Global Change Biology

Global Change Biology (2017) **23**, 1258–1281, doi: 10.1111/gcb.13411

A potato model intercomparison across varying climates and productivity levels

DAVID H. FLEISHER[1], BRUNO CONDORI[1], ROBERTO QUIROZ[2], ASHOK ALVA[3], SENTHOLD ASSENG[4], CAROLINA BARREDA[2], MARCO BINDI[5], KENNETH J. BOOTE[4], ROBERTO FERRISE[5], ANGELINUS C. FRANKE[6], PANAMANNA M. GOVINDAKRISHNAN[7], DIEUDONNE HARAHAGAZWE[8], GERRIT HOOGENBOOM[4], SOORA NARESH KUMAR[9], PAOLO MERANTE[5], CLAAS NENDEL[10], JORGEN E. OLESEN[11], PHILLIP S. PARKER[10], DIRK RAES[12], RUBI RAYMUNDO[4], ALEX C. RUANE[13], CLAUDIO STOCKLE[14], IWAN SUPIT[15], ELINE VANUYTRECHT[12], JOOST WOLF[16] and PREM WOLI[17]

[1]*Crop Systems and Global Change Laboratory, USDA-ARS, Beltsville, MD, USA,* [2]*Production Systems and the Environment, International Potato Center, Lima, Peru,* [3]*Desert Agriculture and Ecosystem Program, Kuwait Institute for Scientific Research, Safat, Kuwait,* [4]*Agricultural & Biological Engineering Department, University of Florida, Gainesville, FL, USA,* [5]*Department of Agrifood Production and Environmental Sciences, University of Florence, Florence, Italy,* [6]*Soil, Crop and Climate Sciences, University of the Free State, Bloemfontein, South Africa,* [7]*Central Potato Research Institute, Shimla, India,* [8]*Production Systems and the Environment, International Potato Center SSA, Nairobi, Kenya,* [9]*Centre for Environment Science and Climate Resilient Agriculture, Indian Agricultural Research Institute, New Delhi, India,* [10]*Leibniz Centre for Agricultural Landscape Research, Institute of Landscape Systems Analysis, Müncheberg, Germany,* [11]*Department of Agroecology, Aarhus University, Tjele, Denmark,* [12]*Department of Earth and Environmental Sciences, KU Leuven University, Leuven, Belgium,* [13]*NASA Goddard Institute for Space Studies, New York, NY, USA,* [14]*Biological Systems Engineering, Washington State University, Pullman, WA, USA,* [15]*Earth System Science and Climate Adaptive Land Management, Wageningen University and Research Center, Wageningen, The Netherlands,* [16]*Plant Production Systems, Wageningen University and Research Center, Wageningen, The Netherlands,* [17]*AgWeatherNet Program, Washington State University, Pullman, WA, USA*

Abstract

A potato crop multimodel assessment was conducted to quantify variation among models and evaluate responses to climate change. Nine modeling groups simulated agronomic and climatic responses at low-input (Chinoli, Bolivia and Gisozi, Burundi)- and high-input (Jyndevad, Denmark and Washington, United States) management sites. Two calibration stages were explored, partial (P1), where experimental dry matter data were not provided, and full (P2). The median model ensemble response outperformed any single model in terms of replicating observed yield across all locations. Uncertainty in simulated yield decreased from 38% to 20% between P1 and P2. Model uncertainty increased with interannual variability, and predictions for all agronomic variables were significantly different from one model to another ($P < 0.001$). Uncertainty averaged 15% higher for low- vs. high-input sites, with larger differences observed for evapotranspiration (ET), nitrogen uptake, and water use efficiency as compared to dry matter. A minimum of five partial, or three full, calibrated models was required for an ensemble approach to keep variability below that of common field variation. Model variation was not influenced by change in carbon dioxide (C), but increased as much as 41% and 23% for yield and ET, respectively, as temperature (T) or rainfall (W) moved away from historical levels. Increases in T accounted for the highest amount of uncertainty, suggesting that methods and parameters for T sensitivity represent a considerable unknown among models. Using median model ensemble values, yield increased on average 6% per 100-ppm C, declined 4.6% per °C, and declined 2% for every 10% decrease in rainfall (for nonirrigated sites). Differences in predictions due to model representation of light utilization were significant ($P < 0.01$). These are the first reported results quantifying uncertainty for tuber/root crops and suggest modeling assessments of climate change impact on potato may be improved using an ensemble approach.

Keywords: climate change, crop modeling, model improvement, solanum tuberosum, uncertainty analysis, yield sensitivity

Received 3 May 2016 and accepted 19 June 2016

Correspondence: David H. Fleisher, tel. +1 301 504 6712, fax +1 301 504 5823, e-mail: David.fleisher@ars.usda.gov

Introduction

Climate change, including rises in atmospheric carbon dioxide (CO_2) concentration (C), shifts in air temperature (T), and changes in duration, intensity, and seasonality of rainfall (W), has geospatial and temporal dependencies likely to negatively impact many agriculturally productive regions (Rosenzweig & Parry, 1994; Easterling *et al*., 2000; Hijmans, 2003; Wolfe *et al*., 2007; IPCC, 2014). Global food systems will be increasingly challenged to sustainably address food security needs, given rising populations and competing demands for arable land, water, and other natural resources (Godfray *et al*., 2010; Foley *et al*., 2011; Garnett *et al*., 2013). Crop models have been used to study climate change effects in order to assess risk and impact, evaluate adaptation strategies, and/or test management options for various agricultural commodities (e.g., Rosenzweig & Parry, 1994; Matsuoka *et al*., 2001; Hansen, 2005; Haverkort & Verhagen, 2008; Challinor *et al*., 2009; White *et al*., 2011; Haverkort *et al*., 2013). A majority of simulations and conclusions were drawn on the basis of a single model. However, multiple models exist for most of the globally important crops. These models were generally developed independently of one another and thus differ with respect to the type and manner in which soil–plant–atmospheric science is encapsulated and parameterized in the source code (Tubiello & Ewert, 2002). This variation among model structure leads to additional questions regarding the efficacy of using simulation results from a single model as a basis for evaluating agricultural responses (Wolf, 2002; Rosenzweig *et al*., 2013), and recent literature suggests that a collection, or ensemble, of such models may provide a more nuanced and accurate reflection of reality than any individual one (Martre *et al*., 2015). This study focuses on quantifying the variation, or uncertainty, among potato crop models, including an assessment of predicted climate change responses at multiple production locations.

Potato (*Solanum tuberosum* L.) ranks as the third food crop worldwide, behind rice and wheat, with respect to quantity of production and consumption (Haverkort & Struik, 2015). Despite this agronomic importance, White *et al.* (2011) found that only seven of 231 original research articles simulating effects of climate change on crops included potato. Production is dispersed broadly around the world and encompasses highly contrasting locations with respect to soils, climate, and management. This wide adaptation range coincides with large variation in the plant's cultivated ploidy, particularly as compared with other crops (Watanabe, 2015). Potato is also distinct from other commodities due to a highly indeterminate growth habit and lack of distinct developmental stages (Ewing & Struik, 1992), traits, which challenge crop modelers to establish biologically meaningful relationships between climate, growth, biomass partitioning, and phenology (Ewing & Sandlan, 1995; Vos, 1995). These challenges lead to large variations across potato models with respect to structures and approaches used to simulate growth and development (Table 1). The crop is frequently categorized as drought and T sensitive, with yield declines occurring at moderate levels of soil moisture depletion (Borah & Milthorpe, 1962; Van Dam *et al*., 1996) and with cooler optimal T for yield and rate processes as compared with other economically important commodities (Van Loon, 1981; Stalham & Allen, 2004; Timlin *et al*., 2006). Similar to other plants with a C3 biochemical photosynthetic pathway, potato exhibits positive responses to increasing C when other cultural factors are nonlimiting (e.g., Baker & Allen, 1994; Finnan *et al*., 2005; Kaminski *et al*., 2014). Thus, potato crop models intended to evaluate impacts of climate change from a global perspective must be capable of simulating growth and development sensitivities, including associated temperature and drought stresses, to varying C, T, and W levels across a broad range of genetic and management factors. Accurate representation of these climate factors may present an additional challenge as many crop models were not originally developed to respond to C (White *et al*., 2011).

Systematic protocols for crop model intercomparison studies were recently proposed by the Agricultural Model Intercomparison and Improvement Project (AgMIP) (Rosenzweig *et al*., 2013), including comparison of simulation results against common datasets using partially and fully calibrated crop models. Common themes were reported among these studies. In a comparison of 27 models participating in the AgMIP Wheat Pilot, Asseng *et al*. (2013) observed major differences due to crop model structure and parameter values when comparing simulated vs. experimental yields and assessing sensitivity to climate change across multiple production sites. Both partially and fully calibrated models were able to reproduce observed experimental data, but uncertainty was reduced after full calibration. The median and mean from an ensemble approach of these models were consistently more accurate at replicating observed yields across all evaluated sites than any single model (Martre *et al*., 2015). Both error and variation decreased with increasing number of models in an ensemble, although no further declines were observed after at least ten models were randomly included. Uncertainty among models with respect to climate change factors varied according to site and crop management and increased with higher T. Strikingly similar results were observed with other cereal crops by the AgMIP Maize Pilot (Bassu *et al*., 2014) and the AgMIP Rice Pilot (Li *et al*., 2013) except that both pilots measured increased variability among

Table 1 Potato crop models participating in the intercomparison pilot and methodologies used to simulate soil, plant, and atmospheric components

Model	Leaf area/ light interception †	Light utilization ‡	Yield formation §	Crop phenology ¶	Root distri-bution ‖	Environ-mental stresses**	Water stress ††	Heat stress ‡‡	Water dynamics §§	ET ¶¶	Soil CN model ‖‖	CO_2 effects ***	# of cultivar parameters	Reference
AQUACROP V4.0	S	TE	HI	T	LIN	W,A, H,O	S	R	C,R O	P, PM, PT,TW, MAK, HAR, SW	-	TE	10	Steduto *et al.* (2009)
CSM-SUBSTOR-Potato V4.6*	S	RUE	Tn Prt	T,DL, O	CA, O	W,N, A	E	V, R	C	PM,PT	CN, P(3)	PT	5	Jones *et al.* (2003), Hoogenboom *et al.* (2015); Singh *et al.* (1998)
CROPSYST V3	S	TE RUE	HI B	T, DL	LIN	W,N	E		C, R	PM, PT	N, P(4)	TE RUE	>25	Stockle *et al.* (2003)
INFOCROP	S	RUE	B Prt	T,DL,O	LIN	W,N, A, H	E	V, R	C	PM, PT	CN	TE RUE	10	Aggarwal *et al.* (2006)
MONICA V1.01.0	S	P-R	Prt	T,DL, V,O	EXP	W,N, A,H	E, S	V	C	PM	CN, P(6), B	PT, F	15	Nendel *et al.* (2011)
SOLANUM	S	RUE	Prt	T	NR	W,H	E, S	V, R	C	PM	-	RUE	10	Condori *et al.* (2010)
	D	P-R	Prt	T,DL, O	CA,O	W,N, H	E, S	V, R	R			F, T, LF	14	

Table 1 (continued)

Model	Leaf area/ light interception †	Light utilization ‡	Yield formation §	Crop phenology ¶	Root distri- bution ‖	Environ- mental stresses**	Water stress ††	Heat stress ‡‡	Water dynamics §§	ET ¶¶	Soil CN model ‖‖	CO_2 effects ***	# of cultivar parameters	Reference
SPUDSIM v.2.0										PM, PT, O	CN, P(2)			Fleisher *et al.* (2010)
LINTUL V4	D	P-R	Prt B	T, DL	LIN	W,A N	S	V	C	P	-	-	6	Shibu *et al.* (2010)

*Two independent modeling groups used the SUBSTOR model.
†Leaf area development and light interception: S, simple approach (e.g., LAI); D, detailed approach (e.g., canopy layers).
‡Light utilization or biomass growth: RUE radiation use efficiency approach; P-R, gross photosynthesis – respiration; TE, transpiration efficiency biomass growth.
§Yield formation depending on: HI, fixed harvest index; B, total (aboveground) biomass; Tn, number of tuber or tuber growth rate; Prt, partitioning during reproductive stages.
¶Crop phenology is a function of: T, temperature; DL, photoperiod (day length); O, other water/nutrient stress effects considered.
‖Root distribution over depth: LIN, linear; EXP, exponential; SIG, sigmoidal; CA, carbon allocation; NR, no roots-just soil depth zone; O, other approaches.
**Stresses involved: W, water stress; N, nitrogen stress; A, aeration deficit stress; H, heat stress; O, other stresses.
††Type of water stress: E, actual to potential evapotranspiration ratio; S, soil available water in root zone.
‡‡Type of heat stress: Stress applied to V, vegetative organ (source); R, reproductive organ (sink).
§§Water dynamics: C, 'Tipping bucket' capacity approach; R, Richards approach.
¶¶Evapotranspiration: P, Penman; PM, Penman–Monteith; PT, Priestley-Taylor; TW, Turc–Wendling; MAK, Makkink; HAR, Hargreaves; SW, Shuttleworth and Wallace (resistive model).
‖‖Soil CN model: CN, CN model; N, N model; P(x), x number of organic matter pools; B, microbial biomass pool.
***CO_2 effects: RUE, radiation use efficiency; TE, transpiration efficiency; GY, grain yield; CLN, critical leaf N concentration; F, Farquhar model; T, stomatal conductance; PT, photosynthesis and transpiration; LF, Leaf-level photosynthesis–Rubisco (o) on QE and Amax.

models with increasing C as well as T. Marin *et al.* (2015) made similar recommendations with respect to sugarcane and climate change; however, only two models were compared. In non-AgMIP studies, Kabat *et al.* (1995) presented results from seven potato models calibrated against a common dataset and Wolf (2002) assessed differences between two potato models with respect to climate change sensitivity in Europe. In both studies, however, models were calibrated for a single production region, uncertainty was not quantified and, in the former case, did not include climate sensitivity. More rigorous multimodel intercomparisons have not been conducted for potato or other root crops.

Studies of crop model intercomparisons intended to elucidate commonalities in response to climate change can be summarized in terms of four types of analyses. Namely, those that (i) focus on in-depth exploration of simulated data, (ii) explore differences among models related to their structure and parameterization, (iii) quantify differences in uncertainty arising from climate predictions that serve as input data vs. that resulting from the models themselves, and (iv) evaluate vulnerability, adaptation, and resilience (White *et al.*, 2011; Rosenzweig *et al.*, 2013). The primary intent of this study was to address these first two components. Specifically, these aims were to evaluate the simulated data and uncertainties among an ensemble of crop models and, to a lesser extent, test for consistent differences in simulated values due to the manner in which models represent various soil–plant–atmospheric phenomena.

