOFOST) were tested using spring barley (*Hordeum vulgare* L.) data set under varying nitrogen (N) fertilizer rates from three experimental years in the boreal climate of Jokioinen, Finland. This is the largest standardized crop model inter-comparison under different levels of N supply to date. The models were calibrated using data from 2002 and 2008, of which 2008 included six N rates ranging from 0 to 150 kg N/ha. Calibration data consisted of weather, soil, phenology, leaf area index (LAI) and yield observations. The models were then tested against new data for 2009 and their performance was assessed and compared with both the two calibration years and the test year. For the calibration period, root mean square error between measurements and simulated grain dry matter yields ranged from 170 to 870 kg/ha. During the test year 2009, most models failed to accurately reproduce the observed low yield without N fertilizer as well as the steep yield response to N applications. The multi-model predictions were closer to observations than most single-model predictions, but multi-model mean could not correct systematic errors in model simulations. Variation in soil N mineralization and LAI development due to differences in weather not captured by the models most likely was the main reason for their unsatisfactory performance. This suggests the need for model improvement in soil N mineralization as a function of soil temperature and moisture. Furthermore, specific weather event impacts such as low temperatures after emergence in 2009, tending to enhance tillering, and a high precipitation event just before harvest in 2008, causing possible yield penalties, were not captured by any of the models compared in the current study.

INTRODUCTION

Nitrogen (N) is the main nutrient affecting crop growth and yield formation, but at the same time excessive N

* To whom all correspondence should be addressed.
Email: tapio.salo@luke.fi

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

in relation to crop N requirements has substantial negative environmental effects (Sutton *et al*. 2011). Effective N fertilizer management is therefore crucial for optimizing crop yields, maximizing farmer income and minimizing negative environmental impacts from reactive N flows. Climate change will directly and indirectly change crop N demand, N mineralization and leaching (Stuart *et al*. 2011; Patil *et al*. 2012; Doltra *et al*. 2014). This will make management of the agricultural N cycle even more challenging in future than it is at present.

Crop models have been widely applied for optimizing N management in agriculture (Shaffer 2002; Kersebaum *et al*. 2005; Rahn *et al*. 2010; Nendel *et al*. 2013). They have been important means in estimating the effect of N fertilizer application rates and timing on both crop yield (Dobermann *et al*. 2000; Cassman *et al*. 2003; Nykänen *et al*. 2009) and N leaching (Wolf *et al*. 2005; Cannavo *et al*. 2008; Hyytiäinen *et al*. 2011).

For Europe, four crop model inter-comparison studies were set up during 2007–2010 based on the need for evaluating performance of crop models against data from field experiments under current climate as a basis and first step for judging their utility for climate change impact assessments (Monteith 1981; Rötter *et al*. 2011*a*; Wheeler & von Braun 2013). These inter-comparisons were performed within the framework of COST action 734 seeking coverage of the most widely used and accessible crop simulation models: two of the comparisons were for winter wheat (Palosuo *et al*. 2011) and spring barley (Rötter *et al*. 2012) across multiple sites in Europe with limited calibration and one for heat and drought sensitivity of wheat and maize at two Austrian locations (Eitzinger *et al*. 2013). The fourth inter-comparison presented here was conducted using comprehensive barley data sets from one Finnish location and included a comparison of the capability of 11 different models to simulate yield response to varying N inputs.

The main objective of the current model inter-comparison study was to assess and compare the uncertainties and capability of different models in simulating spring barley yield responses to a range of N application rates from none to high (150 kg N/ha) when provided with a comprehensive data set for model calibration. Eleven crop models were run at the Jokioinen site in Finland for two growing seasons of spring barley for model calibration and one season for model testing.

MATERIAL AND METHODS

Models

The 11 crop simulation models included in the inter-comparison were APSIM, CERES, CROPSYST, COUP, DAISY, EPIC, FASSET, HERMES, MONICA, STICS and WOFOST (Table 1). Models were first calibrated and then tested against different N fertilizer rates. Details about the models used can be obtained from the main references gathered by Rötter *et al*. (2012), except for APSIM (Probert *et al*. 1998; Keating *et al*. 2003), COUP (Jansson & Karlberg 2012) and EPIC (Williams 1995). Table 2 provides a characterization of process descriptions on N dynamics for the various models. CROPSYST was operated by two separate teams using different model versions and calibration approaches. The results for these are presented as CROPSYST a and b.

All models were dynamic and worked on a daily time step. Most of them used similar process descriptions of the N cycle. Below, some differences among models in simulating major processes are described. For more details of N processes, the reader is referred to Table 2 and for other biophysical processes, to Palosuo *et al*. (2011) and Rötter *et al*. (2012):

- *N demand*. Most of the models determined crop N demand according to the daily growth and the critical N concentration needed for accumulated biomass (Table 2). Some models applied an optimum N concentration curve (DAISY, HERMES, STICS) or factor (MONICA).
- *N uptake*. Most models took up soil N according to crop N demand, soil available N, soil moisture, root length distribution and N transport rates in soil (Table 2). CROPSYST did not provide grain N content. It evaluated N uptake in total aboveground biomass, but did not estimate the proportion of N in grains. WOFOST used the QUantitative Evaluation of the Fertility of Tropical Soils (QUEFTS) approach to estimate N supply from soil and N uptake by the crop during the growing season (Janssen *et al*. 1990).
- *N stress*. Most of the models estimated N stress from the difference or ratio between daily or cumulative N demand and crop N uptake (Table 2). Lack of N determined by N concentrations dropping below the critical concentration – lead in most models to depressed daily dry matter (DM) growth, reduced leaf area expansion (FASSET, STICS), decreased gross photosynthesis (DAISY), and to enhanced root biomass allocation in some of the models (e.g. COUP, FASSET). STICS also reduced grain

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

Table 1. *Model version applied in the current study, references to papers with model descriptions and model web address*

Model	Version	Applications for cereals	Documentation/accessibility (weblink)
APSIM		Keating *et al.* (2003); Anwar *et al.* (2009)	http://www.apsim.info
CERES	V.4.01.0	Hlavinka *et al.* (2010); Eitzinger *et al.* (2004); Trnka *et al.* (2004)	http://dssat.net/
CROPSYSTa	V. 3.04.08	Donatelli *et al.* (1997); Stöckle *et al.* (2003)	http://modeling.bsyse.wsu.edu/CS_Suite_4/CropSyst/index.html
CROPSYSTb	V.4.13.04	Stöckle *et al.* (2003)	
COUP	V. 3.1	Jansson & Thoms-Hjärpe (1986); Borg *et al.* (1990); Nykänen *et al.* (2009)	http://www2.lwr.kth.se/CoupModel/index.html
DAISY	V. 4.01	Hansen *et al.* (1990); Abrahamsen & Hansen (2000); Hansen (2000)	http://code.google.com/p/daisy-model/
EPIC		Strauss *et al.* (2012)	http://epicapex.tamu.edu/
FASSET	V. 2.0	Berntsen *et al.* (2004); Doltra *et al.* (2011); Sapkota *et al.* (2012)	http://www.fasset.dk
HERMES	V. 4.26	Franko *et al.* (2007); Kersebaum (2007); Hlavinka *et al.* (2014)	http://www.zalf.de/en/forschung/institute/lsa/forschung/oekomod/hermes/Pages/default.aspx
MONICA	V. 1.0	Nendel *et al.* (2011)	http://monica.agrosystem-models.com
STICS	V. 6.9	Corre-Hellou *et al.* (2009); Launay *et al.* (2009)	http://www6.paca.inra.fr/stics_eng
WOFOST	V. 7.1	Rötter *et al.* (2011*b*)	http://www.wofost.wur.nl

number due to N stress at certain developmental stages.

- *N mineralization from soil organic matter*. Most of the models divided the soil organic matter pool into 2–7 interconnected (conceptual) sub-pools of different mineralization rates (Table 2). Some models calculated N mineralization from C turnover (DAISY, FASSET, MONICA), while others calculated N mineralization directly, without any link to C dynamics (HERMES, STICS). Mineralization of soil organic matter was controlled at least by soil temperature and moisture. In addition, in some models soil clay content (MONICA, STICS, DAISY), soil oxygen concentration (MONICA) and soil calcium contents (STICS) were used as control variables for N mineralization.
- *Vertical movement of water in soil*. Eight out of 11 models used a simple capacity or tipping bucket approach for soil water balance (Table 2). COUP and DAISY applied the more detailed Richard's approach for soil water movement (Richards 1931). CROPSYST users could select either of the approaches for soil water movement, and capacity approach was used by CROPSYSTa and Richard's approach by CROPSYSTb.
- *Vertical movement of N in soil*. N movement was based mostly on the convection–dispersion equation or complete mixing of the flowing solute with the resident solute (Table 2). FASSET used a solute leaching intermediate model (Addiscott & Whitmore 1991), and STICS used a reservoir with variable mixing depths (Burns 1974). WOFOST did not calculate N movement or leaching. HERMES and MONICA considered upward movement by capillary rise if a shallow groundwater table was present.
- *Denitrification*. N loss due to denitrification was modelled as a function of soil temperature and water content in all models except WOFOST that did not simulate processes of soil N dynamics. Ammonia volatilization was included in CROPSYST, DAISY, MONICA and STICS only.

Study site

The model inter-comparison study was carried out using data from field experiments conducted in the years 2002, 2008 and 2009 at Jokioinen, southern Finland (60°42′N, 23°30′E 104 m a.s.l). In Jokioinen during 1980–2010, temperature sum (>5 °C base temperature) in May–September averaged 1244 °Cd and the precipitation sum 317 mm, which created typical conditions for spring barley production in a boreal climate. The experiments in 2002 and 2008 were

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

Table 2. *Modelling approaches applied in the current study regarding the major processes determining nitrogen (N) response on crop growth and nitrogen dynamics in soil*

	APSIM	CERES	CROPSYST	COUP	DAISY	EPIC	FASSET	HERMES	MONICA	STICS	WOFOST
Water movement											
Capacity model	x	x	x			x	x	x	x	x	x
Richards			x	x	x						
Soil organic matter, number of pools	3	2	5	2	6	5	7	2	6	3	1
Factors affecting N mineralization											
Soil T	x	x	x	x	x	x	x	x	x	x	
Soil moisture	x	x	x	x	x	x	x	x	x	x	
C/N-ratio	x	x	x	x	x	x	x		x	x	
Inorganic N immobilization	x	x	x	x	x	x	x		x	x	
Denitrifcation	x	x	x	x	x	x	x	x	x	x	
Ammonia volatilization			x		x				x	x	
Nitrogen movement											
Convection-dispersion					x			x	x		
Mixing of solute	x	x	x	x		x	x			x	
Crop N demand											
Daily growth	x	x	x	x	x	x	x	x	x	x	
Critical N concentration	x	x	x	x	x	x	x			x	
Optimum N concentration					x			x	x		
Target yield based											x
Nitrogen stress function											
Daily actual/potential demand	x	x		x	x	x	x			x	
Critical N curve			x					x	x		
Growth response to N stress											
Daily growth decreased		x	x	x	x		x	x	x	x	
Increased root allocation		x		x			x				
Phenological response to N stress	x								x	x	
Development enhanced after anthesis	x								x		

x denotes that approach was used in the model.

conducted on the same field, while the experimental field in 2009 was located 5 km northeast of the former site. Both sites were *c*. 200 m from the river Loimijoki and their soil properties (Table 3) and crop rotation were considered similar enough to allow simulations with the same soil chemical and physical characteristics (Table 4). The soil type was sandy clay in the top layer (0–30 cm) and heavy clay below 30 cm. The soil was classified as Vertic Cambisol (Yli-Halla & Mokma 2001). Temperature (>5 °C), global radiation and precipitation sums from May to September in each growing season are presented in Fig. 1. In the growing season 2002, mean temperature (14·6 °C) was higher than the long-term average (13·1 °C, averaged over 1980–2010), while in 2008 and 2009 mean temperatures were 12·6 and 13·6 °C, respectively. Precipitation during the modelled growing seasons in May–September were lower than long-term average (317 mm, averaged over 1980–2010) amounting to 219, 288 and 246 mm, respectively, in 2002, 2008 and 2009. In 2008, there was a rain event of 47 mm on 7–8 August. For the cultivation of spring cereals in boreal climates it is known that climatic and soil conditions 2 months after sowing are crucial for crop yield establishment and small differences during early growth may have large implications for final yield (Håkansson *et al*. 2002; Pietola & Tanni 2003). After snow melting, water flows to the sub-soil drainage system and soil dries to field capacity, and soil layers below 5 cm only start to dry after soil preparation and sowing. Groundwater level decreases from the depth of 30–40 cm after soil frost melting to below 2 m depth in the middle of the growing season.

