Article (refereed) - postprint

Korhonen, Panu; Palosuo, Taru; Persson, Tomas; Höglind, Mats; Jégo, Guillaume; Van Oijen, Marcel; Gustavsson, Anne-Maj; Bélanger, Gilles; Virkajärvi, Perttu. 2018. **Modelling grass yields in northern climates – a comparison of three growth models for timothy**.

Contact CEH NORA team at
noraceh@ceh.ac.uk

1 Modelling grass yields in northern climates - a comparison of three growth models for timothy

2

3 Panu Korhonen[a*], Taru Palosuo[a], Tomas Persson[b], Mats Höglind[b], Guillaume Jégo[c], Marcel Van

4 Oijen[d], Anne-Maj Gustavsson[e], Gilles Bélanger[c], Perttu Virkajärvi[a]

5

6 [a] Natural Resources Institute Finland (Luke), Halolantie 31A, 71750, Kuopio, Finland

7 [b] Norwegian Institute of Bioeconomy Research (NIBIO), Postvegen 213, 4353 Klepp Stasjon,

8 Norway

9 [c] Agriculture and Agri-Food Canada (AAFC), 2560 Hochelaga Boulevard, Québec City, Québec, G1

10 V 2J3, Canada

11 [d] Centre for Ecology and Hydrology (CEH), Bush Estate, Penicuik EH26 0QB, UK

12 [e] Swedish University of Agricultural Sciences (SLU), SE-901 83 Umeå, Sweden

13

14 [*]Corresponding author: panu.korhonen@luke.fi

15

ABSTRACT

17 During the past few years, several studies have compared the performance of crop simulation models

18 to assess the uncertainties in model-based climate change impact assessments and other modelling

19 studies. Many of these studies have concentrated on cereal crops, while fewer model comparisons have

20 been conducted for grasses. We compared the predictions for timothy grass (*Phleum pratense* L.) yields

21 for first and second cuts along with the dynamics of above-ground biomass for the grass simulation

22 models BASGRA and CATIMO, and the crop model STICS. The models were calibrated and evaluated

23 using field data from seven sites across Northern Europe and Canada with different climates, soil

24 conditions and management practices. Altogether the models were compared using data on timothy

25   grass from 33 combinations of sites, cultivars and management regimes. Model performances with two

26   calibration approaches, cultivar-specific and generic calibrations, were compared. All the models

27   studied estimated the dynamics of above-ground biomass and the leaf area index satisfactorily, but

28   tended to underestimate the first cut yield. Cultivar-specific calibration resulted in more accurate first

29   cut yield predictions than the generic calibration achieving root mean square errors approximately one

30   third lower for the cultivar-specific calibration. For the second cut, the difference between the

31   calibration methods was small. The results indicate that detailed soil process descriptions improved the

32   overall model performance and the model responses to management, such as nitrogen applications. The

33   results also suggest that taking the genetic variability into account between cultivars of timothy grass

34   also improves the yield estimates. Calibrations using both spring and summer growth data

35   simultaneously revealed that processes determining the growth in these two periods require further

36   attention in model development.

Keywords

INTRODUCTION

Process-based crop simulation models that simulate crop growth, development, and yields, while taking into account the interactions between the crop genotype, management and environmental factors are increasingly used to support decision making and planning in agriculture, including aspects related to animal feed and forage production (Kipling et al., 2016). Several studies have recently been published on the comparison of the performance of crop simulation models under different environmental conditions in an effort to improve crop models and climate impact assessment projections and to gain an understanding of the uncertainties related to these assessments (see, e.g. Asseng et al., 2013, Asseng et al. 2015, Bassu et al. 2014, Pirttioja et al. 2015). In addition, there have been model-based evaluations of adaptation options to climate change (Ruiz-Ramos et al., 2017, Chenu et al. 2017).

To date, model comparisons and model ensemble studies have mostly focused on cereal crops and fewer model comparisons have been published for perennial forage grasses. Still, many crop models or crop modules of farm system models, e.g. STICS (Jégo et al., 2013) and APSIM (Keating et al. 2003), can also simulate forage grasses. There are also separate forage grass models (e.g. BASGRA, Höglind et al., 2016; CATIMO, Bonesmo and Bélanger, 2002a) that have comparable process descriptions to those in cereal crop models, such as radiation interception and use efficiency. Forage grass production systems, however, have specific characteristics that should be taken into account in simulation models. Almost all above-ground biomass is harvested several times during growing seasons and consequently the status of the plants after cuttings and regrowth are key issues for forage grass models (Jing et al. 2013). Another important aspect is the dynamically changing feed quality during forage development (Bonesmo et al. 2002b; Gustavsson and Martinsson, 2001). Finally, grass leys are typically perennial, which makes it essential to simulate the growth initiation in the spring (Bélanger et al. 2008) and thus, makes it important to develop the models to simulate relevant processes related to over-wintering particularly for forage grasses at high latitudes, where there is virtually no cold-season growth (Höglind et al. 2016).

64 To our knowledge, those forage grass model comparisons that have been conducted (e.g. Hurtado-Uria

65 et al., 2013) were restricted to regions with rather homogeneous climate conditions, without testing their

66 performance and suitability over a wide range of climate conditions. Even though the forage grass

67 modules of current farm system models and the separate forage grass models that are referred to above

68 are considered process-based, they all include several empirically derived functions. Therefore, one

69 could assume that the predictability of such forage grass models or modules would vary with climatic

70 and other environmental conditions. Hence, a comparison of crop simulation models across a wide

71 range of conditions is a key to strengthening the understanding of the effects of different model process

72 descriptions, i.e. model structures, on yield and quality related output variables.