This study was thus conducted to quantify (i) the variation among a suite of peer-reviewed potato crop models with respect to simulating yield and other growth and development responses at different production sites, (ii) the effect of partial vs. full calibration on model uncertainty at these sites, and (iii) agronomic responses to climate change and associated variability in those predictions when using an ensemble model approach. We hypothesized that (i) the median response of an ensemble of potato models can be used to accurately represent yields at multiple locations, (ii) the median response will be more accurate than any individual model and, by extension, no individual model will consistently be better, or worse, than other contributing models, (iii) calibration will reduce uncertainty, and (iv) uncertainty will increase with the magnitude of shifts in C, T, and W away from baseline values. Results provide insight with regard to the utility of adopting an ensemble approach for potato modeling and identify sources of uncertainty and knowledge gaps within the models, especially in relation to model sensitivity to climatic factors.

Materials and methods

Crop models and datasets

Eight potato models, varying in representation of different soil–plant–atmosphere components, utilized by nine independent modeling groups were included in the AgMIP Potato Pilot (Table 1). These models were selected based on the willingness of the contributing group to conduct the necessary simulations and evidence of prior peer review. The SUBSTOR model was operated independently by two separate users in the pilot and thus represented as two separate models for statistical purposes.

Experimental datasets were obtained from sites within four distinct potato production regions (Table 2). These encompassed high-input (involving automated irrigation systems and chemical fertilizers as in Denmark and the United States) and low-input (rainfed production with organic amendments as in Bolivia and Burundi) management practices, different cultivars, and large variations in soil and climate. The combined genetic, environment, and management differences represented by the four diverse sites provided an extremely broad set of input variables for testing model variation. Each of the four datasets comprised a single growing season within each location at a specific field and included time-series measurements of plant growth and soil conditions throughout the growing season. Measured data included phenology dates; dry mass for tuber, leaf/stem or haulm, and root; and soil nitrogen and water content for some of the sites. These datasets were thus classified as 'gold-rated' sentinel sites for the purpose of crop model intercomparison studies (Boote *et al.*, 2015). Measured evapotranspiration data were not available, but nitrogen uptake was estimated from plant tissue nitrogen concentrations for the crops in the United States and Denmark. Experimental data were obtained via direct communication with the experimentalists or from peer-reviewed publications (Table 2). Daily climatic data were provided from a combination of observed site-specific values and estimated variables for surface wind speed, air humidity (dew point temperature, vapor pressure deficit, and relative humidity at the time of the day of maximum temperatures), maximum and minimum air temperature, solar radiation, and rainfall (Ruane *et al.*, 2015b). Estimated data were derived from the agricultural modeling version of the Modern Era Retrospective Analysis for Research and Applications (AgMERRA) climate forcing dataset (Rienecker *et al.*, 2011; Ruane *et al.*, 2015a) to fill in gaps in the observed station records. The result was a merged dataset covering the 1980–2009 planting years, providing at least 30 full growing seasons of bias-corrected estimates to obtain interannual variability estimates for each site in addition to the single, or baseline, experimental year of historical weather.

Simulation protocols

Simulations were conducted in two phases, partial (P1) or full (P2) calibration. In P1, modelers were provided with the minimum information required to run the models for each site.

Table 2 Field experiment locations, soil classification, crop management, and typical climate values for the four site locations used in the pilot

Experiments	Bolivia	Burundi	Denmark	United States
Location	Chinoli	Gisozi	Jyndevad	Washington
Latitude	−19.63	−3.57	54.90	46.22
Longitude	−65.37	29.68	9.13	−119.32
Altitude (masl)	3450	2091	15	520
Environment	Low-yielding rainfed	Medium-yielding rainfed	Medium-yielding irrigated	High-yielding irrigated
Reference	None	Harahagazwe *et al.* (2011)	None	None
Soils				
Soil type (FAO)	Leptosols	Histosols	Podzols	Arenosols
Soil type (USDA)	Sandy loam	Loam	Sand	Sand
Lower limit*	0.075	0.179	0.051	0.061
Drained upper limit*	0.185	0.339	0.151	0.160
Bulk density (g cm^{-3})*	1.58	1.00	1.49	1.33
Rooting depth (cm)	50	100	150	120
Crop management				
Variety (*S.t.* ssp *tuberosum*)	Desiree	Victoria (CIP381381.20)	Kaptah	Ranger Russet
Residue type	Barley	None	None	None
Plant density per ha	47619	41667	40404	50700
Row spacing (cm)	70	80	75	86
Plant spacing (cm)	30	30	33	23
Planting date (DOY†)	301	193	107	77
Manure applied (t/ha)	5	20	0	0
Inorganic applied fertilizer (kg N/ha)‡	80	100	180	366
Total irrigation (mm)	0	0	104	816
Phenology				
Emergence (DOY)	341	224	136	111
Harvest date (DOY)	91	339	267	235
Season duration (days)	156	154	161	159
Experimental year§	1997/1998	2007	1990	2004
Mean temperature (°C)	15.96	16.45	13.81	14.53
Cumulated precipitation (mm)	339.8	529.0	470.8	164.4

*Characteristics of uppermost soil layer.
†DOY, day of year.
‡Combined organic and inorganic.
§Obtained over the growing season from field data reported by experimentalist.

The dataset included daily climate data, initial soils conditions, management applications (e.g., planting dates, densities, irrigation, and/or fertilizer applications), dates of emergence and harvest, and cultivar name, maturity class, and photoperiod sensitivity. No information was provided regarding experimentally observed yields or time-series data, and thus, modelers were forced to estimate model calibration factors based on limited knowledge. In P2, modelers were provided with all P1 information plus experimentally obtained yields and time-series data. Interpretation, preparation, and development of input data for each model were conducted at the discretion of each modeling group according to agreed-upon guidelines for the pilot. The majority of calibration approaches were based on adjusting values for cultivar specific parameters by minimizing least-squared differences between observed and simulated yields. Some modeling groups reported difficulties in matching real-world data, particularly for soil water contents, and used additional approaches as noted in Table 3.

Simulation runs for baseline-year and multiple years were conducted for each phase at each sentinel site using baseline 'historical' weather data. In addition to the historical weather series, during P2, the 30-year weather data were modified on a daily basis to include variations in C, T, and/or W climatic factors to test model sensitivities to climatic factors. These consisted of five different levels of C (360, 450, 540, 630, and 720 ppm), five levels of T (−3, 0, +3, +6, or + 9 °C below/above baseline 24-h maximum/minimum air temperatures), and three levels of

Table 3 Summary of calibration approaches used for each modeling group for full calibration phase 2

Model	General approach	Data used*	Comments from group
AQUACROP	Manual	– Phenology – Biomass – Canopy coverage	• Quality/availability of data posed challenges especially for low-input sites without knowledge of nutrient/water stress • Adjusted WUE coefficient to obtain high yields for US site
CROPSYST	Manual	– Phenology – Biomass	• Calibration approach varied with each site based on availability of phenology, biomass, and nitrogen uptake data.
INFOCROP	Manual	– Phenology, – Biomass – LAI (for United States)	• Calibration conducted for phenology first
LINTUL	Manual	– Phenology – Biomass – Soil water holding properties	• Calibration conducted in three steps: – phenology (dates) – adjustment of RUE parameter (biomass) – adjustment of soil parameters that influence plant available water
MODSUBSTOR	Iterative approach using uncertainty estimator	– Biomass	• Low-input sites problematic as calibration generally requires nonstressed data
MONICA	Least-square iterative procedure	– Phenology – Biomass	• Unable to reach high yields from the United States
SOLANUM	Manual	– Canopy cover – Biomass	• Use of biomass component varied with availability of information at each site • Calibration for Burundi was more difficult than other locations
SPUDSIM	Least-square iterative procedure	– Phenology – Biomass – Canopy cover	• Low-input sites challenging to calibrate without knowledge of nutrient/water stresses
SUBSTOR	Least-square iterative procedure	– Phenology – Biomass	• None

*Biomass included total, leaf and stem, and/or yield dry and fresh mass.

W (−30%, 0, and +30% decreases/increases in daily precipitation from baseline). These levels were chosen as per Rosenzweig *et al.* (2013) to represent the range of potential 21st-century climate changes and to enable cross-crop comparisons with other model intercomparison studies (e.g., Mcdermid *et al.*, 2015). A full 5 × 5 × 3 factorial combination of these modified 30-year weather datasets was conducted at each site (i.e. 5 × 5 × 3 × 30 = 2250 simulation runs per model per site). For each year, initial soil conditions were reset to the original levels of water, organic matter, and nitrogen contents as per the corresponding field experiment. For irrigated locations in Denmark and the United States, modelers implemented automatic irrigation so as to maintain the soil water content within the top 50 cm at 90% field capacity or higher.

Data analysis

Simulated responses including end-of-season tuber yield and total biomass (kg dry mass ha^{-1}), harvest index (HI, yield divided by total mass minus roots), cumulative evapotranspiration (ET, mm), water use efficiency (WUE, yield basis, kg m^{-3}), nitrogen uptake (N_{up}, kg N ha^{-1}), and dates of tuber initiation (TI), 50% maximum tuber bulking (T50), and maturity (MAT) were evaluated. Relative percent error between simulated and observed values was computed across all models as in Eqn (1), where R_i represents corresponding response (e.g., tuber yield, ET, and etc.), $S_{i,m}$ is the simulated value of response i from model m, O_i is the observed experimental value for response i, and n is the total number of models that simulated a value for R_i:

$$R_i = \frac{\sum_{m=1}^{n} |(S_{i,m} - O_i)|/O_i}{n} \times 100 \quad (1)$$

Three methods were used to compare ensemble model performance: (i) analysis of variance, (ii) box–whisker plots, and (iii) coefficients of variation. Analysis of variance was conducted separately for calibration phase, simulation run, and response variable within each site using the GLM procedure in SAS software (SAS for Windows 9.4, SAS Institute, Inc., Cary, NC, USA). Main factors were site (n = 4) and model group (n = 9) for the baseline-year runs, plus climate factors C (n = 5), T (n = 5), and/or W (n = 3) and weather-years (n = 30) for the multiyear runs. These factors were considered fixed effects. Weather-years were nested within site for all multiyear runs. Interactions among C, T, W, and site were tested. Mean separations for significant main factors were conducted using the Sidak method (Sidak (1967), α = 0.05). Effects of model structure, grouped into three categories (RUE, TE, or P-R) according to their representation of light utilization (Table 1), in response of simulated tuber yield and ET to C, T, or W within each site were also tested. As this resulted in an unbalanced design (i.e., four RUE based models, two TE, and three P-R), type II sums of squares were used to evaluate for significance of the model structure term (Langsrud, 2003). Box and whisker plots, with boxes delimiting 25 and 75th percentiles and whiskers extending to 10th and 90th, were used to evaluate the distribution among model responses for each response and site with respect to simulated mean and median. Coefficient of variation (CV, %) was calculated across simulated model responses within each site. Results were initially averaged for each model across the 30 years for multiyear runs. Thus, reported CVs indicate variability among the model means and not that due to interannual responses within a given model.

The relationship between the uncertainty of predicted tuber yield and an ensemble, or grouping, of models (ranging from two to nine) used to simulate the response was explored separately P1 and P2 at each site. A random selection of 90 potential model combinations was obtained among all possible groups of models as determined from Eqn (2), following a similar approach used in other crop model intercomparison pilots (Asseng *et al.*, 2013; Bassu *et al.*, 2014; Li *et al.*, 2015):

$$C_{m=2\ldots9} = \frac{9!}{m!(9-m)!} \tag{2}$$

where m is the number of models in an ensemble and C_m is the potential number of combinations. CV was calculated for each group of models and used to determine the number needed to reduce uncertainty below a minimum of 17% as representative of observed potato yields under typical field conditions (Vermeer, 1990).

Results

Baseline-year responses

Predicted end-of-season tuber dry yield and total dry mass were not significantly different among models (Table 4, baseline-year responses), but phenology, nitrogen, and water use values varied ($P < 0.05$ or less). Results were affected by site and the distribution of simulated values within each location provided additional insight into ensemble model performance (Fig. 1). For example, median tuber yield responses for phase P1 were overpredicted for low-input sites of Bolivia and Burundi and underpredicted for the high-input sites in Denmark and USA. Only 16.7% of the model yield predictions, when aggregated across all four sites, fell within the 17% CV threshold of expected uncertainty for experimentally determined yield values. CVs ranged between 30% and 50% across sites for tuber dry yield (Table 5). Median values closely approximated observed yields after calibration (Fig. 1), and an average 80.6% of models, or seven of nine across the sites, fell within the threshold CV. With the exception of Bolivia, CVs for simulated values in this P2 phase were <14% (Table 5).

Relative errors between observed and simulated values declined for the majority of simulated responses from P1 to P2 (Fig. 2). These included nearly threefold reductions in errors for tuber dry yield and total biomass. Harvest index errors decreased to a lesser extent

Table 4 Analysis of variance summary* for partial (P1) and full (P2) calibration phases within baseline-year and multiyear runs. Simulated responses included tuber and total dry mass, harvest index (HI), seasonal evapotranspiration (ET), nitrogen uptake (N_{up}), water use efficiency (WUE), and dates of tuber initiation (TI), 50% maximum tuber bulking (T50), and tuber maturity (MAT). Factors included differences between 4 sites (S), 9 model (M) levels except for N_{up} (n = 7), TI and MAT (n = 7), and T50 (n = 4), and 30 weather-years (Y) for multiple year runs at each site

	P1									P2								
Factors	Tuber	Total	HI	N_{up}	ET	WUE	TI	T50	MAT	Tuber	Total	HI	N_{up}	ET	WUE	TI	T50	MAT
Baseline-year																		
S	***	***	ns	**	***	ns	*	***	***	***	***	*	***	***	***	***	*	***
M	ns	ns	**	*	*	ns	ns	***	***	ns	ns	ns	*	**	**	*	ns	**
Multiple years																		
S	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***
M	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***
Y (S)†	***	***	***	***	***	***	***	***	*	***	***	*	***	***	***	ns	***	***
M x S	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***

*Significance: ***<0.001, **<0.01, *<0.05, and ns>0.05.
†Weather-years were nested within site.