The data from 2002 originated from a spatial soil mapping experiment (Farkas *et al*. 2006), where soil physical conditions were measured and compared with crop growth from the 20 replicates. In 2008 and 2009, data were collected from a fertilizer experiment, with mineral N fertilizer rates from 0 to 150 kg N/ha at 30 kg N/ha intervals (designated as N0, N30 and so on up to N150). The experimental design was randomized complete blocks with four replicates in both years. Plot size was 25 × 3 m^2 and yields were harvested from 50 m^2 in 2008 and 35 m^2 in 2009. Preceding crops were barley for 2008 and 2009, and spring wheat for 2002. Initial soil inorganic N content was measured from 0–30 cm soil layer and was 20–24 kg N/ha at the time of sowing. Previous measurements from a similar clay soil (Pietola *et al*. 1999; Pietola & Tanni 2003) and sampling in spring 2009 provided information that the soil inorganic N in layer 30–90 cm was only 8–10 kg N/ha: this information was also provided for the modellers. Soil water content was slightly below field capacity (<40% vol.) at the time of sowing.

In the experimental years, compound nitrogen : phosphorus : potassium (NPK) fertilizer (17% N as ammonium nitrate) was placed between every second seed row at the time of sowing. Distance between seed rows was 12·5 cm. Spring barley cultivar Annabell (two-row) was used in all experimental years and the target seeding density was 500 plants/m^2. In all experiments, the plots were kept weed-free and no interference from pests or diseases was observed. Management practices followed the best practices in the region and regulations of the Finnish Agri-environmental programme (Finnish Ministry of Agriculture and Forestry 2007).

Field observations

Each year, the leaf area index (LAI) was measured with a Sun Scan Canopy Analyser (Delta-T Devices, SS1, Cambridge, UK). In 2002 and 2008, LAI was measured four times in all treatments and in 2009 nine times in the N90 treatments and three times in all remaining treatments. Total above-ground biomass was sampled from all N treatments 41 days after sowing (DAS) in 2008 and 53 DAS in 2009. In 2009, nine sequential biomass measurements were taken from the N90 treatment. Biomass samples were collected from a 0·25 m^2 square in each plot and stubble was cut to a height of 2 cm above ground level. Grain yields were measured from 35 to 50 m^2 areas within each plot with a combine harvester and all grain yields were calculated as dry weight. To measure harvest index, total above-ground biomass samples were taken in 2009 at harvest from N90 treatments. Individual grain weight and grain N concentration were measured from all N treatments in 2008 and 2009. Grain N uptake was calculated from grain N concentration and grain yield. Soil water content was measured in 2002 and in 2009 with gypsum blocks and tensiometers from 25 to 40 cm in 2002 and from 10, 20 to 30 cm in 2009. In 2008, soil water content measurements with time-domain reflectrometry probes were available from a nearby field experiment on similar clay soil (46% clay and 25 g organic C/kg soil in 0–20 cm) from the depths of 0–30 cm and 0–60 cm. This experiment was conducted for comparing different soil management methods, and the soil water content measurements were used

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

Table 3. *Detailed soil properties, soil inorganic nitrogen (N) in spring (Nmin) and grain N uptake of the unfertilized treatments in 2008 and 2009*

Year and replicates (n)	Soil properties (0–20 cm) in the experimental years					
	2002 Mean	$n = 20$ S.D.	2008 Mean	$n = 4$ S.D.	2009 Mean	$n = 4$ S.D.
Clay (mass%)	47	5·1	46·8		52·6	
Silt (mass%)	26	3·6	24·2		16·1	
Sand (mass%)	27	7·4	29·0		31·3	
Organic carbon (mass%)	2·9	0·31	3·2	0·71	3·5	0·18
Organic nitrogen (mass%)			0·20	0·044	0·21	0·010
pH	6·4	0·17	6·2		5·9	
Calcium (mg/dm^3)	3168	295·3	3304		2788	
Potassium (mg/dm^3)	136	22·6	155·4		365·7	
Magnesium (mg/dm^3)	380	80·4	410·2		534·2	
Phosphorus (mg/dm^3)	19	3·9	23·7		9·1	
Nmin in spring 2008 (kg/ha) in 0–20 cm			20			
Nmin in spring 2009 (kg/ha) in 0–20 cm			17	3·1	21	2·4
N uptake of grains (kg/ha) in 2008			51	4·0		
N uptake of grains (kg/ha) in 2009			38	3·5	24	1·9

Table 4. *Soil characteristics of the study site used in the modelling exercise*

Layer	Sand (%)	Silt (%)	Clay (%)	C_{org} (%)	Total N (%)	fc – wp root zone (mm)*	Saturated HC (cm/h)†	pH	Soil Bulk density (kg/dm^3)
0–30	27	26	47	2·9	0·3	54	95	6·4	1·3
30–90	7	30	63	0·6	0·06	90	85	6·0	1·4

HC, hydraulic conductivity; N, nitrogen.

* mm water at field capacity (fc) and wilting point (wp) in specific root zone.

† Hydraulic conductivity including macropores.

from conventional mouldboard ploughing as that was the soil management of the field used for model calibration in 2008. Barley was estimated to be suffering from water stress when plant available water capacity (field capacity – wilting point) was decreased to 30%, which would correspond to 27 and 33% of volumetric water contents in 0–30 and 0–90 cm, respectively.

Setup of model inter-comparison

Information available for modellers for calibration

First, modellers were asked to calibrate their models for the experimental years 2002 and 2008. The data provided for this step (Table 5) included daily weather, crop management data (e.g. information on previous crop, tillage, sowing date and density, fertilization and harvest) and information on some basic soil properties (Table 4). Modellers were also provided with information on key phenological dates for spring barley during the growing periods. These included dates of sowing, emergence, flowering, yellow ripeness and harvest (Table 5). In addition, development of the LAI and the measured grain yield for the years 2002 and 2008 were provided for calibration. Calibration results of models for phenological dates, LAI and biomass development, grain yield and yield N uptake were gathered from the individual modellers for further analysis.

Model testing

For the test year 2009, similar input data as for calibration years were delivered to the modellers, except for LAI and yield data. It was agreed that the crop phenology and soil parameters derived from the calibration were used in the testing exercise. After simulating crop growth for the test year, modellers provided their

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

Table 5. *Input data provided to the models*

Category	Variable	Type
Meteorological data	Minimum temperature	Daily minimum (°C)
	Maximum temperature	Daily maximum (°C)
	Relative air humidity	Daily average (%)
	Vapour pressure at 6 am	kPa
	Global radiation	Daily sum (MJ/m^2)
	Wind speed	Daily average (m/s)
	Precipitation	Daily sum (mm)
Soil data (0 cm to maximum rooting depth)	Texture	Per layer clay, silt, sand (mass%)
	Corg	Per layer (mass%)
	C : N ratio	Per layer (no units)
	Bulk density	Per layer (m^3/m^3)
	pH	Per layer (no units)
	Field capacity	Per layer (m^3/m^3)
	Wilting point	Per layer (m^3/m^3)
	Total pore space	Per layer (m^3/m^3)
	Saturated hydraulic conductivity	Per layer (cm/h)
	Maximum rooting depth	(cm)
Crop data	Crop density	Plants/m^2
	Emergence	doy (=day of year)
	Flowering (or heading)	doy
	Yellow ripeness	doy
	Leaf area index*	m/m
	Yield*	kg DM/ha
Initial status	Water content	Per layer (m^3/m^3)
	Soil mineral N	Per layer (kg/ha)
Management	Sowing date	doy
	Harvest date	doy
	N fertilization	doy, fert. type, amount (kg N/ha)
	Tillage	doy, type, depth (m)
	Previous crop sowing/harvest	doy
	Previous crop yield	(kg/ha)

DM, dry matter.

* Data not provided for the test year 2009.

results on phenology, LAI and biomass development, yield and N uptake for analysis. After the first check of results, modellers were allowed to correct their results only if errors in inputs or other mistakes in the simulation setup were identified.

Methods for evaluating model performance and assessing uncertainties

For assessing and comparing model performance a set of statistical indicators was calculated in line with those reported by Palosuo *et al.* (2011) and Rötter *et al.* (2012). Grain DM yields and grain N content simulated by the various models and multi-model mean (MMM) were compared with observed values. Using multi-model ensembles has proved to provide robust estimates of growth and yield indicators (Palosuo *et al.* 2011; Rötter *et al.* 2012; Asseng *et al.* 2013) and especially can provide better estimates than any individual model when multiple output variables are considered (Martre *et al.* 2015).

The root mean square error (RMSE) was taken as a measure of relative average difference between the model estimates and measurements: it describes the average absolute deviations between the simulated and observed values.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2} \qquad (1)$$

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

where N is the number of estimate–observation pairs, P_i is the model prediction and O_i is the observed value of the model i.

Mean bias error (MBE) was taken as an indicator of under- or over-estimation, i.e. the direction and magnitude of bias:

$$\mathrm{MBE} = N^{-1} \sum_{i=1}^{N} (P_i - O_i) \quad (2)$$

The index of agreement (IA) developed by Willmott (1981) was used as a more general indicator of model efficiency.

$$\mathrm{IA} = 1 - \frac{N \times \mathrm{MSE}}{\mathrm{PE}} \quad (3)$$

where $\mathrm{PE} = \sum_{i=1}^{N} \left(|\dot{P}| + |\dot{O}| \right)^2$ and where $\dot{P} = P_i - \overline{O}$ and $\dot{O} = O_i - \overline{O}$, and $\overline{O}$ is the mean of the observed variable. Vertical bars indicate absolute values and MSE equals Eqn (1) without square root. The IA is the ratio of the sum of squared errors (N × MSE) and the largest potential error (PE) with respect to the mean of observed values. This parameter is sensitive to the model mean and to the peak values, and insensitive to low magnitude values (Willmott 1981). The IA can have values within the range [0, 1] and values closer to 1 indicate better simulation quality.

For the N fertilizer experiments, measurements from replicate plots ($n = 4$) were taken, allowing for an analysis of the uncertainties in observed LAI, yield and grain N uptake due to measurement errors and spatial variability in field conditions. Ranges and RMSE of simulated values were compared with standard deviations of observations.

Simulated yield responses to N were also analysed using the quadratic model that is commonly used to describe the relationship between fertilizer rates and yield responses (Schabenberger & Pierce 2002):

$$y = y_0 + aN + bN^2 \quad (4)$$

where y is the grain yield, y_0 is the grain yield at $N = 0$, a and b are fitted constants, and N is the N fertilizer rate. The quadratic model was fitted to simulated grain yield of all crop models, MMM and observed data using the NLIN procedure of SAS software (SAS Institute Inc., Cary, NC, USA). The economically optimal N rate was calculated with 1€/kg N fertilizer cost and 0·15 €/kg DM price of barley grains, based roughly on N fertilizer and grain prices in Finland in 2000s, and then income was maximized by deducting fertilizer cost from the value of the barley yield. Maximizing Eqn (5) resulted in Eqn (6), which could be solved using only the fitted constants from Eqn (4).

$$f(N) = 0.15 \times (y0 + aN + bN^2) - N \times 1 \quad (5)$$

$$N = \frac{1 - 0{\cdot}15a}{0{\cdot}15 \times 2b} \quad (6)$$

where f(N) is the gross profit (€/ha) the farmer earns due to N fertilizer.