73 Moreover, varying the genetic variability and climate and soil conditions within the calibration and

74 evaluation datasets could provide knowledge about the calibration procedures as well as knowledge

75 about model application strategies. Persson et al. (2014) found that the observed dry matter yield for

76 one variety (cv Grindstad) was more accurately predicted by LINGRA model (the predecessor of the

77 BASGRA model) when the parameters were calibrated against data from several locations within a

78 region with heterogeneous climate and soil conditions than when the parameters were calibrated against

79 data from only one location. However, to the best of our knowledge no study has been published where

80 the effects varying the genetic variability in calibration datasets on grassland dry matter yield were

81 evaluated. Such a study providing knowledge about model sensitivity to genetic variability could be

82 used to arrange field trial data for model calibration. It could also give useful information about how

83 calibrate and apply grassland models for genetically heterogeneous conditions, such as in estimations

84 of regional or national grassland productivity.

85 Timothy (Phleum pratense L.) is one of the most important forage grass species in the cold temperate

86 climate zone of the northern hemisphere, including Canada and the Nordic countries in Europe.

87 Management of timothy swards varies considerably according to the climate and soil conditions where

88 it is grown and with its end use. Timothy is grown either in pure stands or mixed with other grasses or

89 leguminous species for three years or longer before it is ploughed up and reseeded or rotated with an

90 annual crop, usually a cereal crop. When used to feed dairy cows, timothy is often cut and harvested at

91   the mid-heading stage to optimize the nutritive value and yield. When used as feed for beef cattle and

92   sheep, timothy is usually cut at later stages, which generally results in higher yields but a lower nutritive

93   value. The number of harvests per year usually varies from two to four depending on the cutting

94   strategy, the cultivar-specific characteristics such as the development rate and its effect on nutrient

95   composition, as well as the climate and weather conditions. In addition, plant characteristics, such as

96   the maximum tiller height, pattern of development of vegetative and reproductive tillers and the

97   leaf/stem ratio, vary between cultivars, which have been bred to meet different regional climate

98   conditions and management practices (Virkajärvi et al., 2010). Considering the range of environmental

99   conditions, alternative management strategies and the genetic variability of timothy, a model

100   comparison with timothy data covering different environmental conditions, a wide set of different

101   cultivars and alternative management options would provide material for critical testing of crop

102   simulation models.

103   The overall aim of this study was to assess and compare the ability of simulation models to accurately

104   simulate the growth and yield of the first and second annual cuts of timothy under different

105   environmental conditions. To this end, the performance of two grass simulation models, BASGRA and

106   CATIMO and the soil-crop model STICS were assessed with a comprehensive experimental dataset

107   collected from across Northern Europe and Canada with varying management practices. The three

108   models were calibrated either specifically for each cultivar (cultivar-specific calibration) or for a

109   number of cultivars all together (generic calibrations) and the performance of the models with both

110   calibrations was tested.

MATERIALS AND METHODS

Models

113   The three models simulating the growth and development of timothy as a function of weather, soil and

114   crop management factors included in the comparison were: CATIMO (R-version 1.0; Bonesmo and

115   Bélanger, 2002a, b; Jing et al. 2012 and 2013), BASGRA (version 2014; Höglind et al., 2001; van Oijen

et al., 2005; Höglind et al., 2016) and STICS (v8.4; Brisson et al., 1998, 2008; Jégo et al., 2013; Jing et al., 2017). All three models use the radiation use efficiency (RUE) approach instead of calculating the photosynthesis and respiration in detail and they use a simple leaf area index (LAI) to calculate light interception (Table 1). They all cover soil water and N effects in the simulation of grass growth, but the effect of excess water is only taken into account by the STICS model. The representation of soil processes differs considerably between the three models. While in the STICS model there are detailed sub-routines for soil water and N balances, the CATIMO and BASGRA models are simpler with a description of the root zone as one single homogenous layer. Tillering dynamics and vernalisation are simulated only in BASGRA. The CATIMO model starts biomass accumulation in spring when the daily mean temperature exceeds the base temperature (default 5 °C) (Table 1). In this study, where over-wintering was not included, we initiated the spring growth in the BASGRA simulation the first time of the year that the mean air temperature exceeded 5°C for five consecutive days or longer. In the STICS simulation, biomass accumulation starts after model initialization as soon as the daily average air temperature is above the minimum threshold temperature for net photosynthesis (parameter *temin*). In the case of multi-annual simulations, winter dormancy processes can also be simulated with STICS, but they were not used in this study.

Table 1. Approaches used by the three models for the major processes determining crop growth and development.

	Models		
Process	BASGRA	CATIMO	STICS
Leaf area development and light interception[a]	S	S	S
Light utilization[b]	RUE	RUE	RUE
Root distribution over depth[c]	-	-	Sig
Drought stress[d]	ETa/ETp	ETa/ETp	ETa/ETp
Water dynamics[e]	C	C	C
Evapo-transpiration[f]	PM	PM	P
Effect of nitrogen[g]	NSNS	RNC	RNC
Tillering dynamics[h]	C	-	-
Vernalisation[i]	SV	-	-
Start of spring growth[j]	5D	5Drm	1D
Regrowth dynamics[k]	SSDG	RDG	RDG
Soil C/N model[l]	CN, P(3)	N	CN, P(3), B

[a] Leaf area development and light interception: S = simple approach (e.g. leaf area index (LAI)

[b] Light utilization or biomass growth: RUE = simple (descriptive) radiation use efficiency approach

[c] Root distribution over depth: Sig = sigmoidal

[d] Drought stress: ETa/ETp = actual to potential evapotranspiration ratio

[e] Water dynamics approach: C = capacity approach

[f] Method to calculate evapotranspiration: P = Penman; PM = Penman–Monteith

[g] Effect of nitrogen: RNC = relative nitrogen concentration as the ratio of the actual N concentration to the critical N concentration (Bélanger & Richards, 1997), NSNS = N source/N sink balance dependent growth