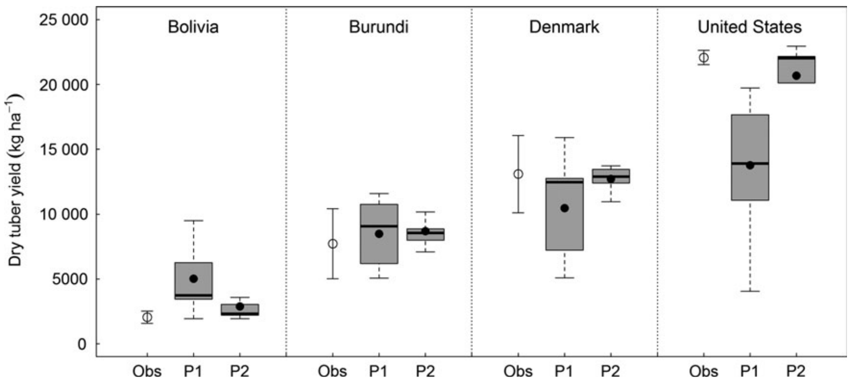

Fig. 1 Variation in simulated end of season tuber yield dry mass among nine potato crop models at four sites corresponding to year of experiment. P1 and P2 calibration phases were shown along with observed (Obs) tuber yield values and standard deviations. Observed (open circles) and simulated (filled circles) means and median values (solid line) as indicated.

Table 5 Coefficients of variation (%) for averaged model responses per site for calibration phases P1 and P2 and baseline-year vs. multiple year model runs. Abbreviations were defined in Table 4

Simulated response	Bolivia	Burundi	Denmark	United States	Bolivia	Burundi	Denmark	United States
	P1 baseline-year				P2 baseline-year			
Tuber	50.0	30.3	36.3	35.2	46.2	13.0	7.5	12.4
Total	50.7	27.5	24.0	33.9	40.9	8.8	9.7	17.4
HI	12.7	14.6	15.9	13.6	16.3	10.4	8.5	6.6
N_{up}	59.5	22.0	40.4	66.8	56.6	29.5	27.9	39.4
ET	25.7	25.3	24.1	19.7	52.3	30.9	38.1	20.9
WUE	49.7	29.5	38.7	32.3	48.8	63.1	33.6	21.2
TI	64.2	7.8	18.1	20.0	2.8	8.4	1.6	11.7
T50	66.9	6.4	7.7	7.8	122.2	6.3	1.7	5.9
MAT	7.3	2.9	2.6	2.9	23.7	4.7	8.5	7.5
	P1 multiple years				P2 multiple years			
Tuber	58.3	37.7	31.7	31.8	53.6	29.1	18.8	13.3
Total	54.7	38.3	21.2	32.6	39.6	26.3	16.7	18.5
HI	20.0	9.1	13.7	12.4	25.1	13.2	9.6	8.2
N_{up}	64.5	47.2	33.4	58.7	49.1	45.8	35.0	42.5
ET	29.0	28.1	26.0	33.5	43.3	38.5	44.3	23.4
WUE	45.2	35.4	31.7	33.0	51.6	77.4	40.0	21.1
TI	59.9	7.1	13.5	13.9	19.9	6.0	4.9	10.3
T50	64.8	8.6	9.0	8.7	70.0	10.1	1.3	11.2
MAT	21.2	22.5	4.0	5.4	26.0	5.0	7.2	9.6

than yield and total dry mass. The error for maximum canopy coverage, which served as an indirect measure of leaf area index, slightly declined and nitrogen uptake error was <20% following full calibration. However, errors for TI and T50 increased after full calibration.

CVs for resource responses (i.e., N_{up}, ET, and WUE) were generally larger than the dry matter values, particularly after full calibration (Table 5). These measures of uncertainty decreased for N_{up} for Denmark and the United States, but slightly increased for Burundi, and the uncertainty for ET predictions increased across all four sites (Table 5). Phenology responses were not substantially influenced by calibration, although variation was apparent among models after calibration for estimates of the T50 value for Bolivia, a result attributed to one outlying model (not shown). There were still significant differences among model simulations with respect to maturity date following full calibration (Table 4). On the whole, calibration appeared to increase, or not affect, the variance among models for resource use responses while decreasing that for dry matter and phenology, and this result was emphasized for low-input managed sites of Bolivia and Burundi.

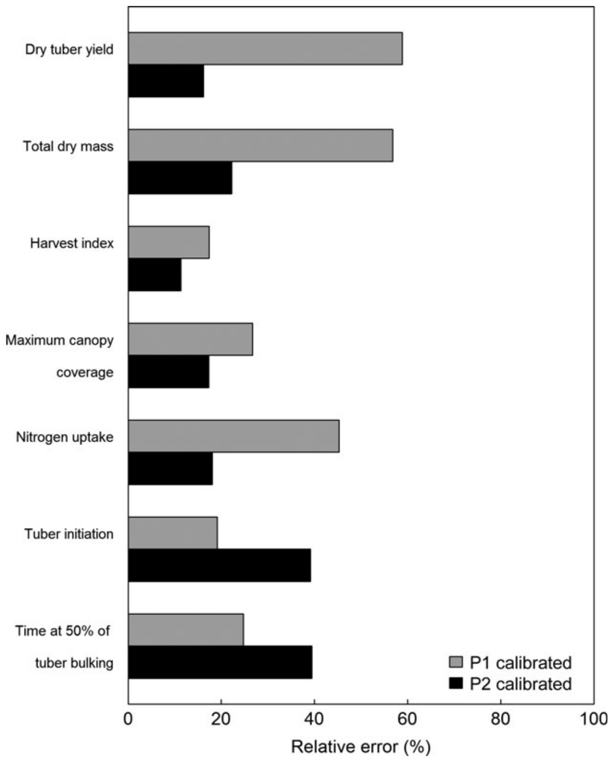

Fig. 2 Relative percent error between the simulated response (averaged across potato crop models) vs. the observed value aggregated across all four experimental sites. Grey and black bars were reported for partial P1 and full P2 calibration phases, respectively. Nine models were included in all responses except for maximum canopy coverage (n = 5), nitrogen uptake (n = 7), and tuber initiation and tuber bulking dates (n = 7). Nitrogen uptake was only reported experimentally for Denmark and United States.

Multiyear responses with baseline weather and sensitivity to climatic factors

Main factors and interactions between model and site under baseline multiyear runs were significant ($P < 0.05$ or less) for all responses, including dry tuber yield and total biomass, at both calibration phases (Table 4). Significance of the model and site interaction indicated that the manner in which model simulations varied with each other depended on geographic location. The distributions of model predictions, as reflected by CV (Table 5), were broader than baseline-year results but exhibited similar trends. CVs were generally lower and declined after calibration, for dry matter vs. resource responses, with phenology unaffected except for the Bolivian site. Variation among model predictions remained higher for low- vs. high-input sites.

Multiyear runs conducted for P2 also included factorial combinations of climatic factors C, T, and W. Interactions between model and each of these factors were significant ($P < 0.01$) for all responses and sites, indicating the difference between model simulations varied according to the levels of C, T, and/or W (ANOVA results not shown). Phenology, including tuber initiation and maturity dates, was influenced by W and T, but not C. Responses to C generally had lower CVs as compared with T or W (Table 6) and this measure of uncertainty was consistent throughout all C levels at each site location (i.e., CV did not substantially increase with increasing C level). Variation among model simulations was again higher for resource use vs. dry matter responses, particularly for low-input (Bolivia and Burundi) vs. high-input (Denmark and United States) sites (Table 6). Nearly, all models simulated an increase in tuber dry yield in response to rising C (Fig. 3), the exception being one model showing a lack of response for Bolivia and Burundi. Thus, median response to increasing C levels at each site was positive for simulated dry yield, with increases between baseline (360 ppm) and maximum (720 ppm) C levels of 12–41% (Fig. 3, Table 7). The highest relative yield increases were observed at the lower-input sites of Bolivia and Burundi. Median responses for total dry mass followed similar trends, the result being little change in harvest index among the models (Table 7). Nitrogen uptake increased in proportion to dry matter with a range of 0.3–21%, ET declined slightly between 1% and 6%, and thus, WUE also increased between 11% and 37%.

Variation among models to changing T was typically higher than C or W for all simulated responses (Table 6). Unlike C responses, CVs increased within a given site as T increased or decreased at the −3 °C level, from the baseline value. Dates of tuber initiation and tuber maturity exhibited strong sensitivity to T and large variation among models (Fig. 4). Tuber initiation occurred between 12 and 18 days earlier for United States and Denmark as T increased from baseline, but was delayed for all sites at the −3 °C level. CVs were generally greater at the coldest and warmest levels of T (Fig. 4). Tuber maturity exhibited similar patterns, with warmer temperatures resulting in substantially earlier maturity dates for all four sites. Most models indicated declining tuber dry yields with warming T, but three of nine models showed an opposite trend for Bolivia (not shown). Median dry matter responses from all nine models indicated warming temperatures were detrimental to potato tuber dry yields, with losses ranging from 29% to 45% with respect to baseline air temperatures (Fig. 5). Similar trends were observed for total mass (Table 7), but harvest index declined from 5% to 9% depending on site. ET responses were more variable and increased for some sites with warming T (Table 7).

Table 6 Coefficients of variation (%) for single climate factor changes (atmospheric carbon dioxide (C), temperature (T), or rainfall (W)) for multiyear averaged model response across each site (Bo – Bolivia; Bu – Burundi; De – Denmark; US – United States) at full (P2) calibration. Simulated responses included tuber and total dry mass, harvest index (HI), nitrogen uptake (N_{up}), seasonal evapotranspiration (ET), and water use efficiency (WUE)

	C																							
	Tuber				Total				HI				N_{up}				ET				WUE			
C	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US
360	54	27	22	13	40	25	18	18	25	13	10	8	49	46	35	42	38	34	34	23	56	60	29	23
450	53	29	21	15	40	25	16	19	25	14	10	8	51	43	32	40	38	34	34	22	52	58	26	23
540	54	30	23	17	43	25	17	20	25	15	10	7	53	40	31	39	38	35	35	22	51	56	24	24
630	56	30	22	18	45	25	16	21	24	15	10	7	56	38	30	40	39	35	36	22	52	56	23	24
720	57	31	25	18	48	25	18	21	24	16	10	7	58	36	30	38	39	35	37	22	52	55	23	25
	T																							
T	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US
−3	57	28	26	26	41	23	19	24	36	18	12	13	56	28	25	39	34	30	28	22	58	68	35	34
0	54	27	22	13	40	25	18	18	25	13	10	8	49	46	35	42	38	34	34	23	56	60	29	23
+3	53	37	30	16	40	37	26	22	20	13	9	7	49	67	51	50	43	41	43	28	57	51	29	23
+6	63	47	35	17	52	48	35	26	19	15	8	9	61	87	66	58	50	48	51	32	64	46	31	27
+9	80	68	41	25	66	68	46	31	22	21	12	17	68	109	82	68	57	54	57	36	80	63	41	41
	W																							
W	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US
−30	47	45	32	23	33	45	29	19	25	16	11	17	50	66	42	40	49	39	36	23	62	71	35	31
0	54	27	22	13	40	25	18	18	25	13	10	8	49	46	35	42	38	34	34	23	56	60	29	23
+30	65	16	18	13	52	13	13	18	25	12	9	8	59	40	33	41	32	31	33	22	55	50	27	23

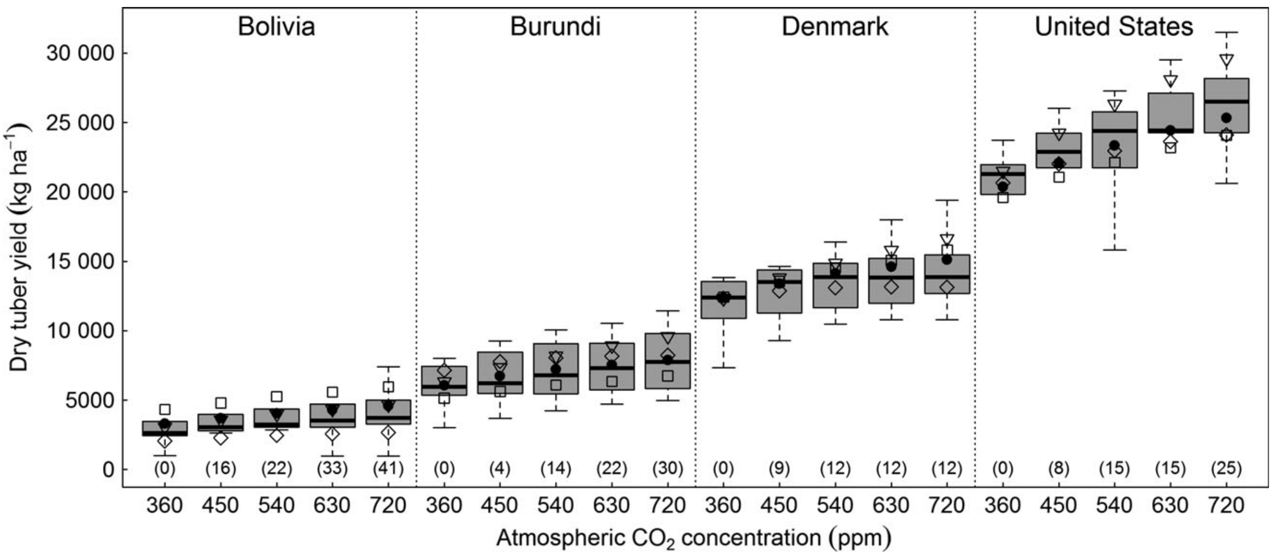

Fig. 3 Atmospheric CO_2 concentration (ppm) effect on tuber dry yield (kg ha^{-1}) simulated by nine models across 30 years at Bolivia, Burundi, Denmark and United States. Model structure mean was represented by open diamond for P-R approach, open square for RUE, and open triangle for the models with TE approach (details in Table 1). Values in parentheses indicate % median change from baseline value of 360 ppm. Additional symbols as described in Fig. 1.