RESULTS

Experimental years

The general weather conditions of the three seasons included in the current study were similar during the growing period for barley (Fig. 1). According to measured data from the sites and from the nearby field experiments (data not shown), soil moisture remained sufficient over all seasons in the 0–60 cm soil layer. However, measurements indicated dry soil (soil water content <30% of plant available water capacity) from 29 to 38 DAS in the uppermost (0–30 cm) soil layer in year 2002. In 2008 and 2009, the top layer dried <30% of plant available water capacity at 42 and 50 DAS, respectively. In the deeper (30–60 cm) soil layer, soil water contents decreased <30% of plant available water capacity around 75 DAS in 2002 and 2009 but at 55 DAS in 2008.

The observed maximum LAI of all N treatments (Fig. 6) and the LAI dynamics in N90 (data not shown) were similar in 2008 and 2009. Measured DM yields in 2002, 2008 and 2009 for the standard Finnish N rate (78 kg N/ha in 2002 and 90 kg N/ha in 2008 and 2009) were 4010 kg/ha (S.D. = 161 kg/ha, $N = 20$), 4430 kg/ha (S.D. = 305 kg/ha, $N = 4$) and 4560 kg/ha (S.D. = 109 kg/ha, $N = 4$), respectively. Twenty observation sites on the same field in 2002 showed a considerable variation (range 3390–4100 kg/ha). In 2008, measured grain yield without N fertilizer was considerably higher, 3200 kg/ha (S.D. = 268 kg/ha, $N = 4$) compared with non-fertilized yield of 2009 (1840 kg/ha, S.D. = 152 kg/ha, $N = 4$) or average yields in the non-fertilized controls of the Finnish field experiments at 2400 kg/ha (S.D. = 800 kg/ha, $N = 60$) (Valkama *et al.* 2013). Yield response to N fertilizer was very different between 2008 and 2009, as in 2008 the N rate of 60 kg/ha led to 0·96 of maximum grain yield obtained with 150 kg N/ha, while in 2009 yield response to N fertilizer continued until 150 kg N/ha (Fig. 2).

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

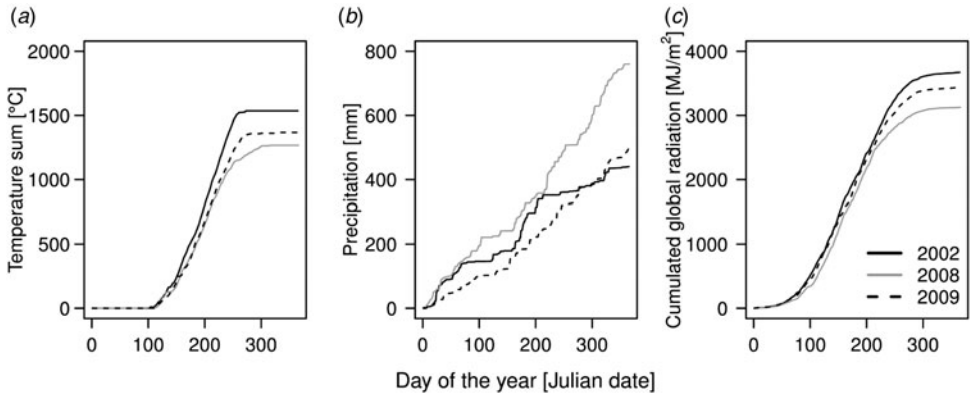


Fig. 1. (*a*) Temperature sum (>5 °C) (°C), (*b*) cumulated precipitation (mm) and (*c*) cumulated global radiation (MJ/m^2) at Jokioinen in 2002, 2008 and 2009.

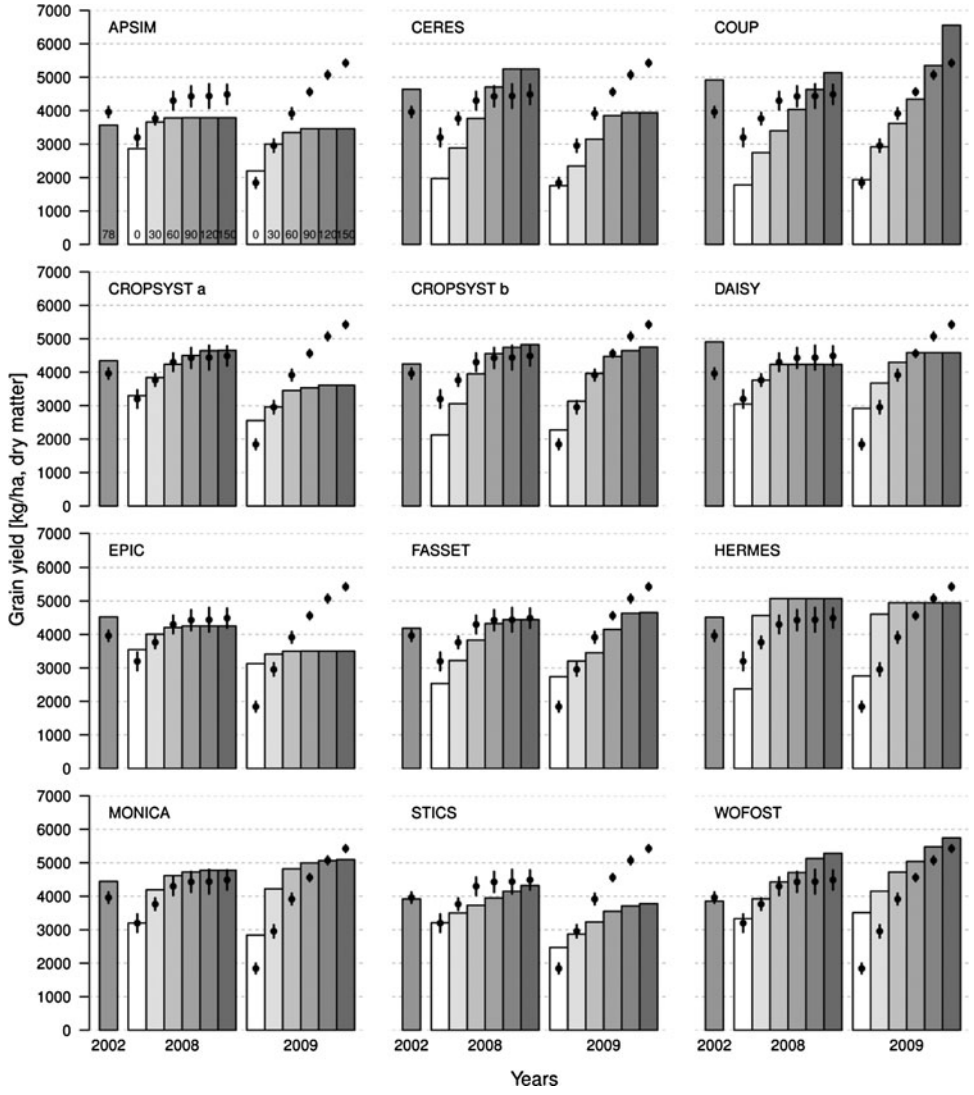


Fig. 2. Simulated (bars) and observed (dots) dry matter grain yield in 2002, 2008 and 2009. The bars on the observed yields show standard deviation.

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

Observed individual grain weight was clearly lower in 2008 (36 mg) than in 2002 (42 mg) and 2009 (47 mg) for treatments with standard N rates (78–90 kg/ha), which may have been the result of a shorter grain-filling period in 2002 and 2008 as compared with 2009. This was also found in data from farms of southern Finland where individual grain weight of cultivar Annabell was 37 mg in 2008 and 44 mg in 2009 (grain quality data of Finnish Food Safety Authority, unpublished). Field observations did not indicate any reasons (e.g. lodging or diseases) that would explain the difference in grain weights. The main observed difference between the growing seasons was the early and vigorous tillering after emergence in 2009 and a heavy rainfall event on 7–8 August 2008. In 2009, the low daily average temperature (7 °C lower than in 2008) 20–25 DAS, together with a 25 mm rainfall event (3–4 June) resulted in early and vigorous tillering. Field observations on 2 July 2009 counted 3–4 tillers in treatments with sufficient N rates (>60 kg/ha). In 2008 a heavy rainfall event caused soil water content to rise from 28 to 40% vol. in the 0–30 cm soil layer after 85 DAS, while in 2009, soil water content in 0–30 cm layer remained below 30% vol. until harvest.

Grain yield

Calibration

Grain yields were calibrated reasonably well in 2002 and 2008 (Figs 2 and 3). In 2002, simulated grain yields ranged between 3570 and 4920 kg DM/ha, i.e. –11 to +23% compared with an observed yield of 4000 kg/ha. In 2008, CROPSYSTa, DAISY, EPIC, FASSET, MONICA, STICS and WOFOST showed good performance with RMSE values lower than 500 kg/ha and IA values > 0·8 in calibration (Fig. 3). CERES, COUP and HERMES produced the highest RMSE values of 720–870 kg/ha in calibration and their IA values were still >0·7. APSIM, COUP and STICS under-estimated the grain yields, whereas HERMES, MONICA and WOFOST over-estimated yields (Fig. 3). In the 2008 experiment, standard deviation of observed grain yields in N treatments varied from 180 to 360 kg/ha, which means that the best model calibrations (CROPSYSTa, DAISY, EPIC, MONICA and STICS) were within the experimental error observed in the field. Multi-model mean produced RMSE values of 220 kg/ha and IA of 0·9 in 2008.

Testing against independent data set

For the test year 2009, the performance of models decreased compared with the calibration year 2008, as indicated by RMSE that approached 1000 kg/ha for several models (Fig. 3). In the group of well-calibrated models, CROPSYSTa, EPIC, STICS and WOFOST produced considerably higher RMSE for the test year. CROPSYSTb, DAISY and FASSET were able to retain their good performance with RMSE values <700 kg/ha and IA values still close to 0·9 for the test year.

In the yield observations within N treatments, standard deviation was low and varied from 70 to 180 kg/ha in 2009. Multi-model mean produced RMSE values of 540 kg/ha and IA value of 0·9 in 2009 (Fig. 3) showing that overall, model performances were not considerably weaker than in the calibration year 2008.

Yield response to Nitrogen

None of the models was able to reproduce the general pattern of observed grain yield fertilizer response of 2008 and 2009 simultaneously, which included a levelling off at medium N rate as in 2008 and a steep response beyond high N rate as in 2009. Observed grain DM yield response to N fertilizer in 2008 levelled off between N60 and N90 although the maximum yield of 4490 kg/ha (S.D. = 293 kg/ha) was recorded in the N150 treatment (Fig. 2). In the test year 2009, the observed grain yield response for N was considerably stronger and continued until N150 reaching 5420 kg/ha (S.D. = 107 kg/ha) of grain yield.

The fitted second degree polynomials showed that CERES, COUP and CROPSYSTb were calibrated to under-estimate grain yield without N fertilizer by >1000 kg/ha than was observed in 2008 (Table 6). The overall calibrated yield response to N in 2008 was best simulated by CROPSYSTa, DAISY, EPIC, MONICA, STICS, FASSET and WOFOST (Figs 2 and 3).