[h] Tillering dynamics: C = Three different tiller categories (dependent on internal as well as external factors)

[I] Vernalisation: SV = simple approach (threshold temperature)

[j] 5D = the first day of 5 consecutive days above base temperature, 5Drm = the first day when the running mean of a five-day daily mean temperature is above base temperature; 1D = 1 day above base temperature (start defined by model user)

[k] Regrowth dynamics: RDG = reserve dependent growth, SSDG= source (LAI, reserves) and sink (tillers) dependent regrowth

[l] Soil C/N model: CN = soil CN model, N = soil N model with only mineral N; P(x) = number of organic matter pools; B = microbial biomass pool

Experimental sites

For the model intercomparison, crop data were obtained from field experiments at sites across

Northern Europe and Canada with different cultivars (Table 2, Fig. 1).
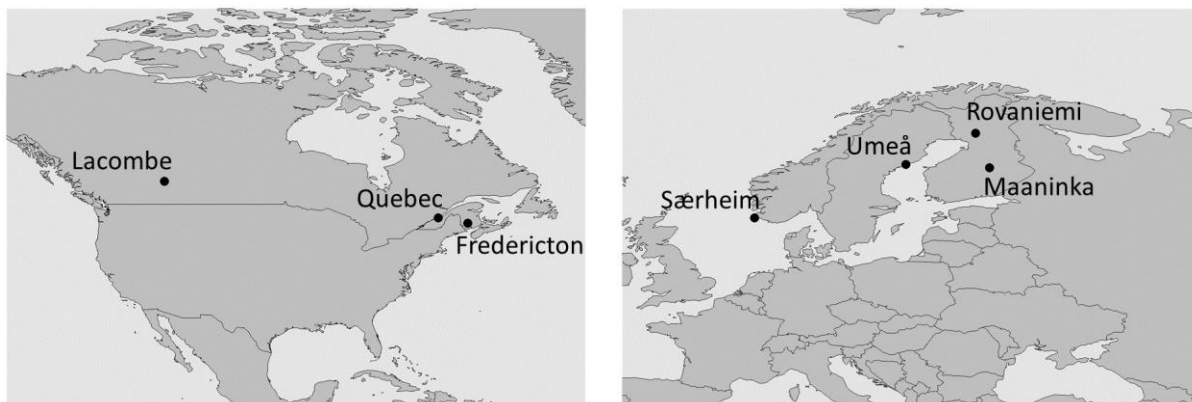
Fig. 1. Locations of the experimental sites.

Together, these sites comprise a wide range of climate, soil and management practices that are

associated with timothy production in its main production regions of Canada and Northern Europe. Data

from two to three growing seasons per site were available and they covered altogether 33 different

combinations of site, year, cultivar and management practices (below called treatments) (Table S1).

The cultivars differed between the sites, except that the cultivar 'Champ' was used in both Fredericton

and Québec (Table 2).

164 **Table 2.** Characteristics of experimental sites and their climatic conditions for the study years.

Location	Position lat/lon/alt (m.a.s.l)	Precipitation (mm yr^{-1})	Mean annual temperature (°C)	Data description	Period	Cultivar	Parallel treatments	Observed crop data*
Maaninka, Finland	63.14N / 27.32E / 90 m	560	4.2	Virkajärvi et al., 2012	2006-2007	Tammisto II	-	DM, sward height, LAI
Rovaniemi, Finland	66.35N / 26.01E / 106 m	610	1.0	Nissinen et al., 2010	1999-2001	Iki	two N application rates	DM, sward height, DM$_{leaf}$, DM$_{stem}$
Umeå, Sweden	63.45N / 20.17E / 12 m	595	3.3	Gustavsson and Martinsson 2001	1995-1996	Jonatan	-	DM, sward height, DM$_{leaf}$, DM$_{stem}$
Særheim, Norway	58.36N / 5.39E / 90 m	1392	8.0	Höglind et al., 2005	2000-2002	Grindstad	early and late cutting regimes	DM, LAI, SLA, tiller density
Fredericton, Canada	45.55N / 66.32W / 35 m	1108	5.7	Bélanger and Richards, 1997	1991-1993	Champ	varying N application rates	DM, DM$_{leaf}$, DM$_{stem}$, LAI
Québec, Canada	46.47N / 71.07W / 75 m	1009	5.3	Bélanger et al., 2008	1999-2001	Champ	three N application rates	DM
Lacombe, Canada	52.28N / 113.44W / 860 m	429	3.5	Jing et al., 2012; Jégo et al., 2013	2004-2005	Climax	-	DM, LAI

165 *Abbreviations: DM = dry matter yield, DM$_{leaf}$ = dry matter yield of leaves, DM$_{stem}$ = dry matter yield of stems, LAI = leaf area index, SLA

166 = specific leaf area

167

Setup of model comparison

Information provided for model calibration and runs

170 Comprehensive data on crop performance and management, as well as soil and weather conditions were

171 obtained from all sites in the study to allow for detailed model calibrations and evaluations. These were

172 conducted for each model independently. We randomly selected one treatment for model evaluation

173 from each treatment type from sites with more than two treatments or years, i.e. Rovaniemi, Særheim,

174 Québec, and Fredericton (in total 9 treatments) and used the remaining 24 for model calibration (Table

175 S1). This approach differs from previous model comparisons (Palosuo et al., 2011; Rötter et al., 2012;

176 Asseng et al. 2013) that applied a "blind test" approach, i.e. model users were not provided with the

177 observed data for crop yields, biomasses or other variables delivered as model results. Here instead, we

178 tried to comprehensively apply all the data available for calibration treatments.