Table 7 Median simulated response of nine models averaged over 30 years per site to varying levels of C, T, and /or W for end-of-season tuber yield, total dry mass, harvest index (HI), nitrogen uptake (N_{up}), evapotranspiration (ET), and water use efficiency (WUE)*

	C																							
	Tuber				Total				HI				N_{up}				ET				WUE			
C	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US
360	2.7	6.0	12.4	21.3	4.5	8.3	16.3	28.4	0.73	0.86	0.84	0.80	100	130	207	350	364	340	403	677	0.95	1.73	3.10	2.90
450	3.1	6.2	13.5	22.9	4.7	9.2	17.0	30.9	0.74	0.86	0.84	0.80	100	130	212	374	364	336	399	673	1.11	1.80	3.29	3.19
540	3.2	6.8	13.9	24.4	4.9	9.7	18.3	32.5	0.74	0.86	0.84	0.80	100	130	220	395	363	332	393	669	1.19	1.84	3.41	3.51
630	3.5	7.3	13.8	24.4	5.4	9.8	18.6	35.1	0.74	0.86	0.84	0.80	101	130	238	412	363	327	387	665	1.24	1.89	3.53	3.79
720	3.7	7.7	13.9	26.5	5.9	10.0	18.7	35.5	0.74	0.86	0.84	0.80	101	130	235	423	363	323	380	661	1.30	1.93	3.62	4.03
	T																							
T	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US
−3	2.9	6.8	12.8	21.6	4.6	8.6	16.0	27.4	0.66	0.87	0.82	0.79	96	131	215	393	379	379	360	633	0.77	1.47	2.96	2.88
0	2.7	6.0	12.4	21.3	4.5	8.3	16.3	28.4	0.73	0.86	0.84	0.80	100	130	207	350	364	340	403	677	0.95	1.73	3.10	2.90
+3	2.6	5.3	11.0	19.0	3.8	6.3	13.5	24.7	0.73	0.86	0.87	0.80	67	113	178	302	362	317	345	665	0.81	1.58	3.15	2.84
+6	2.2	3.7	9.6	15.4	3.2	5.3	13.5	20.5	0.75	0.86	0.84	0.79	57	91	139	259	347	294	362	727	0.76	1.19	2.78	2.55
+9	1.7	2.7	8.8	11.7	2.5	3.5	11.2	16.3	0.73	0.80	0.76	0.77	76	70	116	221	331	309	399	713	0.55	0.76	1.88	2.02
	W																							
W	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US	Bo	Bu	De	US
−30	2.3	5.8	12.2	21.3	3.6	7.6	16.6	26.6	0.69	0.86	0.84	0.80	52	83	191	417	285	313	372	674	0.81	1.59	3.13	2.91
0	2.7	6.0	12.4	21.3	4.5	8.3	16.3	28.4	0.73	0.86	0.84	0.80	100	130	207	350	364	340	403	677	0.95	1.73	3.10	2.90
+30	2.9	6.8	12.2	21.3	4.5	8.7	15.6	28.4	0.74	0.86	0.85	0.80	102	130	222	350	424	342	411	680	0.85	1.77	2.99	2.91

*Dry mass units in Mg ha^{-1}, nitrogen uptake in kg ha^{-1}, transpiration in mm, and water use efficiency in g l^{-1}.

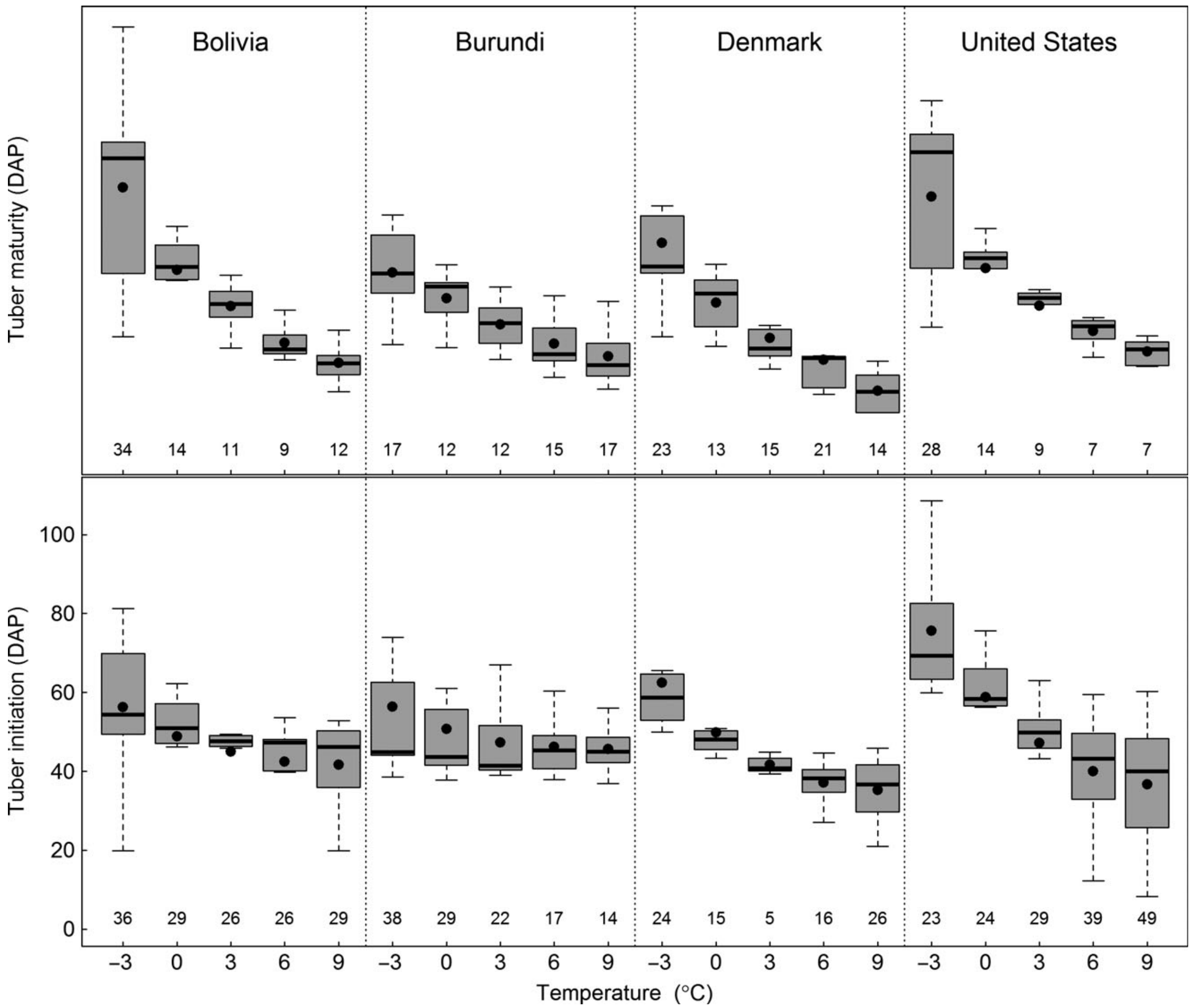

Fig. 4 Temperature effect on tuber initiation and tuber maturity at Bolivia, Burundi, Denmark and United States based on 30-year simulations with seven models. Values below box-whisker plots are CV (%) across the models within each temperature combination. Mean baseline temperatures (where T = 0) during the growing cycle were 15.96, 16.45, 13.81 and 14.53 °C in Bolivia, Burundi, Denmark and United States, respectively. Additional symbols as in Fig. 1.

As a result of dry matter and ET patterns, WUE declined with an across site range between 30% and 56%. N_{up} declines were similar to dry mass.

There was little influence of rainfall on simulated dry matter production for Denmark and the United States, as both implemented automatic irrigation as part of the management practice (Table 7), but low-input sites showed tuber yield increases as large as 26% in response to increasing rainfall (Fig. 6). As with dry matter responses, water use was only slightly impacted for high-input sites, but increased for low-input sites. N_{up} also averaged a 46% increase for both low-input sites (Table 7). With the exception of Bolivia, CVs decreased as rainfall increased, indicating there was less variation among models under less stressed production conditions (Table 6).

Multiyear tuber dry yield and ET responses to climate change factor interactions

Interactive effects between model and combinations of either C × T, C × W, or T × W were significant ($P < 0.01$ or less) at each site (not shown). Contour plots for relative changes from baseline values for tuber dry yield and ET simulated at historical C, T, and W levels were aggregated according to low-input (Burundi and Bolivia) and high-input (Denmark and USA) managed sites. These show a relative dry tuber yield increase at

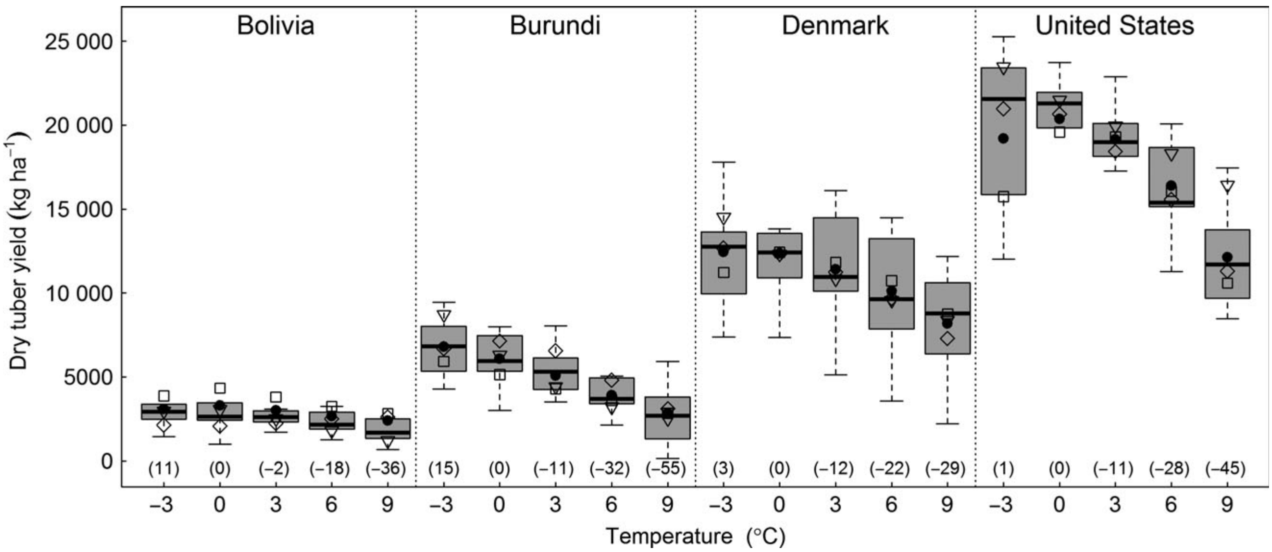

Fig. 5 Temperature effect on tuber dry yield (kg ha^{-1}) based on 30-year simulations with nine models at Bolivia, Burundi, Denmark and United States. Model structure mean was represented by open diamond for P-R approach, open square for RUE, and open triangle for the models with TE approach (details on Table 1). Values in parentheses indicate % median change from baseline value of +0 °C. Additional symbols as in Fig. 1.

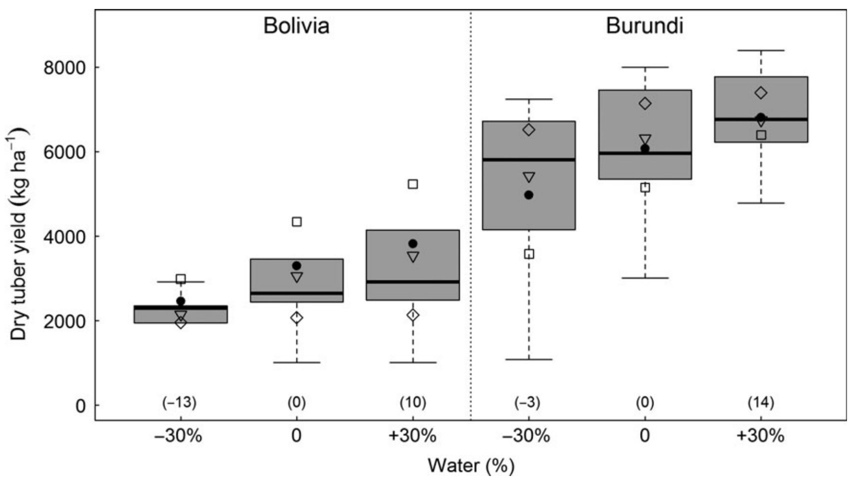

Fig. 6 Precipitation effect on tuber dry yield (kg ha^{-1}) based on 30-year simulations with nine models at Bolivia, Burundi. Model structure mean was represented by open diamond for P-R approach, open square for RUE, and open triangle for the models with TE approach (details on Table 1). Values in parentheses indicate % median change from baseline value of 0%. Additional symbols as in Fig. 1.

higher C and lower T and a decrease with warmer T at all C levels for all sites (Fig. 7). The largest proportional increase in yields occurred at the baseline T and higher C levels for high-input sites, but at a broader T range for the lower-input sites. Rainfall had little influence on the relationships between C and T, but yields increased for low-input sites with higher amounts (Fig. 7). The changes in tuber dry yield in response to warming T (Mg °C^{-1}) were also compiled within selected C x W combinations at each site (Table 8). In contrast to the contour plots, these values were based on changes in response from 0 to 9 °C increases within a given C level (i.e., not referenced to historical baseline values at 360 ppm). Rates were always negative and were at the largest absolute values at the highest C levels, ranging from −0.14, −0.44, −0.51, and −1.04 Mg tuber dry yield °C^{-1} for Bolivia, Burundi, Denmark, and the United States, respectively, at a C level of 720 and baseline W. This indicated greater sensitivity to warming T at each site as C increased. As expected, Denmark and United