The impacts of weather and management data provided on simulated yields were similar in 2008 and 2009 according to most models and they produced similar moderate N response curves for both years (Fig. 2 and Table 6). Based on these similar N response curves, DAISY, EPIC, HERMES, MONICA and WOFOST over-estimated the grain yield without N fertilizer by *c.* 1000 kg/ha or more (Fig. 3, Table 6). Despite using the same soil parameterization data and having similar weather conditions, APSIM, CERES, CROPSYSTa, DAISY, EPIC, MONICA and

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

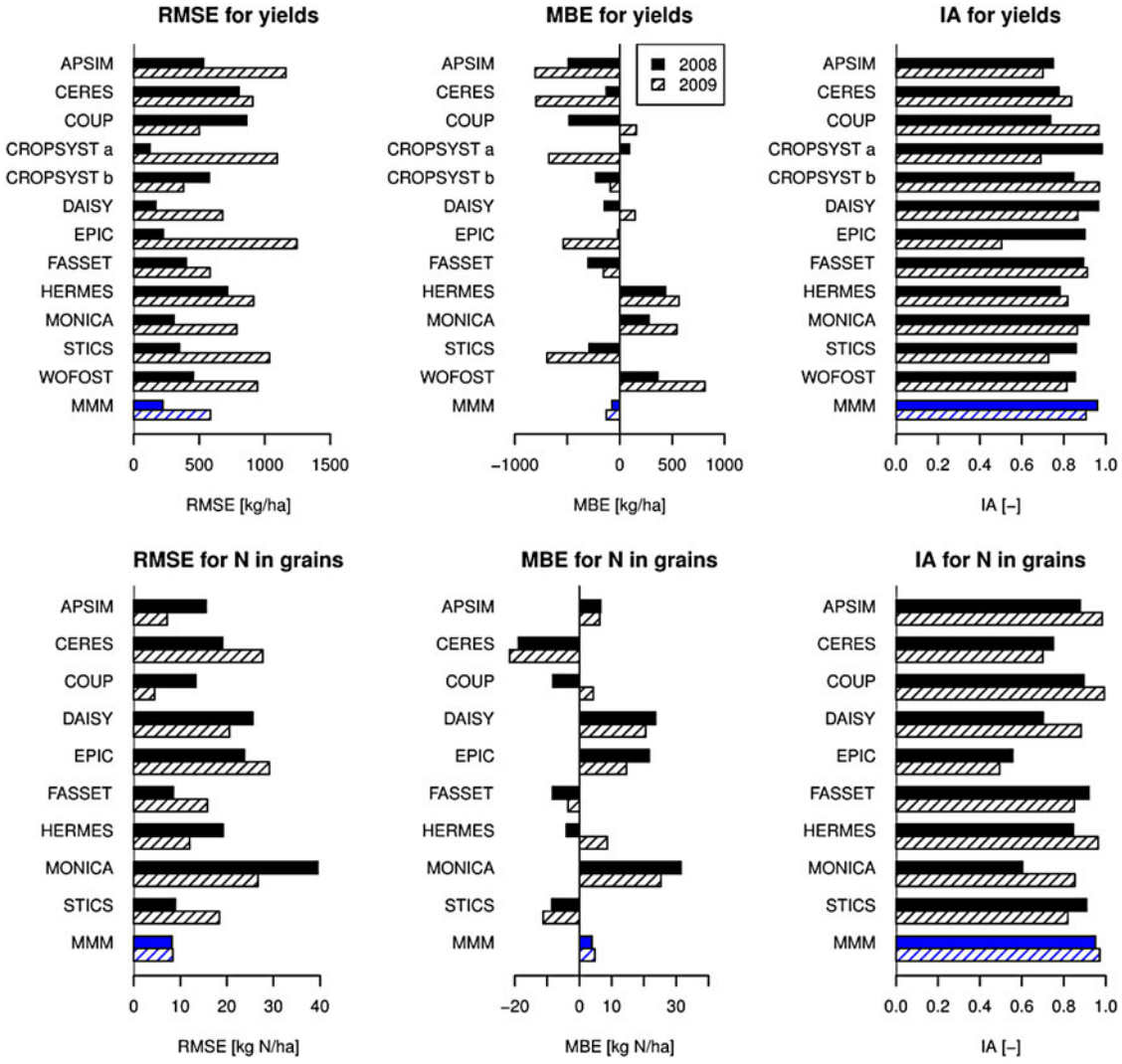


Fig. 3. Statistical indicators of model performance for 2008 (■) and 2009 (▨). MMM shows the multi-model mean.

STICS estimated lower yield without N fertilizer in 2009 compared with 2008, as was also observed in the field experiment. However, the largest yield decrease among the models was only 700 kg/ha, while the actual yield decrease in the field was 1350 kg/ha.

Yield estimates of HERMES, MONICA and WOFOST were closest to the observed maximum yield of 5420 kg/ha (S.D. = 107 kg/ha), whereas COUP over-estimated maximum yield by 1130 kg/ha. COUP provided exceptional results as it was the only model that performed better in the test year than in calibration year 2008, and it showed a steep yield response to N in both years.

The majority of models simulated grain yield response to N fertilizer to level off at the same 90 kg N/ha N rate as in 2008 (Fig. 2, Table 6). However, COUP, STICS, FASSET and WOFOST estimated a higher yield response to N, levelling off only towards 120–150 kg N/ha in both 2008 and 2009. Economically optimum N rates were mostly lower than N rates giving the highest grain yields (biological optimum, Table 6). In 2009, yield response to N was both biologically and economically profitable until N150. FASSET and WOFOST succeeded in increasing their economically optimal N rates to N150 in 2009. The economical optimum of MMM was also increased from N90 in 2008 to N120 in 2009.

Nitrogen uptake

Information on N uptake in grain was not made available to the modellers in the calibration step. In the calibration year 2008, observed grain N uptake from the plots without N fertilizer averaged 51 kg N/ha

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

Table 6. *Second degree polynomials fitted for the model simulated, multi model mean (MMM) and observed grain dry matter yields in 2008 and 2009*

	2008				2009			
	a	max Y	N maxY	N Econ	a	max Y	N maxY	N Econ
APSIM	2987	3910	102	64	2273	3567	110	79
CERES	1875	5431	174	146	1648	3996	143	114
COUP	1821	6128	291	226	2036	>6476	>150	>150
CROPSYSTa	3304	4662	137	91	2545	3643	122	77
CROPSYSTb	2085	4839	140	116	2248	4748	141	114
DAISY	3107	4353	107	76	2922	4669	118	92
EPIC	3591	4315	108	54	3163	3547	104	10
FASSET	2501	4477	136	105	2687	5770	348	217
HERMES	2693	5413	103	90	3039	5232	102	86
MONICA	3300	4886	111	85	2979	5219	113	94
STICS	3216	4793	332	99	2452	3807	167	98
WOFOST	3339	5436	204	138	3525	6028	228	159
MMM	2818	4655	136	102	2627	4520	148	109
Observed	3209	4527	119	83	1850	5519	182	152

a is yield without N fertilizer (kg DM/ha); max Y is the highest grain yield (kg DM/ha) obtained with the N rate; N maxY (kg N/ha); N Econ is the N rate (kg N/ha) to get the best economical result when the price ratio of 1 kg fertilizer to 1 kg grain is 6·67 (1/0·15), obtained by maximizing the differential of yield value (yield from second degree polynomial × grain price) and fertilizer cost.

(s.d. = 4 kg N/ha). The highest grain N uptake was 93 kg N/ha (s.d. = 6 kg N/ha) in the N150 treatment. In 2009, grain N uptake from the plots without N fertilizer averaged 25 kg N/ha (s.d. = 2 kg N/ha). The highest observed grain N uptake was 106 kg N/ha (s.d. = 1 kg N/ha) in N150 treatment.

Models followed the pattern of increased N uptake in grain due to increased N fertilizer rates in both years (Fig. 4). The RMSE values of the models ranged from 10 to 40 kg N/ha in 2008 (Fig. 3) and from 5 to 30 kg N/ha in 2009, while RMSE of the MMM was 8 kg N/ha in both years. DAISY, EPIC and MONICA had a tendency to over-estimate the grain N uptake and CERES had a tendency to underestimate the grain N uptake. The IA values were close to 0·8 for all models except EPIC.

All models, except EPIC, produced almost linear grain N uptake responses to N rate (Fig. 4), which were well described by the second-degree polynomials (Table 7). EPIC showed high and HERMES low grain N uptake of unfertilized barley in both years (Fig. 4 and Table 7). COUP and CERES simulated *c.* 25 kg N/ha uptake from unfertilized treatments, which under-estimated grain N uptake in 2008 but fitted well in 2009. STICS estimated best the crop N uptake in unfertilized barley, but STICS under-estimated N uptake in grain at the high N rates in 2009. MONICA and DAISY tended to over-estimate N in grains at high N rates.

Above-ground N uptake at maturity was reported by all models, except EPIC and WOFOST. On average, models allocated 0·68 of total above-ground N uptake to grains. The lowest proportion of N in grains from total above-ground N uptake, 0·56, was simulated by STICS and the highest, 0·79, by HERMES.

Soil water content

Similar to observations in 2002, CERES, DAISY, HERMES, MONICA, STICS and WOFOST estimated low soil water contents at 29–38 DAS (data not shown). In 2008, MONICA and STICS estimated a short period of water stress at 30 DAS in the 0–30 cm soil layer, which was not indicated by the measurements. Half of the models (CERES, CROPSYSTa, DAISY, EPIC, HERMES and MONICA) estimated, in line with observations that the uppermost (0–30 cm) soil water content decreased below 28% vol. (30% of plant available water left in soil) at 40 and 50 DAS in 2008 and 2009, respectively. For the 31–90 cm soil layer, all models estimated the soil water content higher than 33% vol. (30% of plant available water left in soil).

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

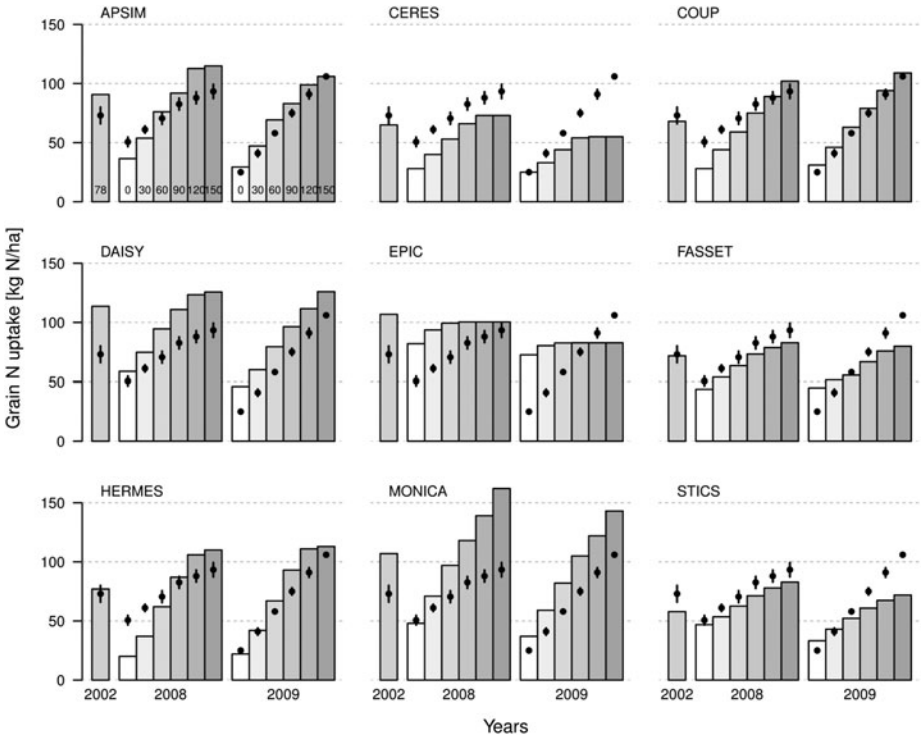


Fig. 4. Measured nitrogen (N) in grain (dots) and model estimates (bars) in 2002, 2008 and 2009. CROPSYSTa, CROPSYSTb and WOFOST did not simulate grain N values. The bars on the observed N in grain show standard deviation.

Table 7. *Second degree polynomials to describe nitrogen (N) uptake (kg N/ha) in grains as function of fertilizer N rate (kg N/ha) (N in grains* $= a + b \times N\ rate + c \times (N\ rate)^2$*)* for the model simulations, multi model mean (MMM) and observations in 2008 and 2009*

	2008				2009			
	Coefficients				Coefficients			
	a	*b*	*c*	RMSE	*a*	*b*	*c*	RMSE
APSIM	34·4	0·80	−0·0016	3·1	28·2	0·76	−0·0015	1·5
CERES	26·4	0·57	−0·0017	1·8	23·4	0·45	−0·0016	1·7
COUP	27·8	0·55	−0·0004	0·4	30·6	0·55	−0·0002	0·4
CROPSYSTa	–	–	–	–	–	–	–	–
CROPSYSTb	–	–	–	–	–	–	–	–
DAISY	56·6	0·77	−0·0019	2·2	45·0	0·59	−0·0003	1·1
EPIC	83·1	0·35	−0·0016	1·2	73·6	0·20	−0·0009	1·0
FASSET	43·3	0·40	−0·0009	0·7	44·4	0·22	0·0002	1·7
HERMES	16·3	0·91	−0·0018	4·0	18·5	1·01	−0·0023	4·0
MONICA	47·9	0·82	−0·0004	0·9	36·6	0·79	−0·0006	1·0
STICS	46·0	0·30	−0·0004	0·8	32·9	0·37	−0·0007	0·4
WOFOST	–	–	–	–	–	–	–	–
MMM	42·9	0·60	−0·0011	1·2	36·9	0·55	−0·0009	1·1
Observed	50·2	0·42	−0·0009	1·1	24·6	0·57	−0·0002	0·4

Root mean square error (RMSE) shows the error between second degree polynomial equation and the modelled results.
* *a*, intercept; *b*, coefficient of the first degree term; *c*, coefficient of the second degree term.