179 The input data included observed daily weather (i.e. minimum and maximum temperature, precipitation,

180 relative air humidity, wind speed and global radiation), sward and soil management (cut height, cutting

181 and fertilisation dates and amounts) and basic information on soil properties (soil texture, hydraulic

182 characteristics and the pH). Standard deviations (SD) were calculated for the Maaninka and Rovaniemi

183 crop datasets based on original replicates. For other datasets, replicate data were not available and the

184 SDs were calculated assuming a coefficient of variation with a value of 0.05 for the yield, 0.1 for LAI,

185 and 0.5 for other observations.

186 <u>Simulation setup</u>

187 The models were run on an annual basis, starting simulations for each growing season in the spring.

188 The starting dates for the simulations were determined with the typical criteria used with each model.

189 For the CATIMO model the start of simulations was determined by a threshold temperature for the

190 running mean of a five-day daily mean temperature (default 5 °C). For BASGRA, the simulations

191 started the first time of the year that there were at least five consecutive days with a mean daily air

192 temperature above 5°C. For the STICS model, simulations were initialized on 15 April at all sites,

193 except Rovaniemi (10 May) to account for the very short growing season and at Særheim (1 March) to

194 account for the mild winter conditions prevailing on the south-west coast of Norway. Although timothy

195 swards are perennial, winter seasons were not simulated because not all of the compared models covered

196 the relevant winter processes, and the focus in this study was on comparing models for the growth

197 estimates during the growing seasons. Simulations were carried out using a daily time step in all models

198 and they lasted until the last observation of summer growth in the observed data.

Calibration of the models

The models were calibrated by adjusting their parameters so that their output fitted the provided observed data either: 1) separately for each cultivar (*cultivar-specific calibration*), or 2) using data for all cultivars at all sites together (*generic calibration*). The BASGRA and CATIMO models were calibrated using Bayesian techniques. Both calibrations were performed by means of Markov Chain Monte Carlo (MCMC) sampling using the Metropolis algorithm (Van Oijen et al., 2005). Beta distributions for all parameters under calibration were used. The chain length, i.e. the number of model runs was 150,000 with BASGRA and varied from 150,000 to 200,000 with CATIMO. The prior ranges and the maximum a posteriori (MAP) values of the parameters for BASGRA and CATIMO are listed in Tables S2 and S3, respectively. 75 for BASGRA and 8 parameters for CATIMO were adjusted for both cultivar-specific and generic calibrations. Parameters that were not included in the calibration and the prior probability distributions of calibrated parameters for CATIMO were taken from earlier CATIMO studies (Bonesmo et al., 2002a, b; Jing et al., 2011 and 2012). The prior probability distributions for BASGRA were based on common knowledge about the parameters and assumptions taking into account results from previous model evaluations for the cultivar Grindstad under conditions representing northern Europe (Persson et al., 2014).

The STICS model calibration was done using the STICS built-in optimization tool based on a simplex algorithm (Nelder and Mead, 1965). For the generic calibration, 17 parameters (Table S4) were simultaneously calibrated by minimizing the RMSE between simulated and measured biomass and the LAI. Values of other parameters were those used in Jégo et al. (2013). For the cultivar-specific calibration, a sub-selection of 6 parameters (identified as "cultivar parameters" in STICS) was calibrated, while for the other 11 parameters the same values as in the generic calibration were used.

Methods used for evaluating model performance

First, the model performances for the timothy growth were evaluated by comparing sequential above-ground biomass observations with the corresponding simulated values during growth (Fig. 2). This was done for spring growth (from the start of growth in spring to the first cut) and summer growth (from the first cut to the second cut) periods separately. Then, the model performance for the first and second cut

226   yields were assessed, i.e. the observed and simulated above-ground biomass for the latest observed data

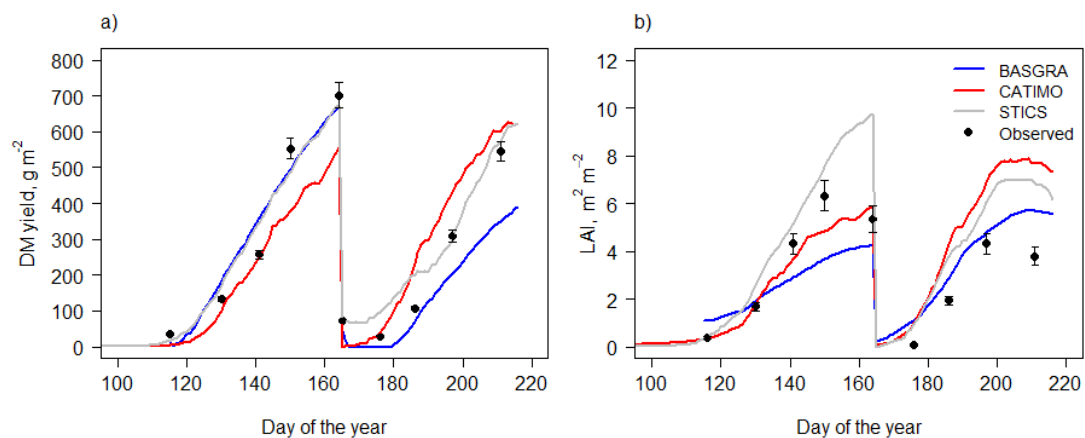227   point of spring and summer growths were compared.

Fig. 2. Simulated (cultivar-specific calibration) and observed (a) above-ground biomass (dry matter) and (b) leaf area index (LAI), using three timothy models at Særheim in 2001 (treatment 11). Error bars represent ±1 standard deviation. The above-ground biomass dynamics for all sites are shown in Figure S1.

228

229   The models and the calibration methods were evaluated based on their prediction of above-ground

230   biomass and LAI dynamics as well as first and second cut yields. Observed measurements were

231   considered true values, though we acknowledge that there is always uncertainty in the measurements.

232   The root mean square error (RMSE) was used to show the average difference between the model

233   estimates and the measurements.