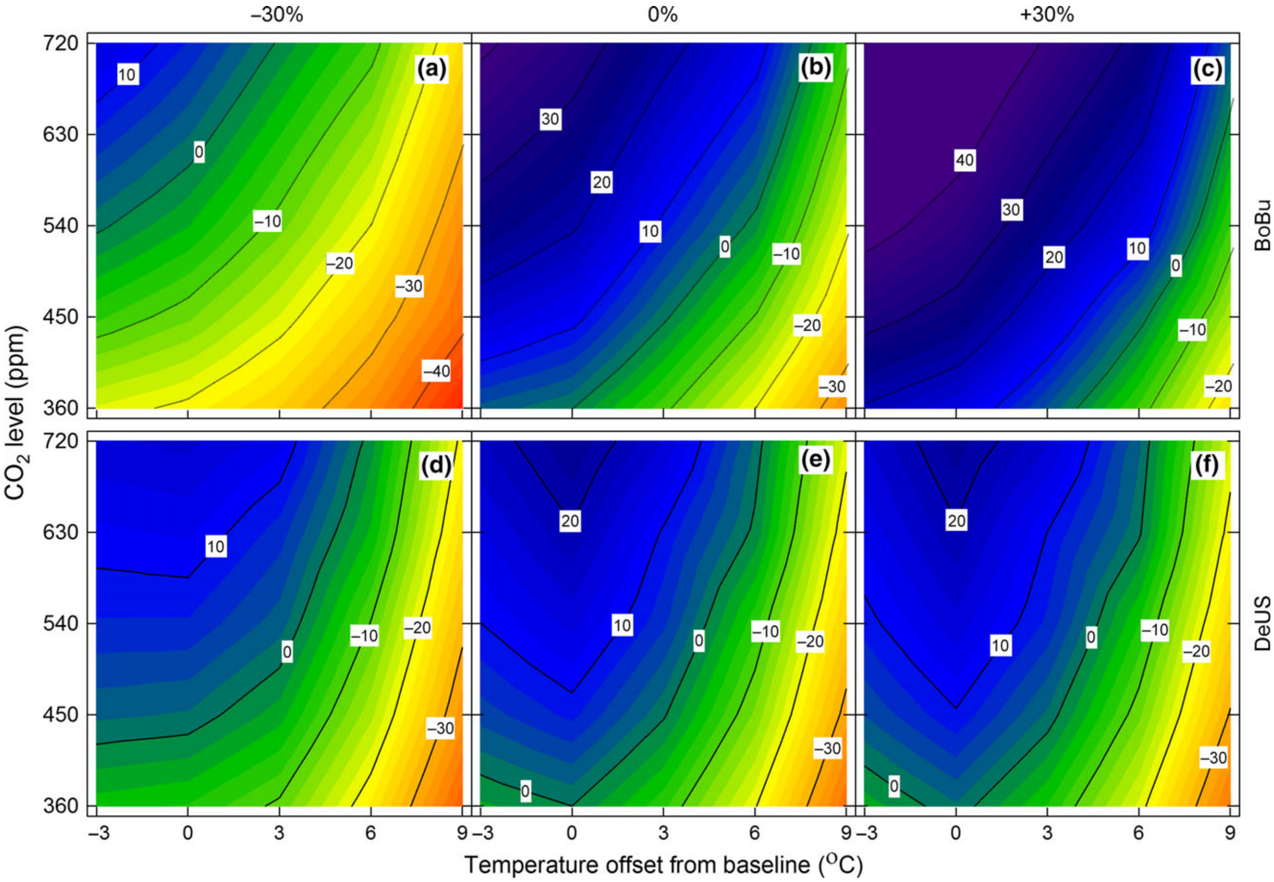

Fig. 7 Relative changes (%) in tuber dry yield from baseline C, T, and W values based on 30-year simulations averaged over nine potato crop models with respect to varying C and T factors. Graphs (a), (b), and (c) represent values averaged across low (Bolivia-Burundi, BoBu) sites and (d), (e), (f) across high (Denmark-United States, DeUS) sites. Graphs (a), (d) indicate −30%, (b), (e) 0%, and (c), (f) +30% decreases/increases in daily rainfall amounts from baseline values.

States showed little sensitivity to rainfall as these sites were irrigated. However, yields were more sensitive to increasing T within each C level when more water was available for the low-input sites.

CVs among models for absolute tuber dry yield (trends for relative yield were nearly identical) showed distinct patterns of uncertainty between the low-input tropical sites and high-input temperate ones (Fig. 8). Variation for low-input sites was roughly twice that of high-input ones at all three levels of W. High-input sites exhibited larger CVs in response to lowest and highest levels of T, but these variations were <6%. In contrast, CVs increased a minimum of 20% at all three W levels for low-input sites as T increased. Most variation among models was observed at the −30% rainfall amount, but little interaction with W and T or C was apparent. Uncertainty in response to C level appeared to predominantly interact with low T in high production sites (i.e., uncertainty increased at −3 °C with higher C level). The largest CVs generally coincided with the largest predicted decrease in yields (at +6 or +9 °C) (Figs 7 and 8).

ET responses exhibited somewhat different patterns than for dry tuber yield and varied between sites (Fig. 9). Relative ET showed the strongest response to T, declining as T increased for low-input sites. Elevated C levels had a larger influence on ET at cooler temperatures, and increasing rainfall resulted in higher ET at lower T and C. As expected, ET at high input, irrigated sites showed little influence due to increasing W; however, warming T, in contrast to low-input sites, slightly increased water use although the effect was moderated at higher C levels. At the tropical low-input sites, sensitivity of ET to increasing T within C levels was negative with larger declines observed at the lower C levels (Table 8). This sensitivity was positive, however, for Denmark and the United States, with values as high as 4.5 mm ET $°C^{-1}$ for Denmark at 360 C level. CVs among averaged model ET predictions were similar to those for yield, except there was little response to C

Table 8 Median simulated tuber dry yield and evapotranspiration (ET) changes in response to increasing T* from nine models averaged over 30 years per site for selected C x W combinations

C	W	Bolivia	Burundi	Denmark	United States
		Change in Tuber dry yield (Mg °C^{-1})			
360	−30	−0.07	−0.29	−0.43	−0.82
	0	−0.10	−0.36	−0.47	−0.92
	+30	−0.12	−0.39	−0.47	−0.91
540	−30	−0.09	−0.35	−0.48	−0.88
	0	−0.13	−0.42	−0.51	−0.99
	+30	−0.14	−0.45	−0.49	−0.97
720	−30	−0.11	−0.37	−0.49	−0.90
	0	−0.14	−0.44	−0.51	−1.03
	+30	−0.14	−0.48	−0.48	−1.02
		Change in ET (mm °C^{-1})			
360	−30	−3.28	−2.24	4.24	2.17
	0	−3.85	−2.07	4.45	2.31
	+30	−3.80	−1.67	4.39	2.44
540	−30	−3.17	−1.97	4.16	2.47
	0	−3.59	−1.76	4.20	2.61
	+30	−3.62	−1.29	4.30	2.74
720	−30	−3.09	−1.71	4.02	2.75
	0	−3.41	−1.39	4.04	2.86
	+30	−3.44	−0.96	4.17	2.98

*Changes based on changes in response variable from 0 to 9 °C within each C × W level.

(Fig. 10). CVs increased with lower rainfall amounts only for low-input sites.

Model structure relationship with variability and ensemble model predictions

Models were grouped into three categories (RUE, TE, or P-R) according to their representation of light utilization (Table 1) in order to assess the influence of model structure on simulated dry tuber yield and ET. Model structure and interactions with all climatic factors were significant for all sites ($P < 0.001$). The ranking of tuber yield or ET means based on order of magnitude was consistent across C, T, or W levels within the same site. For example, tuber yields from models based on RUE were consistently higher than P-R or TE models for Bolivia, P-R and TE models were generally higher than RUE for Burundi, and TE was larger than P-R, which in turn was higher than RUE for the United States for C (Fig. 3), T (Fig. 5), or W (Fig. 6) comparisons. Simulated responses for Denmark were closer to the median for all model structures except at higher C and lower T levels. In Burundi, the relationships were also affected by C level (Fig. 3). ET responses were consistently larger for RUE models, followed by TE and then P-R across all sites (Fig. 11).

The relationship between the number of potato crop models in a modeling ensemble and associated reduction in uncertainty was computed. Distinct differences were observed between P1 and P2 phases and low- and high-input sites (Fig. 12). Among possible model combinations, a minimum of two models was sufficient to achieve uncertainty below the 17% cutoff at either calibration phase for the high-input sites. For lower-input sites, five of the P1 models were recommended or three P2 models to get below this uncertainty threshold. Using nine fully calibrated models from P2 reduced tuber yield uncertainty to <10% in both low- and high-input sites.

Discussion

Uncertainty and model calibration

The median dry tuber yield from the ensemble of potato models was consistently closer to observed values than any individual model at either calibration phase. The ensemble of models also accurately replicated most observed production responses; however, only the fully calibrated models matched experimental tuber yields across all four locations (Figs 1 and 2). Responses from partially calibrated models (P1) were within two standard deviations of observed values for two of the four sites (Fig. 1). The two outlying locations comprised the lowest (Bolivia) and highest (United States) measured yields, but also had relatively less experimental error. Observed productivity for the United States exceeded typical regional irrigated yields (USDA-NASS, 2015) by roughly 40%, the outcome being that all partially calibrated models underestimated the yields. Observed yields for Bolivia were consistent with FAO statistics (FAO, 2015), yet only a single model fell within experimental standard deviations. Because cultivars differed among sites, insufficient data were available to ascertain what specific limitations in the modeling approaches could be attributed to the differences between partial and full calibrations. However, a majority of cultivated potato varieties vary with respect to degree of determinacy. Vegetative growth in an indeterminate variety can occur after the formation of yield or storage bearing organs (Manrique *et al.*, 1990). From a modeling perspective, indeterminacy influences carbon allocation between above- and belowground organs and, via persistence of leaf area, may lead to shorter or longer growth durations depending on crop or soil water and nutrient status, temperature, photoperiod, physiological age of seed, and other conditions. Lack of fully characterized agronomic data for the more extreme sites coupled with this phenotypic trait, therefore, may be partly responsible

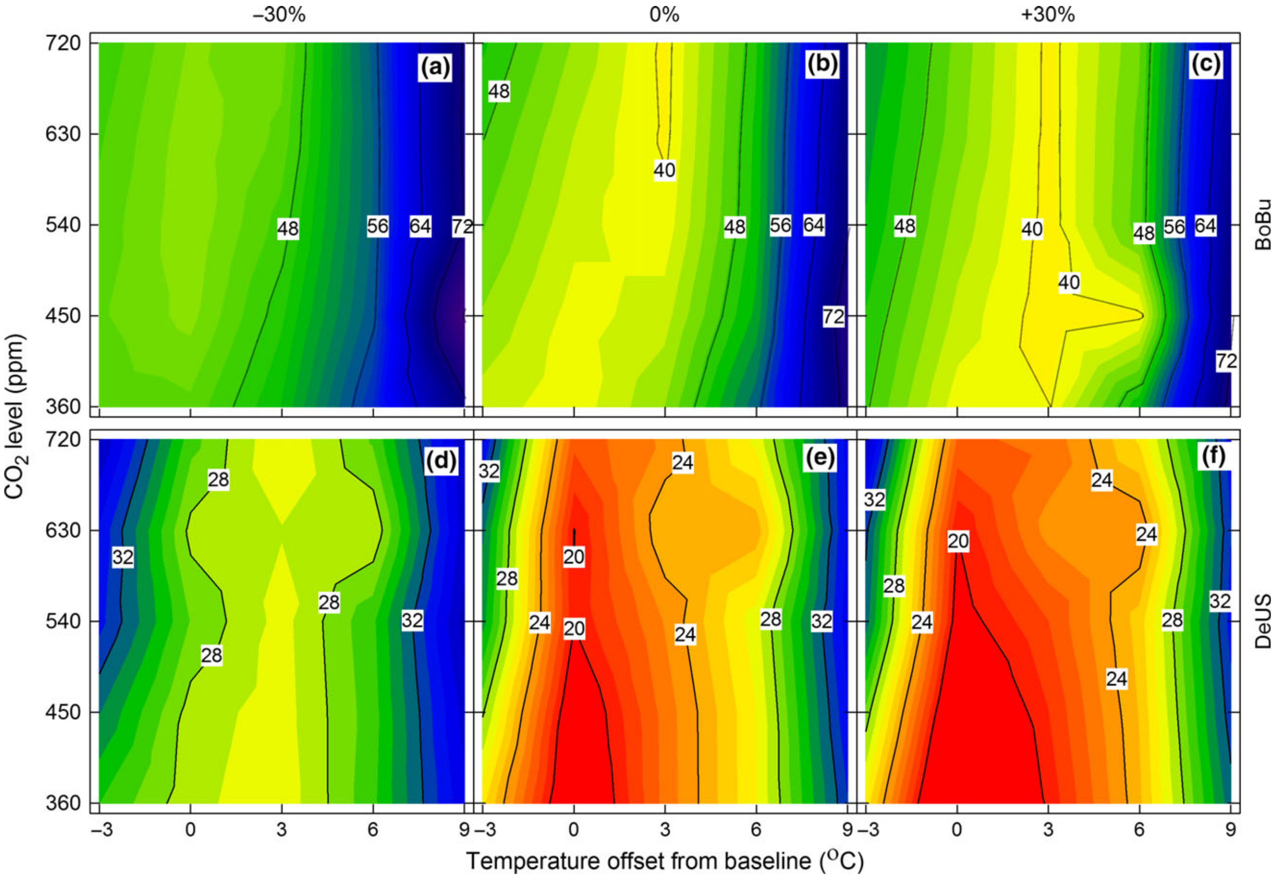

Fig. 8 CVs (%) of tuber dry yield based on 30-year simulations averaged over nine potato crop models with respect to varying C and T factors. Graphs (a), (b), and (c) represent values averaged across low (Bolivia-Burundi, BoBu) sites and (d), (e), (f) across high (Denmark-United States, DeUS) sites. Graphs (a), (d) indicate −30%, (b), (e) 0%, and (c), (f) +30% decreases/increases in daily rainfall amounts from baseline values.

for the current results. Once provided with full access to all experimental data, nearly all models accurately replicated the dry tuber yields at both sites (eight of nine models for Bolivia and seven of nine for the United States). Such results suggest that a minimum calibration dataset for potato should include dry matter data in addition to phenological observations for the more extreme production conditions, in contrast with other modeling pilots. This is an important finding to keep in mind when interpreting results from projected climate change impacts on potato production as complete sets of site calibration data are typically not available (Palosuo *et al.*, 2011). Either way, CVs among model yield predictions declined substantially going from partially to fully calibrated models (Table 5), and the majority of locations were associated with CV values below the field variation threshold of 17%. Thus, in contrast with other model intercomparison studies (Challinor *et al.*, 2009; Palosuo *et al.*, 2011; Rötter *et al.*, 2012; Asseng *et al.*, 2013; Bassu *et al.*, 2014; Li *et al.*, 2015), full model calibration not only reduced variation among the ensemble of contributing models, but also improved the median accuracy. Note that full calibration appeared to increase the uncertainty among models with respect to phenology data (Table 4), but this was largely attributed to challenges modeling groups had in replicating yield data for the Bolivian site (Table 3). As such, this result was reflective of the differences among potato models with respect to simulating developmental progress (Table 1) as well as difficulties in simulating low-yielding locations. In the latter case, calibration strategy may have focused more on matching simulated results to observed dry matter production as opposed to phenological events, creating the increase in variability for some developmental responses.