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

Crop phenology

Calibration results (2002 and 2008) for phenology (Fig. 5) did not match perfectly with the phenological data provided for calibration. The RMSE calculated on anthesis and maturity dates in 2002 and 2008 was 6–8 days for APSIM, CERES, CROPSYSTb, DAISY, FASSET, STICS and WOFOST but for COUP, CROPSYSTa, EPIC, HERMES and MONICA it was 10 days or more. Taking into account the uncertainties often related to observations of anthesis and, more generally, under partially N-stressed conditions, the results obtained by the models can be considered fairly satisfactory at this point. Model estimates for the length of growing period from sowing (growth stage (GS) 00 according to the BBCH scale; BBCH (Biologische Bundesanstallt Für Land-Und Forstwirtschaft) 1997) to physiological maturity (GS 90) were on average 93 and 100 days for 2002 and 2008, respectively, compared with the observed lengths of the growing period, 97 and 100 days. In particular, COUP, CROPSYSTa and EPIC showed fast phenological development (Fig. 5). Simulated emergence (GS 12) differed among the models only in APSIM, which estimated a 15-day longer period for emergence than the other models.

In the test year 2009, model estimates for the length of the growing season were on average 103 days, which is close to the observed length of the growing period, 102 days (Fig. 5). COUP, CROPSYSTa and EPIC were 5–10 days faster and DAISY, FASSET and STICS were 7 days slower in their development than observed for 2009.

The phenological development of MONICA and STICS reacted to low N supply by speeding up the development. In STICS this difference between unfertilized and N fertilized treatments was only one day, whereas MONICA estimated 11 days shorter growing period in 2008 and 9 days shorter in 2009 for the low-N treatments. In MONICA, anthesis began at the same time for all N treatments and physiological maturity was achieved faster with low N supply. While observed data showed no clear phenological response to N treatments, but differences of a few days may have gone unnoticed in the field observations.

Leaf area index, biomass and harvest index

Leaf area index

There were considerable differences between modelled and observed LAI (Fig. 6 and Table 8), even though the modellers were provided with observed LAI in 2002 and 2008 from all N treatments for calibration. In 2002, the observed maximum LAI was 3·7 (s.d. = 0·75) and in 2008 it increased with the increase in N rates from 3·2 (s.d. = 0·48) to 7·1 (s.d. = 0·07) (Fig. 6). The LAI responses to N fertilization during both calibration seasons were best estimated by HERMES (RMSE = 1·2) and STICS (RMSE = 1·3). COUP, DAISY, EPIC, FASSET and WOFOST tended to over-estimate the maximum LAI whereas CROPSYSTa and MONICA mostly underestimated it (Fig. 6). The maximum LAI response to N in 2008 was modelled well by CROPSYSTb, FASSET and WOFOST with RMSE values <1·0 (Table 8).

In the test year 2009, observed maximum LAI increased from 1·5 (s.d. = 0·33) in the unfertilized treatment to 6·9 (s.d. = 0·35) in N150 treatment. Most models could not reproduce that but estimated similar maximum LAIs, between 4 and 5, as was observed in 2008 (Fig. 6). The models (COUP, FASSET and WOFOST) that simulated high maximum LAI, as was observed in 2008, fitted best to high LAI observations of 2009 (Fig. 6). The best estimates of the LAI curve of N90 in 2009 were simulated by APSIM, CROPSYSTb, DAISY and HERMES, all having RMSE < 1·0 (Table 8). FASSET estimated the steep LAI response to N rates best in the test year and its RMSE was 1·0 (Table 8).

Most models simulated similar maximum LAI responses to N for 2008 and 2009 and the majority were able to simulate an increase in maximum LAI with increasing N rates (Fig. 6).

The exceptions were EPIC, which estimated high LAI maxima with no response for N rate (RMSE in 2008 was 1·9 and in 2009 was 2·1), and APSIM, DAISY and MONICA, which estimated low maximum LAI and showed no or minimal response to increasing N rates.

Biomass

Modellers were not provided with the data on above-ground biomass dynamics. The simulated total above-ground biomass varied considerably among models (Fig. 7). Calibration at fertilizer N rates of 78 kg N/ha in 2002 and 90 kg N/ha in 2008 resulted in a wide range of estimates of total above-ground biomass at maturity, from 5970 kg/ha to 12380 kg DM/ha. APSIM provided generally the highest total above-ground biomass estimates and COUP the lowest.

In the test year, simulated total above-ground biomass at maturity at N rate of 90 kg N/ha averaged 8960 kg/ha (s.d. = 1673 kg/ha) across all models, and this was close to the observed total above-ground biomass of 7560 kg DM/ha (s.d. = 1476 kg/ha). There

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

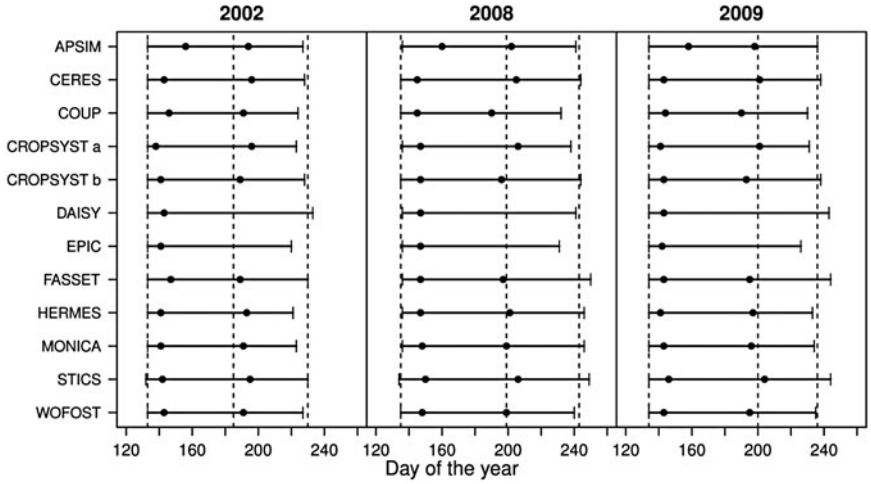


Fig. 5. Comparison between observed and simulated periods from sowing to physiological maturity as bar lines. Observed (dotted lines) and simulated (dots) dates of emergence (growth stage (GS) 12), start of anthesis (GS 60) (completed anthesis –6 days) and physiological maturity (GS 90) in day of the year.

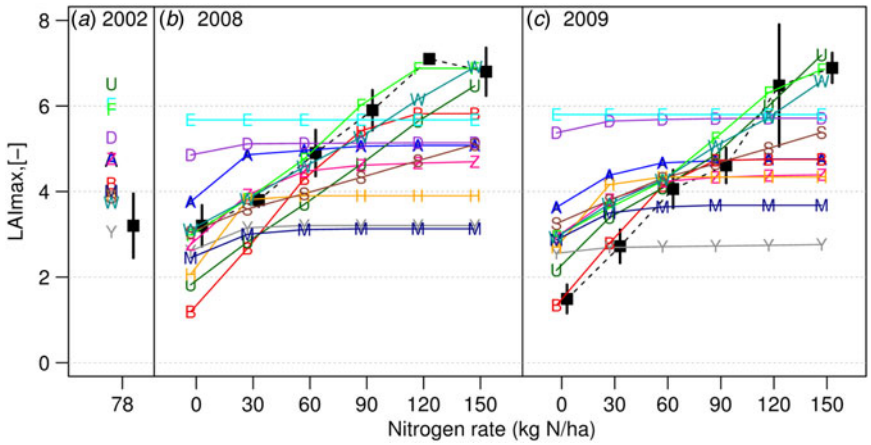


Fig. 6. Maximum leaf area index (LAI) of models and measurements in (*a*) 2002, (*b*) 2008 and (*c*) 2009. Models included are APSIM (A), CERES (B), COUP (U), CROPSYSTa (Y), CROPSYSTb (Z), DAISY (D), EPIC (E), FASSET (F), HERMES (H), MONICA (M), STICS (S) and WOFOST (W). Observations are marked with black squares and their standard deviation with black segment lines.

was a wide range between the models, from 5980 to 10950 kg/ha. The three lowest estimates were from COUP, EPIC and CROPSYSTa, and the three highest were from FASSET, DAISY and APSIM.

All models increased their total above-ground biomass with increasing N rates. In 2008, N rate from 0 kg/ha to 150 kg/ha increased total above-ground biomass 27 kg DM/kg N (S.D. = 11·5 kg DM/kg N) and in 2009 64 kg DM/kg N (S.D. = 10·4 kg DM/kg N) when averaged over all models.

Maximum root biomass estimates ranged from 500 kg/ha of CERES to 3220 kg/ha in EPIC (Fig. 8), while the maximum root biomass estimates were not available from APSIM, CROPSYSTa and STICS. Root biomass estimates of other models except for CROPSYSTb responded to N fertilization with increasing biomass with increasing N rate. Steepest responses were given by COUP, EPIC and FASSET (Fig. 8).

Harvest index

In most models, harvest index (HI) estimates ranged from 0·4 to 0·6 (data not shown), which is plausible when compared with figures reported for spring cereals in Finland (Peltonen-Sainio *et al.* 2008) or the observation of 0·6 in 2009 harvest of N90. The only models that were out of this range were COUP with HI as high as 0·7 and APSIM with low HI of 0·3. COUP, STICS and

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

Table 8. *Model performance (root mean square error (RMSE) values) of estimated leaf area indices (LAIs) calculated based on data from two typical treatments of nitrogen (N) fertilization (78 or 90 kg N/ha) and for all treatments*

	2002 (N78) and 2008 (N90)	2009 (N90)	2008 all N rates	2009 all N rates
Comparisons*	8	12	24	27
APSIM	1·98	0·76	1·14	1·18
CERES	1·88	1·06	1·15	1·19
COUP	2·62	1·19	1·91	1·26
CROPSYSTa	1·48	1·74	1·42	1·89
CROPSYSTb	1·58	0·83	0·91	1·27
DAISY	2·72	0·90	1·68	1·55
EPIC	3·41	1·67	1·86	2·10
FASSET	1·84	1·04	0·86	0·97
HERMES	1·20	0·76	1·20	1·36
MONICA	1·86	1·26	1·69	1·59
STICS	1·33	1·16	1·03	1·37
WOFOST	1·49	1·10	0·80	1·25

* Number of comparisons between observations and modelling results.

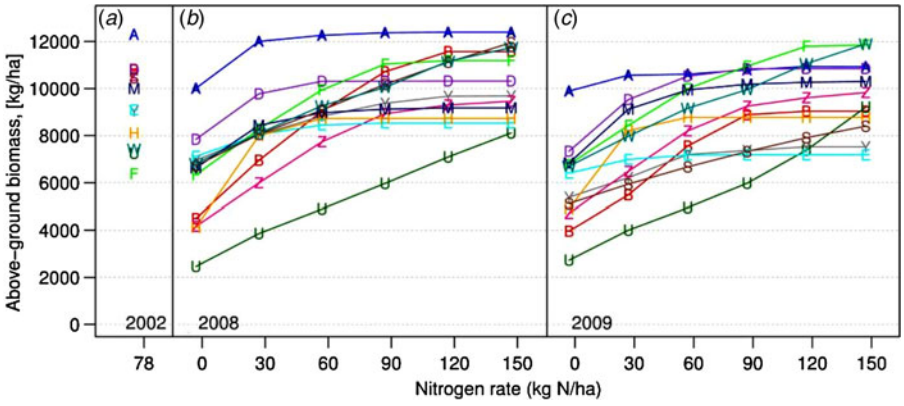


Fig. 7. Modelled total above-ground biomass at physiological maturity (growth stage (GS) 90) in (*a*) 2002, (*b*) 2008 and (*c*) 2009. Models included are APSIM (A), CERES (B), COUP (U), CROPSYSTa (Y), CROPSYSTb (Z), DAISY (D), EPIC (E), FASSET (F), HERMES (H), MONICA (M), STICS (S) and WOFOST (W).