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

234   where O_i is the observed value and P_i is the model prediction and N is the number of estimate-

235   observation pairs. In addition, normalized RMSE values (RMSE%) were calculated by dividing RMSEs

236   by the mean values of observations.

237   The relative mean bias error (rMBE) was used as an indicator to describe whether the model-predictions

238   were over- or underestimates and their relative magnitude.

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

239   Simulated yield responses to spring N fertilisation were analysed using correlations between fertiliser

240   application rates and simulated final spring growth yields for all experiments and by comparing these

241   results to corresponding correlations for observed yields.

RESULTS

242

Dynamics of above-ground biomass and leaf area index

243

244   The start of spring growth varied between the models with STICS simulating a 21-day earlier start of

245   spring growth on average than the other two models with the cultivar-specific calibration (Fig. S1).

With the cultivar-specific calibration, the spring growth was simulated most accurately by STICS (rMBE = -0.01). BASGRA (rMBE = -0.12) and CATIMO (rMBE = -0.13) tended to underestimate the above-ground biomass (Fig. 3).

a)
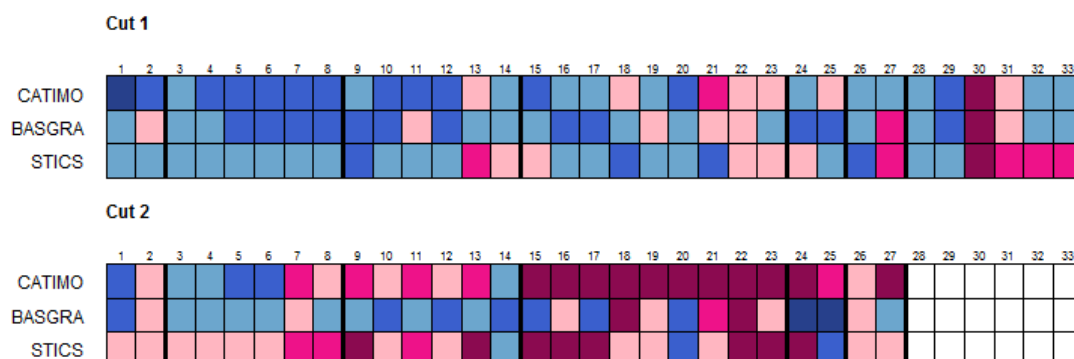
b)
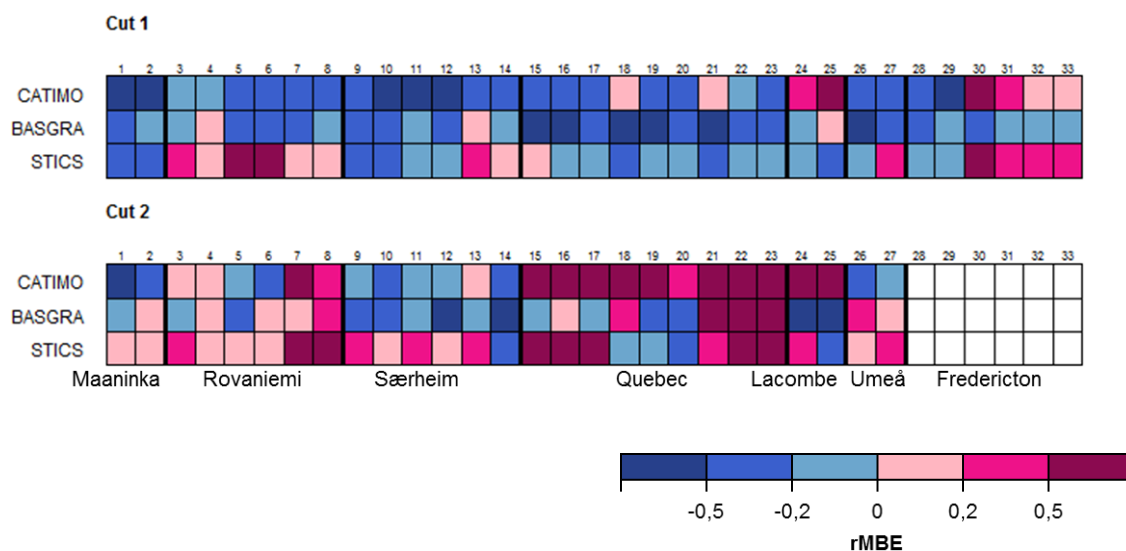
Fig. 3. Relative mean bias error (rMBE) of the simulated above-ground biomass of three timothy models grouped in quantiles with overestimations shown in red shades and underestimations shown in blue shades (see colour scale) during the spring growth (Cut 1) and summer growth (Cut 2) for all 33 treatments simulated with (a) cultivar-specific calibration and (b) generic calibration.

The summer was best reproduced by the BASGRA model, albeit with an underestimation (rMBE = -0.16), whereas STICS (rMBE = 0.22) and CATIMO (rMBE = 0.27) overestimated the above-ground

252 biomass for first summer growth. In terms of RMSEs, the model performance for the spring growth for

253 the treatments used for model evaluation was best in the STICS model (Table 3). For the summer growth

254 and the resulting second cut, however, the RMSEs in the BASGRA and STICS models were similar but

255 fewer than in the CATIMO model (Table 3).

256 **Table 3**. The mean of site-specific root mean square errors (RMSE) for the simulated above-ground

257 biomass (g DM m^{-2}) from the calibration and model evaluations of three timothy models during the

258 spring and summer growth periods.