Variability among partially and fully calibrated models was higher for multiyear runs (Table 5) to the extent that models were significantly different for all reported responses (Table 4). Thus, models responded differently with respect to interannual weather variability even when calibrated using the same set of

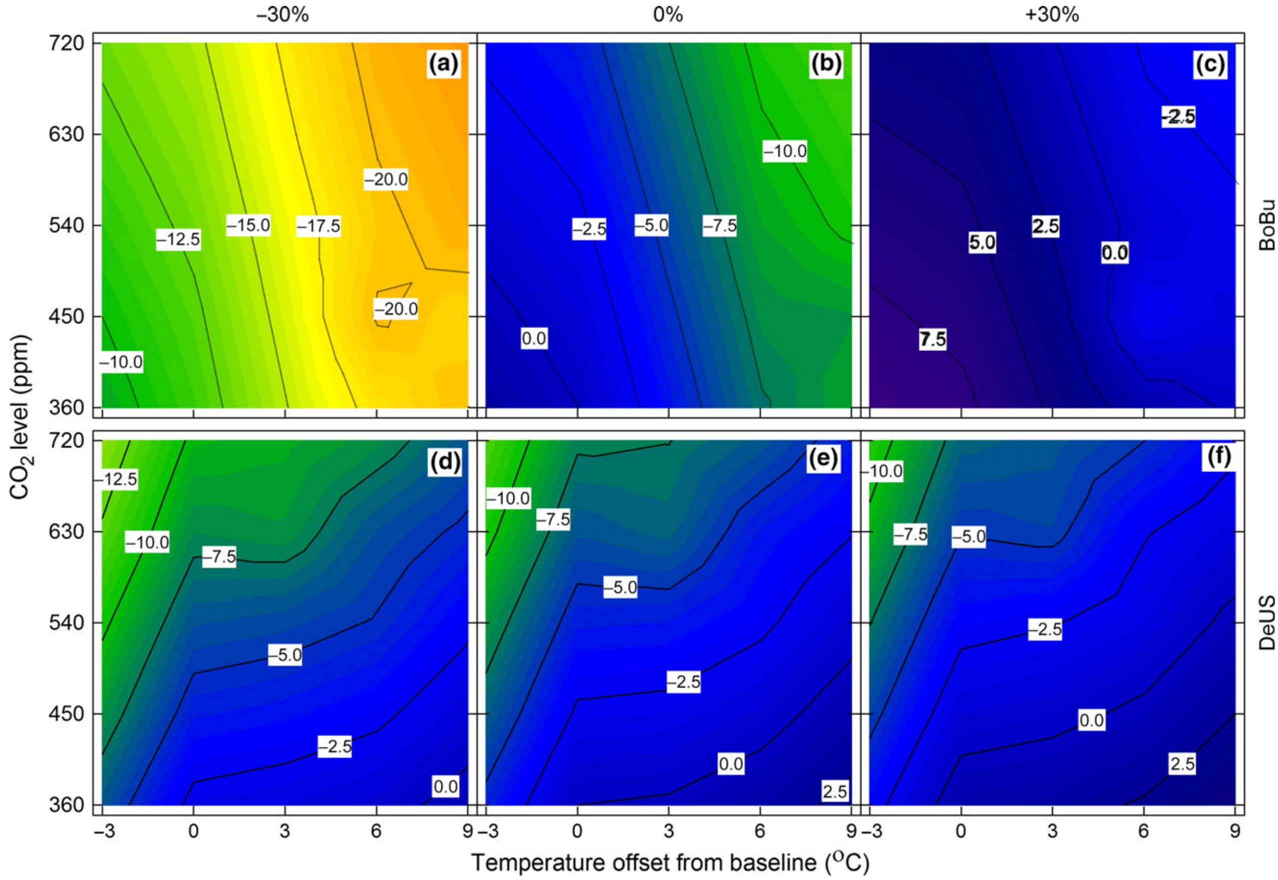

Fig. 9 Relative change in evapotranspiration (%) based on 30-year simulations averaged over nine potato crop models with respect to varying C and T factors. Graphs (a), (b), and (c) represent values averaged across low (Bolivia-Burundi, BoBu) sites and (d), (e), (f) across high (Denmark-United States, DeUS) sites. Graphs (a), (d) indicate −30%, (b), (e) 0%, and (c), (f) +30% decreases/increases in daily rainfall amounts from baseline values.

experimental data. Model uncertainty was consistently higher for resource use responses as compared with dry matter values (Table 5), and CVs actually increased for ET and WUE following full calibration. This was perhaps not surprising, considering that models were calibrated to dry matter data, but not to ET, WUE, or N_{up} data. For example, calibration strategies used by the majority of models primarily focused on using phenology data first and then adjustment of varietal coefficients to minimize differences between observed and simulated dry matter (Table 2). Time-series data for plant tissue nitrogen content was provided for United States and Denmark locations, but additional information regarding soil water and nitrogen contents were only available for the United States. Only one of the nine modeling groups utilized the nitrogen data, and a different group mentioned use of soil water information as part of their calibration process (Table 3). Therefore, it was unclear whether modelers would have explicitly included errors between simulated and observed water and/or nitrogen use as an additional part of their calibration strategies if such data were available. In this case, reasons for differences in model variation between resource and dry matter responses were confounded between calibration/dataset limitations and the range of approaches used by the models with respect to water and nutrient uptake and utilization. Variations of the Penman–Monteith approach for ET and carbon–nitrogen ratios for nitrogen balance were used by most models (Table 1), but implementation of these approaches differed. Differences in simulation methods for soil nitrogen dynamics, nitrogen and water movement, and uptake were likely also responsible for a substantial portion of this variation, and the various methods used to describe root distribution would seem to reflect this point. In-depth characterization involving calibration with and without detailed resource data would provide more certainty to this question, but datasets containing the necessary detail were unavailable.

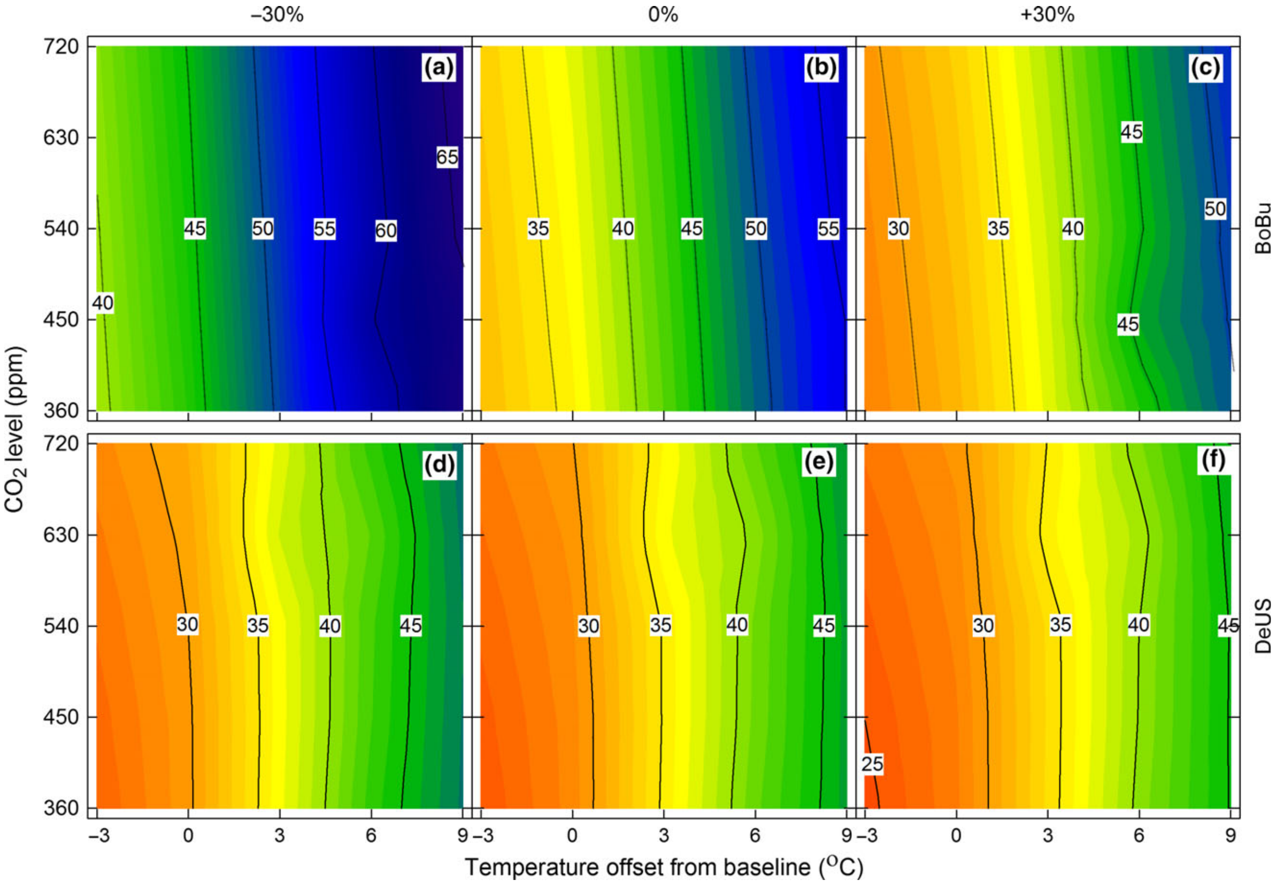

Fig. 10 CVs (%) of evapotranspiration based on 30-year simulations averaged over nine potato crop models with respect to varying C and T factors. Graphs (a), (b), and (c) represent values averaged across low (Bolivia-Burundi, BoBu) sites and (d), (e), (f) across high (Denmark-United States, DeUS) sites. Graphs (a), (d) indicate −30%, (b), (e) 0%, and (c), (f) +30% decreases/increases in daily rainfall amounts from baseline values.

The relationship between cross-model variation and the number of crop models in an ensemble provides an approximate quantification of the uncertainty associated with making future predictions (Asseng *et al.*, 2013). It was expected that fewer P2 models would be needed in the ensemble to meet the common field variation threshold as these showed closer agreement with one another for dry matter simulations. However, this difference in projected ensemble model numbers between P2 and P1 is likely to be minimal when forecasting resource responses instead of dry matter as CVs for ET, N_{up}, and WUE did not improve following full calibration (Table 5).

The number of crop models in an ensemble approach required to meet the common dry yield field variation threshold differed from other AgMIP model intercomparison pilots. A study with wheat ensembles found an average two to three fully calibrated models would be required under historical climatic conditions when averaged across all sites (Asseng *et al.*, 2013). Results for maize included strong site dependency with a range between three and 17 models depending on location (Bassu *et al.*, 2014). Li *et al.* (2015) reported a minimum of eight partially calibrated, or five fully calibrated, models were required across all sites for rice. The present results were closer to the wheat findings, with two or three fully calibrated models recommended for high- or low-input sites, respectively (Fig. 12). However, these other studies utilized a smaller threshold for the natural uncertainty level (13.5% for wheat and maize and 15% for rice) than the 17% used for potato. If a lower threshold of 13.5% were desired, roughly seven fully calibrated models would be required for Bolivia and Burundi sites, and this level of uncertainty could not be met for these low-input sites using partially calibrated models. An additional survey was conducted over a broad range of production and cultural conditions (Van Dam *et al.*, 1996; Miglietta *et al.*, 1998; Fangmeier *et al.*, 2002; Timlin *et al.*, 2006; Condori *et al.*, 2010; Harahagazwe, 2016, Personal communications) to determine whether the variation in potato tuber dry yields from field experiments differed between

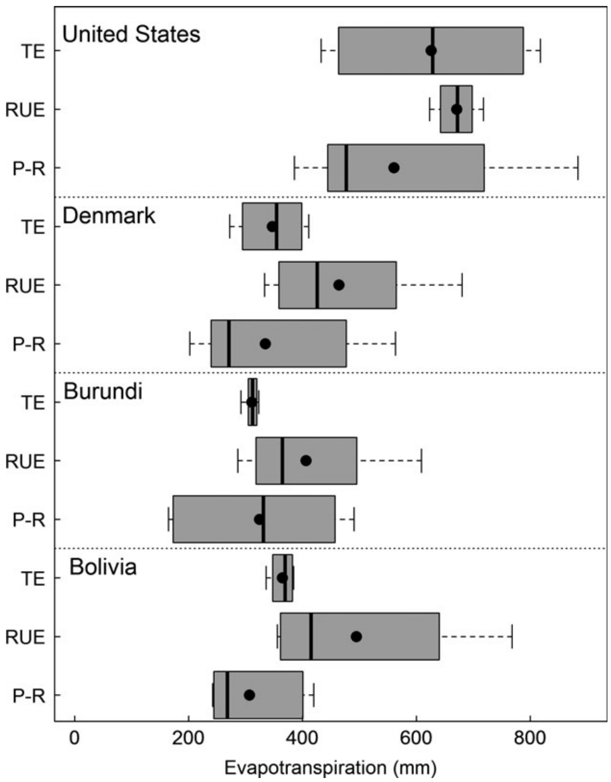

Fig. 11 Box-and-whisker plots of simulated evapotranspiration among 30-year averages of nine-models at each site aggregated across C levels based on model structure (type) as in Table 1.

so-called low- vs. high-input locations. A wide range of variance was found across these studies, between 10% and 28.1%, but no patterns were associated with environment, management, or genetic aspects. It was concluded that the 17% level from Vermeer (1990) was therefore a reasonable threshold for identifying the adequacy of a potato model ensemble.