WOFOST were the only models that estimated a slight decrease in HI with increasing N rates.

DISCUSSION

Overall model performance in estimating yield response to Nitrogen

Previous model comparisons have shown that minimal calibration of crop models can lead to a high degree of uncertainty of yield estimates (Palosuo *et al.* 2011; Rötter *et al.* 2012; Asseng *et al.* 2013). The current study shows that more detailed data provided for model calibration does not necessarily result in high accuracy of model simulations when the models are applied to new situations, in particular when simulations are done for sub-optimum management such as N-limited growth, and for environmental conditions that deviate strongly from the calibration conditions, as in the current study. The RMSE values of grain yield were three times higher in test year 2009 compared with calibration year 2008. Interesting topics for future research are how sensitive the model-estimated yield responses are to the selected calibration data and its quality, and to what extent model results can be extrapolated to new situations, beyond the environmental 'data space' of the calibration years.

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

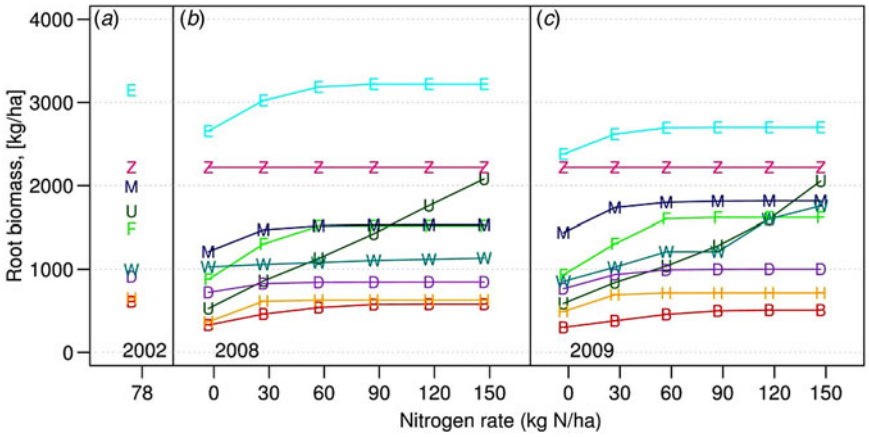


Fig. 8. Modelled root biomass at physiological maturity (growth stage (GS) 90) in (*a*) 2002, (*b*) 2008 and (*c*) 2009. Models included are CERES (B), COUP (U), CROPSYSTb (Z), DAISY (D), EPIC (E), FASSET (F), HERMES (H), MONICA (M) and WOFOST (W).

It was decided to use years for model calibration and testing in chronological order, which is a typical set-up for modelling studies including model calibration and projections for future. The selected years most likely affected the simulated N responses as the yield response to N fertilizer was very different between the years. In further studies, it would be useful to avoid unnecessary restrictions dividing experimental years between calibration and testing. Furthermore, testing the sensitivity of the modelling results on different calibration and test years could also be a useful addition in future modelling studies.

Despite the problems of the individual models to maintain their good performance from calibration to test year, the MMM yield response to N matched well with observations; only three models had lower RMSE (380–580 kg/ha) in 2009 than MMM (590 kg/ha). The MMM relies on the assumption that results of a large group of models will converge close to the 'true value', mainly by error compensation, while the most accurate models might become scaled back to average (Knutti 2010; Asseng *et al.* 2013; Carter 2013: Martre *et al.* 2015). The current results show that MMM cannot always effectively correct systematic errors in model simulations. For instance, as shown by Rötter *et al.* (2012) for a small ensemble (<10) of crop models, due to poor calibration an extreme outlier model can result in a less robust estimate by MMM. In such cases, using the multi-model median could provide a better estimate (Martre *et al.* 2015). Yet, if all the models are (systematically) wrong, as seems to have occurred for estimating LAI in the unfertilized plot in 2009, then the multi-model median will not help. In 2009, all models clearly over-estimated maximum LAI for the N0 plot in 2009. According to the laws of light extinction of leaf canopies, light interception, resulting assimilation, etc., a maximum LAI of 1·5 can at most intercept 0·60 of the light of a canopy with max LAI of 3·0 (de Wit 1965; Goudriaan 1977). Hence, the much lower control yield in 2009 as compared with 2008 and 2002 was to be expected – but all models failed to reproduce observed values. This also clearly calls for in-depth studies to improve models in terms of simulating the dependence of leaf area development on temperature and N availability. Thus, in the current case, due to severe systematic errors in modelling LAI under cold temperatures and low N, the MMM was not able to represent the specific differences of the 2 years either.

Observed differences in yield responses to N between the seasons 2008 and 2009 were quite large even though soil data and weather conditions that were converted to model input did not show large differences. Reasons for these discrepancies could be related to the different ways the models apply to estimate maximum yield and soil N supply.

Maximum yield with optimal N was 20% higher in 2009 than in 2008 and most models under-estimated that response. None of the main factors that have been found to have negative effects on barley yields in Finland, such as early summer drought, excessive rain early in the season, high temperatures around heading or accelerated temperature sum accumulation rates during a period of 2 weeks before heading (Hakala *et al.* 2012), were markedly different in

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

2009 as compared with 2008. Yet there are a number of other adverse weather events that can have a considerable yield impact in a particular year (Van Oort *et al*. 2012; Rötter *et al*. 2013).

After careful re-analysis of the weather data, one adverse weather event was found in the calibration year 2008 and one positive weather event in test year 2009 that can partly explain the difference in observed maximum yields between the 2 years. In 2008, there was heavy rainfall during the last phase of grain filling, which may have led to yield penalties. Although direct grain losses were not reported, the intensive rainfall event may have caused water-saturated soil conditions and interfered with grain filling. Modelled soil water contents increased in all models after heavy rainfall at the beginning of August, but soil water content increased above field capacity only in APSIM and COUP.

In 2009 the positive weather event that resulted in early and vigorous tillering was low daily average temperature after emergence. The observed number of tillers was clearly higher than the reported common value of 1·3 in two-row barley in Finland (Peltonen-Sainio *et al*. 2009). Heads and grains of tillers developed early enough to produce sufficient individual grain size as can be seen from grain weight of 2009. Under northern conditions, grains on side tillers mature later but tillering enhances grain yield especially in two-row barley cultivars (Peltonen-Sainio *et al*. 2009). The number of productive tillers has been noted to increase with moist conditions during early growth (Mäkelä & Muurinen 2012) and with high N rates and cool conditions delaying maturity (Peltonen-Sainio *et al*. 2009). Enhanced tillering increased number of grains and growing conditions were sufficient to increase yield potential. Tillering is not explicitly included in process descriptions of any of the models and this effect was not covered by them indirectly either.

The avoidance or enhancement of stresses caused by drought, diseases, pests, nutrients (other than N), etc. is necessary when modelling N-limited growth. In the current study, disease and pest stresses were ruled out by excluding such years and treatments according to the observations. Considering soil water contents, water deficits cannot be ruled out with certainty. More exact measurements from plots with diverging N rates, to cover possible water shortages in high-yielding treatments, would have been useful to clarify this.

Models were calibrated against the relatively high yield of the N0 treatment in 2008. In 2009, unfertilized grain yield was 40% lower and grain N uptake was 26 N kg/ha lower compared with 2008. Soil characteristics provided to modellers for estimating soil N mineralization were few, including only initial soil C and N concentrations. Soil and weather input did not reveal any clear reason for differences in observed soil N supply between 2008 and 2009. Neither weather conditions in previous winters or pre-sowing in spring revealed any major differences. Detailed simulation results of most models showed that in 2009 soil water content in 0–30 cm at 40–60 DAS was lower than in 2008, which indicates a short period of possible water stress. Information on soil N mineralization was checked from three models: MONICA and STICS provided higher N mineralization in 2008 (63 and 60 kg N/ha) than in 2009 (49 and 40 kg/N ha) from sowing to harvest; COUP on the other hand, showed lower soil moisture content in 2009 than in 2008, but it did not affect the soil N mineralization simulated by the model. Observations of N leaching in a nearby field showed a small leaching event of 3 kg N/ha on 6 June 2009, which could also partly increase the difference of N supply between the years 2008 and 2009. Unknown effects of short water stress periods, leaching events and soil N mineralization are possible causes of different yield responses to N rate between calibration and test year.

Modelling nitrogen dynamics

It appears that the level of detail in process description of N dynamics in soil and plant was not a major factor affecting the different performance of the models. For example, WOFOST with its relatively simple approach using soil N supply estimates based on target yields did not show remarkably weaker performance in comparison with models with more detailed N processes and several soil organic matter pools such as COUP or EPIC. This raises the question of whether, more generally, the models considered in the current study are capable of predicting N processes sufficiently accurately without site-specific calibration as is assumed in many model applications. In the current study, model parameterizations were not compared as models differ widely in their process descriptions. However, an interesting continuation of the study would be to examine calibration effects on simulated responses of multiple models for soil N mineralization, LAI, N uptake, etc. and a wider range of conditions.

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

Barley yields for treatments without N fertilization varied in the current study and they generally seem to vary considerably across non-fertilized sites in Finland (Valkama *et al.* 2013). The main parameters affecting N supply of unfertilized plots are N mineralization from crop residues and soil organic matter. Microbial turnover of C and N, and its interaction with the crop, has long been recognized as a key process for N supply (Kersebaum *et al.* 2007). Simulation of soil N supply, however, requires adequate long-term information of the cultivation history of the site (Petersen *et al.* 2013), since soil organic matter model pools cannot usually be measured directly.

According to Hay & Porter (2006) the main effects of N shortage are found in the plant canopy and expressed by lower shoot numbers, smaller leaves and faster leaf and shoot death. Indeed, field observations showed clear reduction in LAI caused by low N rates. One would thus assume that the models that perform best in their LAI estimations with all N rates, APSIM, CERES and FASSET, would also produce best estimates of DM yield. However, they did not perform better than other models. Furthermore models differed considerably in their evaluations of total above-ground and root biomasses. In future studies, the differences between the models in their total biomass accumulation and allocation should be studied closely.

Sources of uncertainty

The current study shows, like the few comparable studies (Diekkrüger *et al.* 1995), that the same exercise for the same model carried out by different teams can provide very different results. The two CROPSYST teams in the current study applied different model versions (e.g. using two fundamentally different soil water balance routines) and calibration approaches that yielded very different model estimates for yields and other variables. This further emphasizes the important role of calibration for the model results. Model inter-comparisons are thus comparisons of case-specific model realizations and their results cannot easily be generalized.

When calibrating yield response to N rate, the good calibration of unfertilized yield and its N uptake is essential in order to estimate soil N mineralization exactly. In calibration, grain yields of N0 differed from 1780 to 3540 kg/ha among the models, and their total above-ground N uptakes of N0 ranged from 30 to 90 kg/ha. On the other hand, it must be kept in mind that the need of detailed calibrations reduces the generality and possibilities to apply crop models. For instance, if one would force the model to reproduce yield at N0 accurately, this might completely distort its capability to reasonably estimate the yield response curve to increasing N rate. Smith *et al.* (1997) stated that the level of site-specific calibration was one of the main reasons for differences of soil organic matter model performances. They also stated that before models can be used in a truly predictive manner their ability to predict soil organic carbon dynamics without site specific calibration should be improved. For modelling yield response to N fertilizers it would help to have the indigenous soil N supply capacity as an input, as used in Setiyono *et al.* (2011), and then the weather impact on soil N mineralization and N loss processes should be dealt with by the model.