Model	BASGRA				CATIMO				STICS			
Calibration	cultivar-specific		generic		cultivar-specific		generic		cultivar-specific		generic	
	Spring growth	Summer growth	Spring growth	Summer growth	Spring growth	Summer growth	Spring growth	Summer growth	Spring growth	Summer growth	Spring growth	Summer growth
Mean over all treatments (n = 33)	120	73	178	87	113	190	168	142	87	92	126	104
Mean over calibration treatments (n = 24)	90	70	159	86	105	183	167	146	85	95	116	111
Mean over evaluation treatments (n = 9)	199	83	227	93	135	210	171	132	94	81	153	83

259

260

261 The accuracy of the simulated above-ground biomass dynamics for calibration treatments was better

262 than that for the treatments used for evaluating model performance; the RMSEs for the spring growth

263 were from 10% to 55% smaller for the calibration treatments than for the evaluation treatments

264 depending on the model (Table 3).

265 For all three models, the cultivar-specific calibration resulted in greater simulation accuracy than the

266 generic calibration for the spring growth (Fig. 3). The mean RMSEs of all treatments were smaller with

267 the cultivar-specific calibration than with the generic calibration for BASGRA (-33%), CATIMO (-

268 33%) and STICS (-31%) (Table 3). For the summer growth, the mean RMSEs for all treatments were

269 smaller with the cultivar-specific calibration than with the generic calibration for BASGRA (-16%) and

270 STICS (-12%), whereas for CATIMO the mean RMSEs were 34% higher with the cultivar-specific

271 calibration than with the generic calibration (Table 3).

272 All three models performed relatively well in simulating LAI dynamics (Fig. S2). The RMSEs for

273 calibration (altogether over 12 treatments at four sites) and evaluation (over two treatments in Særheim

274 and two in Fredericton) were 1.66 and 1.75 m^2 m^{-2} for BASGRA, 1.72 and 1.57 m^2 m^{-2} for CATIMO

275 and 1.25 and 1.69 m^2 m^{-2} for STICS, respectively. The CATIMO model tended to simulate high LAI

276 values for the summer growth period when LAI data was not available for calibration (Fig. S2).

First and second cut yields

278 For both cultivar-specific and generic calibrations, the first cut yields were mostly underestimated by

279 all models (Fig. 4). The estimates simulated by STICS were closer to the observed yield than those

280 simulated by BASGRA and CATIMO. The first cut performance of all models for the model evaluation

281 treatments was somewhat weaker than their performance for calibration treatments. For the second cut,

282 STICS provided the most accurate yield estimates, while BASGRA somewhat underestimated the

283 yields on average on all sites and CATIMO systematically overestimated the second cut yields in

284 Québec (Fig. 4).

285

286 a)
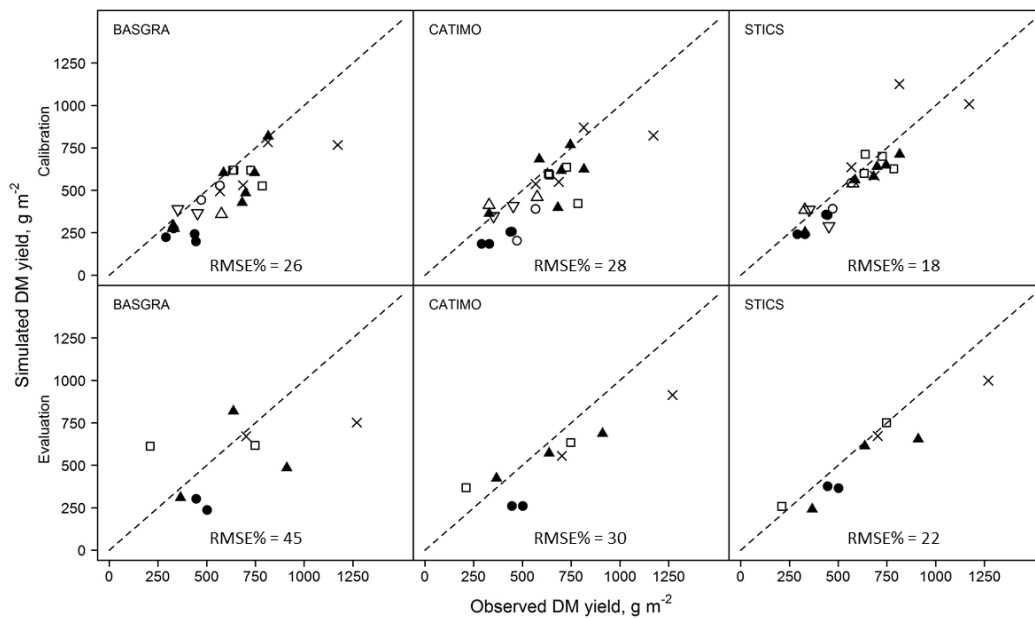
287
288 b)
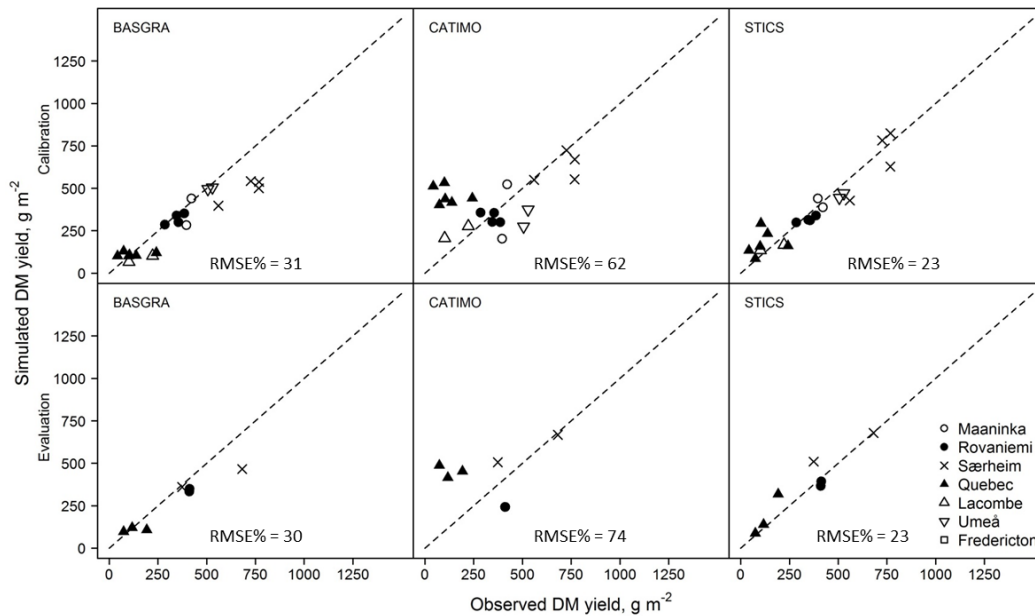
289
290 **Fig. 4**. Simulated (cultivar-specific calibration) and observed yield estimates of the three timothy