Uncertainty with respect to site and climate change factors

As observed in the AgMIP Maize and Rice Pilots (Bassu *et al.*, 2014; Li *et al.*, 2015), variation among models for all responses was consistently higher for low-input (Burundi and Bolivia) vs. high-input (Denmark and the United States) sites (Tables 5 and 6) to the extent that an additional two models would be required within an ensemble approach to meet the natural variation threshold for dry yield (Fig. 12). The source of uncertainty was related to differences in which models simulated the onset and effects of W and nitrogen stress, although differences in environment (shorter photoperiods and warmer T) could not be completely ruled out. Tuber yields in Bolivia and Burundi were likely limited by W availability and nitrogen. Larger interannual weather variability was expected for rainfed conditions due to varying precipitation amounts and associated water stress impacts (Table 5), although even single baseline-year results exhibited larger uncertainty

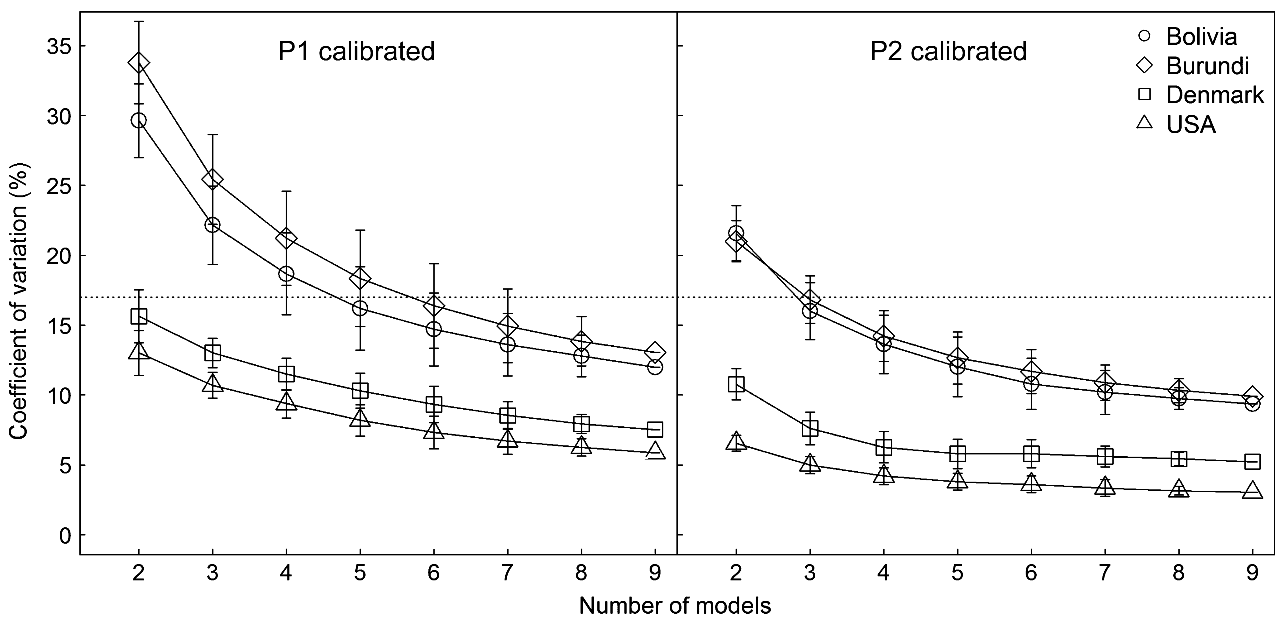

Fig. 12 Uncertainty represented by average coefficient of variation (CV %) from multi-model responses for simulated tuber dry yield for each site for two calibration phases P1 or P2. Error bars show standard deviations in CV across the number of models in each ensemble as chosen in Eqn (1). The dotted line represents the 17% variability threshold for yield as obtained from field experiments (Vermeer, 1990).

compared to high-input sites. This was also reflected in all multiyear simulation runs involving changes in rainfall (W). Variation among models increased as W declined from +30% to −30% of historical values (Fig. 8). In contrast, growth reductions due to limited water and nitrogen availability were likely minimal in Denmark and the United States because of irrigation and use of inorganic fertilizers (Table 2).

Model uncertainty for tuber dry yield and ET increased as treatment factors of T or W moved away from the baseline conditions under which models were fully calibrated (Table 6; Figs 8 and 10). Temperature sensitivity was also reflected in phenology responses (Fig. 4), particularly for tuber initiation. Large variations among the models with respect to high and low T reflect substantial differences in quantification of T effects on development, growth, carbon allocation, and even high temperature stress (Table 1). For example, three of the models for Bolivia, and one for Burundi, simulated tuber dry yield increases, and not decreases, in response to warming T, a result that appeared to be loosely correlated with the timing of tuber initiation for those models. Although variation with respect to C was higher for Bolivia and Burundi (Table 6), this level of uncertainty remained relatively constant with each site at varying C levels, similar to that reported for wheat across most tested locations (Asseng *et al.*, 2013). Only small interactions of C with either T or W were observed (Figs 8 and 10) with slight increases in variation at higher levels of C. This contrasted with maize (Bassu *et al.*, 2014) and rice (Li *et al.*, 2015) studies, which reported increased model uncertainty with rising C levels at the majority of locations tested. The minimum number of models required to keep uncertainty below that associated with commonly reported variation in the field (Fig. 12) was estimated using multiyear model runs with historical weather. As these results showed uncertainty increasing among models under climate change conditions involving shifts in T and W, the minimum ensemble number would be expected to increase as in Asseng *et al.* (2013).

Median responses to climate change factors

The models used in the current pilot were originally developed, tested, and validated for climatic conditions based on ambient (late 20th or early 21st century) C levels (around 360 to 400 ppm). As such, mean ensemble responses for changes in T and W levels, where C was fixed at 360 ppm, were expected to be consistent with the large amount of experimental literature. Increasing T from 0 to +9 °C levels above the historical baseline resulted in large relative yield declines ranging from 29% to 45% across sites (Table 7) and an absolute decrease in yield of 0.11, 0.37, 0.40, and 1.07 Mg ha^{-1} °C^{-1}, or 4.0, 6.1, 3.3, and 5.0% °C^{-1}, for Bolivia, Burundi, Denmark, and United States, respectively. Warming temperatures also resulted in earlier phenology dates and shortened growing season. Such patterns were reflected in the literature (e.g., Borah & Milthorpe, 1962; Marinus & Bodlaender, 1975; Wolf *et al.*, 1990; Timlin *et al.*, 2006). For example, Harahagazwe *et al.* (2012) found an average decrease of 0.44 Mg ha^{-1} °C^{-1} in Burundi when the same variety used in this study was grown in an area 8 °C warmer than Gisozi. Similarly, many cultivated varieties of potato are known to be drought sensitive resulting in yield declines as a result of loss of canopy development (Wolfe *et al.*, 1983; e.g., Costa *et al.*, 1997; Deblonde & Ledent, 2001; Tourneux *et al.*, 2003), and such trends were replicated by median responses for the rainfed locations (Table 7; Figs 4 and 6). Relative increases in simulated tuber dry yield ranged from 12% to 37% across sites when C was doubled from 360 ppm (Table 7). An increase of 0.3, 0.5, 0.4, and 1.5 Mg ha^{-1} 100 ppm^{-1}, or 11.4, 8.3, 3.2, and 6.8% 100 ppm^{-1}, was approximated for Bolivia, Burundi, Denmark, and United States, respectively.

These yield responses to C compared favorably with experimental data obtained for potato over a broad range of C levels (e.g., Schapendonk *et al.*, 2000; Donnelly *et al.*, 2001; Vandermeiren *et al.*, 2002; Finnan *et al.*, 2005; Conn & Cochran, 2006; Högy & Fangmeier, 2009; Kimball, 2011; Fleisher *et al.*, 2013b; Kaminski *et al.*, 2014). Median relative ET declined between 0.3% and 6% across sites representing a decrease of −0.4, −4.8, −6.6, and −4.6 mm 100 ppm^{-1} with respect to Bolivia, Burundi, Denmark, and United States. These small declines were generally below experimentally reported values which ranged between −3% and −12% for well-managed potato under high C conditions (Magliulo *et al.*, 2003; Fleisher *et al.*, 2008a, 2013a), although some controlled environment experiments indicated a slight increase in water use (Wheeler *et al.*, 1999). Harvest index and carbon allocation were reported to shift preferentially to tubers under higher C (Donnelly *et al.*, 2001; Lawson *et al.*, 2001; Fleisher *et al.*, 2008b, 2013b; Högy & Fangmeier, 2009; Kaminski *et al.*, 2014), but such trends were not reproduced in the median model responses at any C level (Table 7).

There is little published data regarding potato responses to C, W, and/or T interactive effects. Simulated C x W interactions on tuber dry yield (Fig. 7) and ET (Fig. 9) were primarily a result of differences between W1 (−30%) vs. W2 and W3 (0 and +30%) levels. Results were intuitive, with tuber dry yields increasing with C and W as noted experimentally (Fleisher *et al.*, 2008a,b, 2013a). C x T ensemble predictions showed a positive C response on yields (Fig. 7) and a negative one on ET (Fig. 9), but this response was generally unable to compensate for more negative impacts of rising T on production. Modeling literature generally shows an increase in ET with rising T (e.g., Goyal, 2004). In the present case, the ensemble response from the models indicated decreased seasonal ET due to a combination of shortened growth duration and decreased leaf area production (data not shown). The extent of yield and ET response to warming T also varied with C level and, in the case of ET, with location (Table 8). As noted in Asseng *et al.* (2013), T affects virtually all processes in the crop, including growth, development, and resource use, while increases in C are primarily associated with increased carbon gain (for C3 crops) and reduced ET in crop models. The present ensemble results confirm this larger influence T will have on simulated yield and water use responses under a warming climate. A representative worst-case climate change scenario illustrates ensemble predictions for C × T × W interactions. In this case, assuming a C of 720 ppm along with +9 °C increase for T and −30% decline in W, percent changes in yield with respect to baseline conditions ranged between −21% and −61% and ET between −3% and −35% across the four sites. As noted from the preceding single-factor discussion, the largest proportion of these relative changes was due to the influence of the highest T level similar to other model intercomparison studies (Makowski *et al.*, 2015), and uncertainty was typically at its highest at the warmest T and highest C levels (Fig. 8). Given the variability among models and differences attributed to geographic location, genotype, and crop management approaches, such ensemble results are perhaps more valuable for assessing relative impacts of climate change factors instead of providing absolute values. Note that these simulations also represent a worst-case scenario as no adaptations were assumed in any climate change scenario.

Consistent differences in model structure (i.e., RUE, P-R, or TE approaches), assessed on light utilization basis (Table 1), for simulated tuber dry yield and ET responses against changes in C, T, or W (Figs 3, 5, 6 and 11) were observed. Model means parsed according to structure held the same high-to-low ranking for either response for C, T, or W changes, but this ranking was location specific. Thus, definitive conclusions regarding the suitability, or accuracy, of a given model approach with respect to a particular geographic region or management style could not be reached. In contrast, Li *et al.* (2015) found RUE-type models gave small changes in rice yields in response to increasing C at one location, but simulated larger responses at others. Across all sites, TE approaches were less variable for both yield and ET, RUE methods exhibited most the uncertainty with respect to yield, and PR the most with respect to ET (data not shown), although this assessment reflects some limitations as only two models used the TE approach. A random process may be the best method for selecting potato models for use in an ensemble approach if studies were to be focused on multiple site locations.

The science of testing and improving crop models for climate change studies is partly constrained by the availability of experimental data, including assessment of potato production in response to climatic factors and other extreme growth conditions. Nonetheless, the present results indicate important findings can be obtained from comparison of simulated data from models trained against common datasets. Likely targets for model improvement were identified. For example, routines for simulating water and nitrogen uptake, utilization, and associated plant stress represent the components contributing to the largest variation. Changes in T resulted in the largest source of uncertainty out of the three climatic factors investigated, indicating the methods in which models implement temperature effects on developmental rates and growth processes represents one of the most important and variable aspects of the models. Simulated harvest index was scarcely influenced by rising C levels in contrast to aforementioned experimental results. This suggests carbon allocation, represents an additional modeling area that can be addressed to improve future forecasts.

Model calibration also plays a significant role. Differences between P1 and P2 indicated that (i) some observed dry matter is likely needed in addition to site characterization data for accurate representation of the highest and lowest yielding sites, (ii) variation among models decreases following calibration but not necessarily for all responses, and (iii) the type, utilization, and quality of available data may influence calibration and subsequent modeling results (e.g., modelers could not incorporate nitrogen or water uptake data in their calibration methods for most sites due to lack of availability). In this case, calibration influenced uncertainty for dry matter responses positively and resource values negatively, but whether this was due to knowledge represented in the model source code or the calibration methodology remains uncertain. This issue also pertains to the estimated number of P1 or P2 models required to meet variation thresholds. If agreement among simulations for yield and other aspects of potato production are highly dependent on calibration, then it can be argued that further exploration of calibration methodology and relationship to climate sensitivity should be explored in more rigorous fashion to understand the distinction between partially and fully calibrated results. This would ultimately provide a higher confidence in simulated effects of climate impacts on potato production where experimental data under natural production conditions are unavailable. These areas of investigation identified here form a basis for ongoing efforts to improve potato models.

References

Aggarwal PK, Kalra N, Chander S, Pathak H (2006) InfoCrop: a dynamic simulation model for the assessment of crop yields, losses due to pests, and environmental impact of agro-ecosystems in tropical environments I. Model description. *Agricultural Systems*, **89**, 1–25.

Asseng S, Ewert F, Rosenzweig C *et al.* (2013) Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, **3**, 827–832.

Baker JT, Allen LH Jr (1994) Assessment of the impact of rising carbon dioxide and other potential climate changes on vegetation. *Environmental Pollution*, **83**, 223–235.

Bassu S, Brisson N, Durand JL *et al.* (2014) How do various maize crop models vary in their responses to climate change factors? *Global Change Biology*, **20**, 2301–2320.

Boote KJ, Porter CH, Jones JW *et al.* (2015) Sentinel site data for model improvement - definition and characterization. In: *Improving Modeling Tools to Assess Climate Change Effects on Crop Response* (eds Hatfield JL, Fleisher D), pp. 1658–1672. ASA, CSSA, and SSSA, Madison, WI, USA.

Borah M, Milthorpe FL (1962) Growth of the potato as influenced by temperature. *Indian Journal of Plant Physiology*, **5**, 53–72.

Challinor AJ, Ewert F, Arnold S, Simelton E, Fraser E (2009) Crops and climate change: progress, trends, and challenges in simulating impacts and informing adaptation. *Journal of Experimental Botany*, **60**, 2775–2789.

Condori B, Hijmans RJ, Quiroz R, Ledent J-F (2010) Quantifying the expression of potato genetic diversity in the high Andes through growth analysis and modeling. *Field Crops Research*, **119**, 135–144.

Conn JS, Cochran VL (2006) Response of potato (*Solanum tuberosum* L.) to elevated atmospheric CO2 in the North American Subarctic. *Agriculture, Ecosystems and Environment*, **112**, 49–57.

Costa LD, Vedove GD, Gianquinto G, Giovanardi R, Peressotti A (1997) Yield, water use efficiency and nitrogen uptake in potato: influence of drought stress. *Potato Research*, **40**, 19–34.

Deblonde PMK, Ledent J-F (2001) Effects of moderate drought conditions on green leaf number, stem height, leaf length and tuber yield of potato cultivars. *European Journal of Agronomy*, **14**, 31–41.

Donnelly A, Craigon J, Black CR, Colls JJ, Landon G (2001) Elevated CO2 increases biomass and tuber yield in potato even at high ozone concentrations. *New Phytologist*, **149**, 265–274.