Although it is challenging to parameterize the size and kinetics of different soil organic N pools of the models (Ros *et al.* 2011; Dessureault-Rompré *et al.* 2013), these would be needed in a detailed calibration. Measurement of soil N mineralization potential, which could show the availability from organic N under optimal conditions, is one measure that would support modelling of N mineralization. However, soil N dynamics and mineralization potential show high spatial variability even within uniformly managed fields (Baxter *et al.* 2003; Kersebaum *et al.* 2005) and data sets of the different seasons are not observed at exactly the same locations. Therefore, measurements of soil mineral N during the growing season under the different N treatments would have been very useful for both model calibration and testing. Additionally, data on N fluxes (atmospheric and leaching losses) would have been helpful to better quantify N availability for crop uptake. Since soil moisture and temperature greatly affect soil N mineralization, the models should be accurate in modelling soil water and temperature dynamics. In particular, the need for better understanding of soil moisture and temperature interactions has been raised due to the task of soil N modelling under different climate change scenarios (Guntinas *et al.* 2012; Novem Auyeung *et al.* 2013). A major challenge is therefore to develop robust procedures for parameterizing the models with the relatively sparse data available for soil input at larger scales. This cannot be addressed without properly designed and representative field experiments conducted over sufficient climatic and soil gradients.

To improve the predictive capacity of the models under a wide range of environmental conditions,

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

there still seems to be a need for better understanding and description of the fundamental processes in the various crop-soil-atmosphere sub-systems (i.e. associated modules or routines). This was illustrated in the current study in particular for modelling leaf area development under colder temperatures and low N. When using crop models to project effect of different N fertilizer rates, coupling a crop model with various soil modules, soil N mineralization, soil water and temperature fluxes and crop growth is essential. Particularly, crop models tend to have completely different assumptions about N supply from soil that can lead to high uncertainty when soil N supply can provide a considerable share of crop N demand. In order to improve modelling of various soil processes and crop N uptake, very detailed experimental data sets from a wide range of climate and soil conditions would be needed. Therefore it is suggested that relevant institutes should organize field experiments on at least two comparable soil types (sand and clay) on different European climate conditions (north, south, west, eastern and central Europe) using one cultivar and the same N doses for at least three seasons. All field experiments should include detailed monitoring of soil water content, soil mineral N content and plant biomass accumulation in addition to yield and N uptake measurements. Finally, modellers should also put more emphasis on automatic calibration procedures to decrease the effect of human error on the calibration process.

The authors wish to acknowledge the financial assistance provided under the umbrella of COST action 734 'Impacts of Climate Change and Variability on European Agriculture (CLIVAGRI)' and the work of individual researchers was funded by various bodies: T. Palosuo, T. Salo and R. Rötter, the strategic project MODAGS funded by MTT Agrifood Research Finland, and projects FACCE MACSUR and NORFASYS (decision nos. 268277 and 292944) funded by the Ministry of Agriculture and Forestry and the Academy of Finland, respectively; K. C. Kersebaum performed parts of the current study under the umbrella of FACCE MACSUR funded by the German Federal Office for Agriculture and Food and COST ES1106; C. Nendel was supported by ZALF in-house funds; J.E. Olesen and R.H. Patil, CRES funded by Danish Strategic Research Council; P. Hlavinka was supported by the Ministry of Education, Youth and Sports of CR within the National Sustainability Program I (NPU I), grant number LO1415; M. Trnka, Crop modelling as a tool for increasing the production potential and food security of the Czech Republic funded by National Agency for Agricultural Research (QJ1310123) and project LD 13030 – Water resources in the Czech Agriculture under the Climate Change conditions - CZECH-AGRIWAT; Mr Antti Ristolainen (2002) and Mr Petri Kapuinen (2008 and 2009) are acknowledged for providing data from their field experiments.

REFERENCES

Abrahamsen, P. & Hansen, S. (2000). Daisy: an open soil-crop-atmosphere system model. *Environmental Modelling & Software* **15**, 313–330.

Addiscott, T. M. & Whitmore, A. P. (1991). Simulation of solute leaching in soils of differing permeabilities. *Soil Use and Management* **7**, 94–102.

Anwar, M. R., O'Leary, G. J., Rab, M. A., Fisher, P. D. & Armstrong, R. D. (2009). Advances in precision agriculture in south-eastern Australia. V. management zones. *Crop and Pasture Science* **60**, 901–911.

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

Baxter, S. J., Oliver, M. A. & Gaunt, J. (2003). A geostatistical analysis of the spatial variation of soil mineral nitrogen and potentially available nitrogen within an arable field. *Precision Agriculture* **4**, 213–226.

BBCH (Biologische Bundesanstallt Für Land-Und Forstwirtschaft) (1997). *Growth Stages of Mono-and Dicotyledonous Plants: BBCH Monograph*. Berlin: Blackwell Wissenschafts-Verlag.

Berntsen, J., Hauggard-Nielsen, H., Olesen, J. E., Petersen, B. M., Jensen, E. S. & Thomsen, A. (2004). Modelling dry matter production and resource use in intercrops of pea and barley. *Field Crops Research* **88**, 69–83.

Borg, G. C., Jansson, P. E. & Linden, B. (1990). Simulated and measured nitrogen conditions in a manured and fertilised soil. *Plant and Soil* **121**, 251–267.

Burns, I. G. (1974). A model for predicting the redistribution of salts applied to fallow soils after excess rainfall or evaporation. *Journal of Soil Science* **25**, 165–178.

Cannavo, P., Recous, S., Parnaudeau, V. & Reau, R. (2008). Modeling N dynamics to assess environmental impacts of cropped soils. *Advances in Agronomy* **97**, 131–174.

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

Carter, T. R. (2013). Multi-model yield projections. *Nature Climate Change* **3**, 784–786.

Cassman, K. G., Dobermann, A., Walters, D. T. & Yang, H. (2003). Meeting cereal demand while protecting natural resources and improving environmental quality. *Annual Review of Environment and Resources* **28**, 315–358.

Corre-Hellou, G., Faure, M., Launay, M., Brisson, N. & Crozat, Y. (2009). Adaptation of the STICS intercrop model to simulate crop growth and N accumulation in pea-barley intercrops. *Field Crops Research* **113**, 72–81.

de Wit, C. T. (1965). *Photosynthesis of Leaf Canopies*. Verslagen Landbouwkundige Onderzoekingen (Agricultural Research Reports) 663. Wageningen, The Netherlands: Pudoc.

Dessureault-Rompré, J., Zebarth, B. J., Burton, D. L., Gregorich, E. G., Goyer, C., Georgallas, A. & Grant, C. A. (2013). Are soil mineralizable nitrogen pools replenished during the growing season in agricultural soils? *Soil Science Society of America Journal* **77**, 512–524.

Diekkrüger, B., Söndgerath, D., Kersebaum, K. C. & McVoy, C. W. (1995). Validity of agroecosystem models a comparison of results of different models applied to the same data set. *Ecological Modelling* **81**, 3–29.

Dobermann, A., Dawe, D., Roetter, R. P. & Cassman, K. G. (2000). Reversal of rice yield decline in a long-term continuous cropping experiment. *Agronomy Journal* **92**, 633–643.

Donatelli, M., Stöckle, C., Ceotto, E. & Rinaldi, M. (1997). Evaluation of CropSyst for cropping systems at two locations of northern and southern Italy. *European Journal of Agronomy* **6**, 35–45.

Doltra, J., Lægdsmand, M. & Olesen, J. E. (2011). Cereal yield and quality as affected by nitrogen availability in organic and conventional arable crop rotations: a combined modeling and experimental approach. *European Journal of Agronomy* **34**, 83–95.

Doltra, J., Laegsmand, M. & Olesen, P. (2014). Impacts of projected climate change on productivity and nitrogen leaching of crop rotations in arable and pig farming systems in Denmark. *Journal of Agricultural Science, Cambridge* **152**, 75–92.

Eitzinger, J., Trnka, M., Hösch, J., Žalud, Z. & Dubrovsky, M. (2004). Comparison of CERES, WOFOST and SWAP models in simulating soil water content during growing season under different soil conditions. *Ecological Modelling* **171**, 223–246.

Eitzinger, J., Thaler, S., Schmid, E., Strauss, F., Ferrise, R., Moriondo, M., Bindi, M., Palosuo, T., Rötter, R., Kersebaum, K. C., Olesen, J. E., Patil, R. H., Saylan, L., Çaldag, B. & Çaylak, O. (2013). Sensitivities of crop models to extreme weather conditions during flowering period demonstrated for maize and winter wheat in Austria. *Journal of Agricultural Science, Cambridge* **151**, 813–835.

Farkas, C., Ristolainen, A. & Alakukku, L. (2006). Spatial variation of soil properties affecting yield. *Cereal Research Communications* **34**, 155–158.

Finnish Ministry of Agriculture and Forestry (2007). *The Rural Development Programme for Mainland Finland 2007–2013*. Helsinki: Finnish Ministry of Agriculture and Forestry.

Franko, U., Puhlmann, M., Kuka, K., Böhme, F. & Merbach, I. (2007). Dynamics of water, carbon and nitrogen in an agricultural used Chernozem soil in Central Germany. In *Modelling Water and Nutrient Dynamics in Soil-Crop Systems* (Eds K. C. Kersebaum, J.-M. Hecker, W. Mirschel & M. Wegehenkel), pp. 245–258. Dordrecht, The Netherlands: Springer.

Goudriaan, J. (1977). *Crop Micrometeorology: A Simulation Study*. Simulation Monographs. Wageningen, The Netherlands: Pudoc, Wageningen.

Guntinas, M. E., Leiros, M. C., Trasar-Cepeda, C. & Gil-Sotres, F. (2012). Effects of moisture and temperature on net soil nitrogen mineralization: a laboratory study. *European Journal of Soil Biology* **48**, 73–80.

Hakala, K., Jauhiainen, L., Himanen, S. J., Rötter, R., Salo, T. & Kahiluoto, H. (2012). Sensitivity of barley varieties to weather in Finland. *Journal of Agricultural Science, Cambridge* **150**, 145–160.

Hansen, S. (2000). *DAISY, a Flexible Soil-Plant-Atmosphere System Model. Equation Section 1*. Copenhagen: The Royal Veterinary and Agricultural University.

Hansen, S., Jensen, H. E., Nielsen, N. E. & Svendsen, H. (1990). *DAISY: A Soil Plant System Model. Danish Simulation Model for Transformation and Transport of Energy and Matter in the Soil-Plant-Atmosphere System*. Copenhagen: National Agency for Environmental Protection.

Hay, R. K. M. & Porter, J. R. (2006). *The Physiology of Crop Yield*, 2nd edn, Oxford: Blackwell publishing.

Hlavinka, P., Eitzinger, J., Smutný, V., Thaler, S., Žalud, Z., Rischbeck, P. & Kren, J. (2010). The performance of CERES-Barley and CERES-Wheat under various soil conditions and tillage practices in Central Europe. *Die Bodenkultur* **61**, 5–17.

Hlavinka, P., Trnka, M., Kersebaum, K. C., Čermák, P., Pohanková, E., Orság, M., Pokorný, E., Fischer, M., Brtnický, M. & Žalud, Z. (2014). Modelling of yields and soil nitrogen dynamics for crop rotations by HERMES under different climate and soil conditions in the Czech Republic. *Journal of Agricultural Science, Cambridge* **152**, 188–204.

Håkansson, I., Myrbeck, Å. & Etana, A. (2002). A review of research on seedbed preparation for small grains in Sweden. *Soil and Tillage Research* **64**, 23–40.

Hyytiäinen, K., Niemi, J. K., Koikkalainen, K., Palosuo, T. & Salo, T. (2011). Adaptive optimization of crop production and nitrogen leaching abatement under yield uncertainty. *Agricultural Systems* **104**, 634–644.

Janssen, B. H., Guiking, F. C. T., van der Eijk, D., Smaling, E. M. A., Wolf, J. & van Reuler, H. (1990). A system for quantitative evaluation of the fertility of tropical soils (QUEFTS). *Geoderma* **46**, 299–318.