291 models for: a) first cut, and b) second cut for the 24 calibrations and nine evaluation treatments.

292 Different study sites are depicted with different symbols. The 1:1 line represents perfect agreement

293 along with the normalized root mean square errors (RMSE%).

294

295 The ranges of the simulated estimates of first and second cut yields between the models were not

296 affected by the calibration method used except for the second cut in Québec where the range clearly

297 widened when the cultivar-specific calibration was used (Fig. 5). The ranges of both calibration methods

298 (generic or cultivar-specific) were close to each other and no systematic differences were observed. In

299 most treatments, the ranges of both calibration methods overlapped, at least partially. However, at some

300 sites, the effect of the calibration methods was more noticeable. For example, in Québec the first cut

301 yield estimates with the generic calibration were clearly lower than the estimates with the cultivar-

302 specific calibration. In Rovaniemi and Lacombe, on the other hand, the generic calibration provided

303 higher first cut yield estimates than the cultivar-specific calibration.

304 The observed first cut yield was within the simulated range for 18 treatments (Fig. 5). For the 15

305 remaining treatments, none of the models/calibrations could achieve the observed first cut yields, which

306 indicates systematic under-prediction in the models. For the second cut, most of the observed yields

307 were included in the simulated range except for one treatment in Særheim (model underprediction) and

308 two treatments in Québec (model overprediction).

309 The yield estimates calculated as means and medians over the three models for first and second cuts

310 using cultivar specific calibration were better in terms of the RMSEs for all sites, than the yield estimates

311 of BASGRA or CATIMO (data not shown). The first cut yield estimates by the STICS model were

312 more accurate or about same performance level than both the multi-model mean and median with both

313 calibration methods.

314
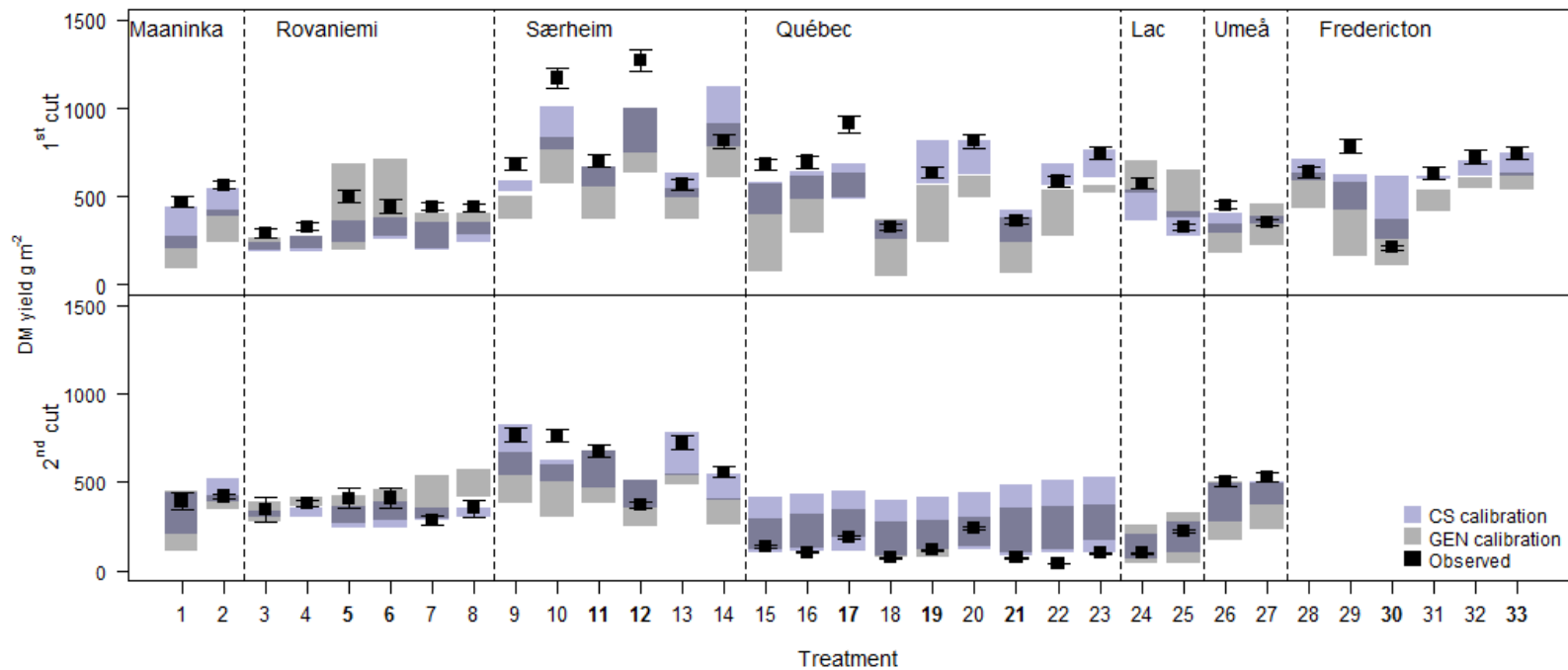
Fig. 5. Yield ranges of observations and simulationsby the three models for the first cut (above) and second cut (below) for all 33 treatments at the seven study sites. The evaluation treatments are marked in bold numbers on the x-axis. Observations are marked with black squares with the standard deviation as error bars. The simulation results with cultivar-specific (CS) and generic (GEN) calibrations are denoted by blue and grey bars, respectively. The site abbreviation Lac denotes the Lacombe study site.