Easterling DR, Evans JL, Groisman PYa, Karl TR, Kunkel KE, Ambenje P (2000) Observed variability and trends in extreme climate events: a brief review. *Bulletin of the American Meteorological Society*, **81**, 417–425.

Ewing EE, Sandlan KP (1995) Biomass partitioning in the potato: considerations for modelling. In: *Modelling and Parameterization of the Soil-Plant-Atmosphere System: A Comparison of Potato Growth Models* (eds Kabat P, Marshall B, Van Den Boroek BJ, Vos J, Van Keulen H), pp. 7–16. Wageningen Pers, Wageningen, the Netherlands.

Ewing EE, Struik PC (1992) Tuber formation in potato: Induction, initiation, and growth. In: *Horticultural Reviews* (ed. Janick J), pp. 89–198. Wiley & Sons, New York.

Fangmeier A, De Temmerman L, Black C, Persson K, Vorne V (2002) Effects of elevated CO_2 and/or ozone on nutrient concentrations and nutrient uptake of potatoes. *European Journal of Agronomy*, **2002**, 353–368.

Fao (2015) FAOSTAT. Production of commodity (World).

Finnan JM, Donnelly A, Jones MB, Burke JI (2005) The effect of elevated levels of carbon dioxide on potato crops. *Journal of Crop Improvement*, **13**, 91–111.

Fleisher DH, Timlin DJ, Reddy VR (2008a) Elevated carbon dioxide and water stress effects on potato canopy gas exchange, water use, and productivity. *Agricultural and Forest Meteorology*, **148**, 1109–1122.

Fleisher DH, Timlin DJ, Reddy VR (2008b) Interactive effects of carbon dioxide and water stress on potato canopy growth and development. *Agronomy Journal*, **100**, 711.

Fleisher DH, Timlin DJ, Yang Y, Reddy VR (2010) Simulation of potato gas exchange rates using SPUDSIM. *Agricultural and Forest Meteorology*, **150**, 432–442.

Fleisher DH, Barnaby J, Sicher R, Resop JP, Timlin DJ, Reddy VR (2013a) Effects of elevated CO2 and cyclic drought on potato under varying radiation regimes. *Agricultural and Forest Meteorology*, **171–172**, 270–280.

Fleisher DH, Wang Q, Timlin DJ, Chun JA, Reddy VR (2013b) Effects of carbon dioxide and phosphorus supply on potato dry matter allocation and canopy morphology. *Journal of Plant Nutrition*, **36**, 566–586.

Foley JA, Ramankutty N, Brauman KA *et al.* (2011) Solutions for a cultivated planet. *Nature*, **478**, 337–342.

Garnett T, Appleby MC, Balmford A *et al.* (2013) Sustainable intensification in agriculture: premises and policies. *Science*, **341**, 33–34.

Godfray HCJ, Beddington JR, Crute IR *et al.* (2010) Food security the challenge of feeding 9 billion people. *Science*, **237**, 812–818.

Goyal RK (2004) Sensitivity of evapotranspiration to global warming: a case study of arid zone of Rajasthan (India). *Agricultural Water Management*, **69**, 1–11.

Hansen JW (2005) Integrating seasonal climate prediction and agricultural models for insights into agricultural practice. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **360**, 2037–2047.

Harahagazwe D, Ledent JF, Rusuku G (2011) Field performance of CIP potato clones in two contrasting environments in Burundi. *African Crop Science Conference Proceedings*, **10**, 393–402.

Harahagazwe D, Ledent JF, Rusuku G (2012) Growth analysis and modelling of CIP potato genotypes for their characterization in two contrasting environments of Burundi. *African Journal of Agricultural Research*, **7**, 6173–6185.

Haverkort AJ, Struik PC (2015) Yield levels of potato crops: recent achievements and future prospects. *Field Crops Research*, **182**, 76–85.

Haverkort AJ, Verhagen A (2008) Climate change and its repercussions for the potato supply chain. *Potato Research*, **51**, 223–237.

Haverkort AJ, Franke AC, Engelbrecht FA, Steyn JM (2013) Climate change and potato production in contrasting South African agro-ecosystems 1. Effects on land and water use efficiencies. *Potato Research*, **56**, 31–50.

Hijmans RJ (2003) The effect of climate change on global potato production. *American Journal of Potato Research*, **80**, 271–280.

Högy P, Fangmeier A (2009) Atmospheric CO2 enrichment affects potatoes: 1. Aboveground biomass production and tuber yield. *European Journal of Agronomy*, **30**, 78–84.

Hoogenboom G, Jones JW, Wilkens PW *et al.* (2015) Decision Support System for Agrotechnology Transfer (DSSAT) version 4.6 (www.dssat.net). Prosser, Washington, DSSAT Foundation.

Ipcc (2014) *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Geneva, Switzerland, IPCC.

Jones JW, Hoogenboom G, Porter CH *et al.* (2003) DSSAT cropping system model. *European Journal of Agronomy*, **18**, 235–265.

Kabat P, Marshall B, Van Den Broek BJ (1995) Comparison of simulation results and evaluation of parameterization schemes. In: *Modelling and Paramerization of the Soil-Plant-Atmosphere System. A Comparison of Potato Growth Models* (eds Kabat P, Marshall B, Van Den Broek BJ, Vos J, Van Keulen H), pp. 441–501. Wageningen Pers, Wageningen, the Netherlands.

Kaminski KP, Kørup K, Nielsen KL, Liu F, Topbjerg HB, Kirk HG, Andersen MN (2014) Gas-exchange, water use efficiency and yield responses of elite potato (*Solanum tuberosum* L.) cultivars to changes in atmospheric carbon dioxide concentration, temperature and relative humidity. *Agricultural and Forest Meteorology*, **187**, 36–45.

Kimball BA (2011) Lessons from FACE: CO_2 effects and interactions with water, nitrogen and temperature. In: *Handbook of Climate Change and Agroecosystems: Impacts, Adaptation, and Mitigation* (eds Hillel D, Rosenzweig C), pp. 87–107. Imperial College Press, London, UK.

Langsrud O (2003) ANOVA for unbalanced data: use Type II instead of Type III sums of squares. *Statistics and Computing*, **13**, 163–167.

Lawson T, Craigon J, Black CR, Colls JJ, Tulloch A-M, Landon G (2001) Effects of elevated carbon dioxide and ozone on the growth and yield of potatoes (*Solanum tuberosum*) grown in open-top chambers. *Environmental Pollution*, **111**, 479–491.

Li D, Liu H, Qiao Y *et al.* (2013) Effects of elevated CO2 on the growth, seed yield, and water use efficiency of soybean (*Glycine max* (L.) Merr.) under drought stress. *Agricultural Water Management*, **129**, 105–112.

Li T, Hasegawa T, Yin X *et al.* (2015) Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. *Global Change Biology*, **21**, 1328–1341.

Magliulo V, Bindi M, Rana G (2003) Water use of irrigated potato (*Solanum tuberosum* L.) grown under free air carbon dioxide enrichment in central Italy. *Agriculture, Ecosystems and Environment*, **97**, 65–80.

Makowski D, Asseng S, Ewert F *et al.* (2015) A statistical analysis of three ensembles of crop model responses to temperature and CO2 concentration. *Agricultural and Forest Meteorology*, **214–215**, 483–493.

Manrique LA, Hodges HF, Johnson BS (1990) Genetic variables for potato. *American Potato Journal*, **67**, 669–683.

Marin FR, Thorburn PJ, Nassif DSP, Costa LG (2015) Sugarcane model intercomparison: structural differences and uncertainties under current and potential future climates. *Environmental Modelling and Software*, **72**, 372–386.

Marinus J, Bodlaender KBA (1975) Response of some potato varieties to temperature. *Potato Research*, **18**, 189–204.

Martre P, Wallach D, Asseng S *et al*. (2015) Multimodel ensembles of wheat growth: many models are better than one. *Global Change Biology*, **21**, 911–925.

Matsuoka Y, Morita T, Kainuma M (2001) *Integrated assessment model of climate change: the AIM approach*, TERRAPUBA.

Mcdermid SP, Ruane AC, Rosenzweig C *et al*. (2015) The AgMIP coordinated climate-crop modeling project (C3MP): methods and protocols. In: *Handbook of Climate Change and Agroecosystems: The Agricultural Model Intercomparison and Improvement Project (AgMIP)* (eds Rosenzweig C, Hillel D), pp. 191–220. Imperial College Press, London, UK.

Miglietta F, Magliulo V, Bindi M, Cerio L, Vaccari FP, Loduca V, Peressotti A (1998) Free Air CO2 Enrichment of potato (*Solanum tuberosum* L.): development, growth and yield. *Global Change Biology*, **4**, 163–172.

Nendel C, Berg M, Kersebaum KC *et al*. (2011) The MONICA model: testing predictability for crop growth, soil moisture and nitrogen dynamics. *Ecological Modelling*, **222**, 1614–1625.

Palosuo T, Kersebaum KC, Angulo C *et al*. (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.

Rienecker MM, Suarez MJ, Gelaro R *et al*. (2011) MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate*, **24**, 3624–3648.

Rosenzweig C, Parry ML (1994) Potential impact of climate change on world food supply. *Nature*, **367**, 133–138.

Rosenzweig C, Jones JW, Hatfield JL *et al*. (2013) The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. *Agricultural and Forest Meteorology*, **170**, 166–182.

Rötter RP, Palosuo T, Kersebaum KC *et al*. (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.

Ruane AC, Goldberg R, Chryssanthacopoulos J (2015a) Climate forcing datasets for agricultural modeling: merged products for gap-filling and historical climate series estimation. *Agricultural and Forest Meteorology*, **200**, 233–248.

Ruane AC, Winter JM, Mcdermid SP, Hudson NI (2015b) AgMIP climate datasets and scenarios for integrated assessment. In: *Handbook of Climate Change and Agroecosystems: The Agricultural Model Intercomparison and Improvement Project (AgMIP)* (eds Rosenzweig C, Hillel D), pp.45–78. Imperial College Press, London, UK.

Schapendonk HCM, Van Oijen M, Dijkstra P, Pot CS, Jordi WJRM, Stoopen GM (2000) Effects of elevated CO2 concentration on photosynthetic acclimation and productivity of two potato cultivars grown in open-top chambers. *Australian Journal of Plant Physiology*, **27**, 1119–1130.

Shibu ME, Leffelaar PA, Van Keulen H, Aggarwal PK (2010) LINTUL3, a simulation model for nitrogen-limited situations: application to rice. *European Journal of Agronomy*, **32**, 255–271.

Sidak Z (1967) Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistics Association*, **62**, 626–633.

Singh U, Matthews RB, Griffin TS, Ritchie JT, Hunt LA, Goenaga R (1998) Modeling growth and development of root and tuber crops. In: *Understanding Options for Agricultural Production* (eds Tsuji GY, Hoogenboom G, Thornton PK), pp. 129–156. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Stalham MA, Allen EJ (2004) Water uptake in the potato (*Solanum tuberosum*) crop. *The Journal of Agricultural Science*, **142**, 373–393.

Steduto P, Hsiao TC, Raes D, Fereres E (2009) AquaCrop—The FAO crop model to simulate yield response to water: I. concepts and underlying principles. *Agronomy Journal*, **101**, 426.

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

Timlin D, Lutfor Rahman SM, Baker J, Reddy VR, Fleisher D, Quebedeaux B (2006) Whole plant photosynthesis, development, and carbon partitioning in potato as a function of temperature. *Agronomy Journal*, **98**, 1195.

Tourneux C, Devaux A, Camacho MR, Mamani P, Ledent J-FO (2003) Effect of water shortage on six potato genotypes in the highlands of Bolivia (II): water relations, physiological parameters. *Agronomie*, **23**, 181–190.

Tubiello FN, Ewert F (2002) Simulating the effects of elevated CO2 on crops: approaches and applications for climate change. *European Journal of Agronomy*, **18**, 57–74.

Usda-Nass (2015) Crop Production 2014 Summary. United States Department of Agriculture National Agricultural Statistics Service, ISSN: 1057-7823, 99p.

Van Dam J, Kooman PL, Struik PC (1996) Effects of temperature and photoperiod on early growth and final number of tubers in potato (*Solanum tuberosum* L.). *Potato Research*, **39**, 51–62.

Van Loon CD (1981) The effect of water stress on potato growth, development, and yield. *American Potato Journal*, **58**, 51–69.

Vandermeiren K, Black C, Lawson T, Casanova MA, Ojanpera K (2002) Photosynthetic and stomatal responses of potatoes grown under elevated CO_2 and/or O_3 - results from the European CHIP-programme. *European Journal of Agronomy*, **17**, 337–352.

Vermeer H (1990) Optimising potato breeding. I. The genotypic, environmental and genotype-enviroment coefficients of variation for tuber yield and other traits in potato (*Solanum tuberosum* L.) under different experimental conditions. *Euphytica*, **49**, 229–236.

Vos J (1995) Foliar development of the potato plant and modulations by environmental factors. In: *Modelling and Parameterization of the Soil-Plant-Atmosphere System: A Comparison of Potato Growth Models* (eds Kabat P, Marshall B, Van Den Broek R, Vos J, Van Keulen H), pp. 21–32. Wageningen Pers, Wageningen, the Netherlands.

Watanabe K (2015) Potato genetics, genomics, and applications. *Breed Science*, **65**, 53–68.

Wheeler RM, Mackowiak CL, Yorio NC, Sager JC (1999) Effects of CO_2 on stomatal conductance: do stomata open at very high CO_2 concentrations? *Annals of Botany*, **83**, 243–251.

White JW, Hoogenboom G, Kimball BA, Wall GW (2011) Methodologies for simulating impacts of climate change on crop production. *Field Crops Research*, **124**, 357–368.

Wolf J (2002) Comparison of two potato simulation models under climate change. II. Application of climate change scenarios. *Climate Research*, **21**, 187–198.

Wolf S, Marani A, Rudich J (1990) Effects of temperature and photoperiod on assimilate partitioning in potato plants. *Annals of Botany*, **66**, 513–520.

Wolfe DW, Fereres E, Voss RE (1983) Growth and yield response of two potato cultivars to various levels of applied water. *Irrigation Science*, **3**, 211–222.

Wolfe DW, Ziska L, Petzoldt C, Seaman A, Chase L, Hayhoe K (2007) Projected change in climate thresholds in the Northeastern U.S.: implications for crops, pests, livestock, and farmers. *Mitigation and Adaptation Strategies for Global Change*, **13**, 555–575.