Jansson, P-E. & Karlberg, L. (2012). *Coupled Heat and Mass Transfer Model for Soil-Plant Atmosphere Systems*. Stockholm: Royal Institute of Technology, Department of Civil and Environmental Engineering.

Jansson, P-E. & Thoms-Hjärpe, C. (1986). Simulated and measured soil water dynamics of unfertilized and fertilized barley. *Acta Agriculturae Scandinavica* **36**, 162–172.

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., Huth, N. I., Hargreaves, J. N. G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J. P., Silburn, M., Wang, E., Brown, S., Bristow, K. L., Asseng, S., Chapman, S., McCown, R. L., Freebairn, D. M. & Smith, C. J. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy* **18**, 267–288.

Kersebaum, K. C. (2007). Modelling nitrogen dynamics in soil–crop systems with HERMES. *Nutrient Cycling in Agroecosystems* **77**, 39–52.

Kersebaum, K. C., Lorenz, K., Reuter, H. I., Schwarz, J., Wegehenkel, M. & Wendroth, O. (2005). Operational use of agro-meteorological data and GIS to derive site specific nitrogen fertilizer recommendations based on the simulation of soil and crop growth processes. *Physics and Chemistry of the Earth, Parts A/B/C* **30**, 59–67.

Kersebaum, K. C., Hecker, J-M., Mirschel, W. & Wegehenkel, M. (2007). Modelling water and nutrient dynamics in soil–crop systems: A comparison of simulation models applied on common data sets. In *Modelling Water and Nutrient Dynamics in Soil–Crop Systems* (Eds K. C. Kersebaum, J.-M. Hecker, W. Mirschel & M. Wegehenkel), pp. 1–17. Dordrecht, The Netherlands: Springer.

Knutti, R. (2010). The end of model democracy? An editorial comment. *Climatic Change* **102**, 395–404.

Launay, M., Brisson, N., Satger, S., Hauggaard-Nielsen, H., Corre-Hellou, G., Kasynova, E., Ruske, R., Jensen, E. S. & Gooding, M. J. (2009). Exploring options for managing strategies for pea-barley intercropping using a modeling approach. *European Journal of Agronomy* **31**, 85–98.

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J. W., Rötter, R. P., Boote, K. J., Ruane, A. C., Thorburn, P. J., Cammarano, D., Hatfield, J. L., Rosenzweig, C., Aggarwal, P. K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Challinor, A. J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L. A., Ingwersen, J., Izaurralde, R. C., Kersebaum, K. C., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J. E., Osborne, T. M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M. A., Shcherbak, I., Steduto, P., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J. W. & Wolf, J. (2015). Multimodel ensembles of wheat growth: many models are better than one. *Global Change Biology* **21**, 911–925.

Monteith, J. L. (1981). Climatic variation and the growth of crops. *Quarterly Journal of the Royal Meteorological Society* **107**, 749–774.

Mäkelä, P. & Muurinen, S. (2012). Uniculm and conventional tillering barley accessions under northern growing conditions. *Journal of Agricultural Science, Cambridge* **150**, 335–344.

Nendel, C., Berg, M., Kersebaum, K. C., Mirschel, W., Specka, X., Wegehenkel, M., Wenkel, K. O. & Wieland, R. (2011). The MONICA model: testing predictability for crop growth, soil moisture and nitrogen dynamics. *Ecological Modelling* **222**, 1614–1625.

Nendel, C., Venezia, A., Piro, F., Ren, T., Lillywhite, R. D. & Rahn, C. R. (2013). The performance of the EU-Rotate_N model in predicting the growth and nitrogen uptake of rotations of field vegetable crops in a Mediterranean environment. *Journal of Agricultural Science, Cambridge* **151**, 538–555.

Novem Auyeung, D. S., Suseela, V. & Dukes, J. S. (2013). Warming and drought reduce temperature sensitivity of nitrogen transformations. *Global Change Biology* **19**, 662–676.

Nykänen, A., Salo, T. & Granstedt, A. (2009). Simulated cereal nitrogen uptake and soil mineral nitrogen after clover-grass leys. *Nutrient Cycling in Agroecosystems* **85**, 1–15.

Palosuo, T., Kersebaum, K. C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J. E., Patil, R. H., Ruget, F., Rumbaur, C., Takac, J., Trnka, M., Bindi, M., Caldag, B., Ewert, F., Ferrise, R., Mirschel, W., Saylan, L., Siska, B. & Rötter, R. (2011). Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. *European Journal of Agronomy* **35**, 103–114.

Patil, R. H., Lægdsmand, M., Olesen, J. E. & Porter, J. R. (2012). Sensitivity of crop yield and N losses in winter wheat to changes in mean and variability of temperature and precipitation in Denmark using the FASSET model. *Acta Agriculturae Scandinavica. Section B: Plant and Soil* **62**, 335–351.

Peltonen-Sainio, P., Muurinen, S., Rajala, A. & Jauhiainen, L. (2008). Variation in harvest index of modern spring barley, oat and wheat cultivars adapted to northern growing conditions. *Journal of Agricultural Science, Cambridge* **146**, 35–47.

Peltonen-Sainio, P., Jauhiainen, L., Rajala, A. & Muurinen, S. (2009). Tiller traits of spring cereals under tiller-depressing long day conditions. *Field Crops Research* **113**, 82–89.

Petersen, S. O., Schjønning, P., Olesen, J. E., Christensen, S. & Christensen, B. T. (2013). Sources of nitrogen for winter wheat in organic cropping systems. *Soil Science Society of America Journal* **77**, 155–165.

Pietola, L. & Tanni, R. (2003). Response of seedbed physical properties, soil N and cereal growth to peat application during transition to conservation tillage. *Soil and Tillage Research* **74**, 65–79.

Pietola, L., Tanni, R. & Elonen, P. (1999). Responses of yield and N use of spring sown crops to N fertilization, with special reference to the use of plant growth regulators. *Agricultural and Food Science in Finland* **8**, 423–440.

Probert, M. E., Dimes, J. P., Keating, B. A., Dalal, R. C. & Strong, W. M. (1998). APSIM's water and nitrogen modules and simulation of the dynamics of water and nitrogen in fallow systems. *Agricultural Systems* **56**, 1–28.

Rahn, C. R., Zhang, K., Lillywhite, R., Ramos, C., Doltra, J., de Paz, J. M., Riley, H., Fink, M., Nendel, C., Thorup-Kristensen, K., Pedersen, A., Piro, F., Venezia, A., Firth, C., Schmutz, U., Rayns, F. & Strohmeyer, K. (2010). EU-Rotate_N – a decision support system – to predict environmental and economic consequences of the management of nitrogen fertiliser in crop rotations. *European Journal of Horticultural Science* **75**, 20–32.

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Physics: A Journal of General and Applied Physics, American Physical Society* **1**, 318–333.

Ros, G. H., Hanegraaf, M. C., Hoffland, E. & van Riemsdijk, W. H. (2011). Predicting soil N mineralization: relevance of organic matter fractions and soil properties. *Soil Biology and Biochemistry* **43**, 1714–1722.

Rötter, R. P., Carter, T. R., Olesen, J. E. & Porter, J. R. (2011*a*). Crop-climate models need an overhaul. *Nature Climate Change* **1**, 175–177.

Rötter, R. P., Palosuo, T., Pirttioja, N. K., Dubrovsky, M., Salo, T., Fronzek, S., Aikasalo, R., Trnka, M., Ristolainen, A. & Carter, T. R. (2011*b*). What would happen to barley production in Finland if global warming exceeded 4 °C? A model-based assessment. *European Journal of Agronomy* **35**, 205–214.

Rötter, R. P., Palosuo, T., Kersebaum, K. C., Angulo, C., Bindi, M., Ewert, F., Ferrise, R., Hlavinka, P., Moriondo, M., Nendel, C., Olesen, J. E., Patil, R. H., Ruget, F., Takac, J. & Trnka, M. (2012). Simulation of spring barley yield in different climatic zones of Northern and Central Europe: a comparison of nine crop models. *Field Crops Research* **133**, 23–36.

Rötter, R. P., Höhn, J., Trnka, M., Fonzek, S., Carter, T. R. & Kahiluoto, H. (2013). Modelling shifts in agroclimate and crop cultivar response under climate change. *Ecology and Evolution* **3**, 4197–4214.

Sapkota, T. B., Askegaard, M., Laegdsmand, M. & Olesen, J. E. (2012). Effects of catch crop type and root depth on nitrogen leaching and yield of spring barley. *Field Crops Research* **125**, 129–138.

Schabenberger, O. & Pierce, F. J. (2002). *Contemporary Statistical Models for the Plant and Soil Sciences*. Boca Raton, FL: CRC Press.

Setiyono, T. D., Yang, H., Walters, D. T., Dobermann, A., Ferguson, R. B., Roberts, D. F., Lyon, D. J., Clay, D. E. & Cassman, K. G. (2011). Maize-N. A decision tool for nitrogen management in maize. *Agronomy Journal* **103**, 1276–1283.

Shaffer, M. J. (2002). Nitrogen modeling for soil management. *Journal of Soil and Water Conservation* **57**, 417–425.

Smith, P., Smith, J. U., Powlson, D. S., McGill, W. B., Arah, J. R. M., Chertov, O. G., Coleman, K., Franko, U., Frolking, S., Jenkinson, D. S., Jensen, L. S., Kelly, R. H., Klein-Gunnewiek, H., Komarov, A. S., Li, C., Molina, J. A. E., Mueller, T., Parton, W. J., Thornley, J. H. M. & Whitmore, A. P. (1997). A comparison of the performance of nine soil organic matter models using datasets from seven long-term experiments. *Geoderma* **81**, 153–225.

Stöckle, C. O., Donatelli, M. & Nelson, R. (2003). CropSyst, a cropping systems simulation model. *European Journal of Agronomy* **18**, 289–307.

Strauss, F., Schmid, E., Moltchanova, E., Formayer, H. & Wang, X. (2012). Modeling climate change and biophysical impacts of crop production in the Austrian Marchfeld Region. *Climatic Change* **111**, 641–664.

Stuart, M. E., Gooddy, D. C., Bloomfield, J. P. & Williams, A. T. (2011). A review of the impact of climate change on future nitrate concentrations in groundwater of the UK. *Science of the Total Environment* **409**, 2859–2873.

Sutton, M. A., Erisman, J. W., Leip, A., van Grinsven, H. & Winiwarter, W. (2011). Too much of a good thing. *Nature* **472**, 159–161.

Trnka, M., Dubrovsky, M. & Zalud, Z. (2004). Climate change impacts and adaptation strategies in spring barley production in the Czech Republic. *Climatic Change* **64**, 227–255.

Valkama, E., Salo, T., Esala, M. & Turtola, E. (2013). Nitrogen balances and yields of spring cereals as affected by nitrogen fertilization in northern conditions: a meta-analysis. *Agriculture, Ecosystems and Environment* **164**, 1–13.

Van Oort, P. A. J., Timmermans, B. G. H., Meinke, H. & van Ittersum, M. K. (2012). Key weather extremes affecting potato production in The Netherlands. *European Journal of Agronomy* **37**, 11–22.

Wheeler, T. & von Braun, J. (2013). Climate change impacts on global food security. *Science* **341**, 508–513.

Williams, J. R. (1995). The EPIC model, 1995. In *Computer Models of Watershed Hydrology* (Ed. V. P. Singh), pp. 909–1000. Highlands Ranch, CO: Water Resources Publications.

Willmott, C. J. (1981). On the validation of models. *Physical Geography* **2**, 184–194.

Wolf, J., Hack-ten Broeke, M. J. D. & Rötter, R. P. (2005). Simulation of nitrogen leaching in sandy soils in the Netherlands with the ANIMO model and the integrated modelling system STONE. *Agriculture, Ecosystems and Environment* **105**, 523–540.

Yli-Halla, M. & Mokma, D. L. (2001). Soils in an agricultural landscape of Jokioinen, south-western Finland. *Agriculture and Food Science in Finland* **10**, 33–43.

https://doi.org/10.1017/S0021859615001124 Published online by Cambridge University Press