321

322   Overall, the first cut yields correlated positively with nitrogen fertilisation in both the observed and

323   simulated data over all studied sites and treatments (Fig. 6). The STICS model with more a detailed

324   description of soil N mineralisation and water balance than the other two models managed to simulate

325   yield responses to increasing N fertilisation more accurately than the BASGRA and CATIMO models,

326   which both underestimated the N fertilisation response.
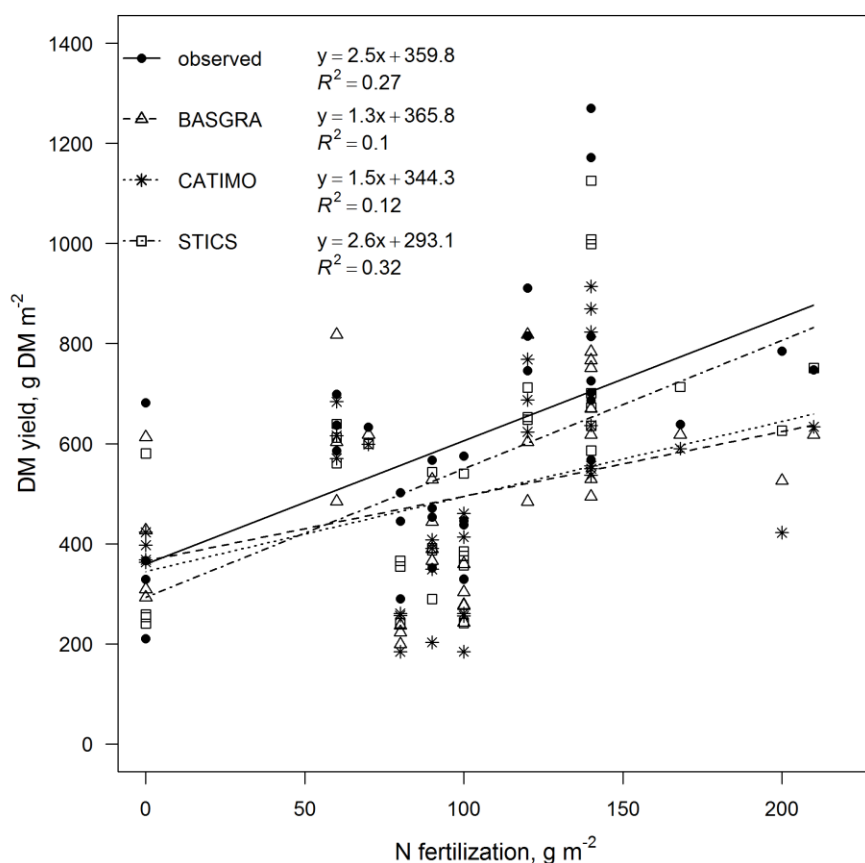
327

328 **Fig. 6**. Regression lines for observed and simulated first cut yields as a function of the N fertilisation

329 rate in the spring.

330

DISCUSSION

Model performances

This study is, to our knowledge, the first comparison of yield simulations from models to simulate timothy, and it shows that the three models estimated the dynamics of above-ground biomass and LAI satisfactorily, although the estimated ranges of the simulation results between the models were quite wide. Despite several sequential observations of the above-ground biomass and LAI provided for model calibration, the model results for many treatments did not match the observed values. The simulated values for summer growth and the resulting yield for the second cut were generally very good, although one model (CATIMO) could not cover the obvious N limitation at one site (Québec).

The simulation accuracy of the three models was roughly the same as the model performances documented in earlier studies of these three models (Jing et al., 2012; Jégo et al., 2013; Persson et al., 2014). The RMSE% for the yields of evaluation treatments with cultivar-specific calibration varied between 22% and 45% for the first cut and between 23% and 74% for the second cut (Fig. 4). In a previous study, Jégo et al. (2013) found the STICS model simulated timothy above-ground biomass at Canadian study sites used for model evaluation to an RMSE% of about 16% for spring growth and 89% for summer growth. The reduction of the values for the optimum and maximum temperature for photosynthesis (*teoptbis* and *temax*) compared to the previous calibration probably contributed to the better simulation of the summer growth in the STICS model.

The CATIMO model, on the other hand, has been found to simulate the yield with an accuracy of 24% (RMSE%) for the first cut and 30% for the second cut (Jing et al., 2012), which are lower figures than those observed in this study. Summer growth dynamics and the second cuts yields were generally well covered in the CATIMO model, but it clearly overestimated the second cut yields in Québec. At that site, there were no N applications for summer growth and the crop growth was probably N-limited. The CATIMO model could not satisfactorily take this into account. The poor accuracy of the CATIMO model in simulating the summer growth, primarily with no applied N, at the Québec site has also been

356   noted by Jing et al. (2012) and also the poor performance for the spring growth with no applied N has

357   been noted by Bonesmo et al. (2005) and Jing et al. (2011).

358   In an earlier study by Persson et al. (2014), the accuracy (RMSE%) of the LINGRA model, the

359   predecessor of BASGRA, was similar for the first cut yield (43%) and higher for the second cut yield

360   (46%) than was observed in this study. The BASGRA model underestimated the first cuts yields, which

361   is in contrast with the overestimation of the first cut yield by the LINGRA model in a study by Persson

362   et al. (2014) where the model was calibrated for Særheim in Norway and validated for several locations

363   in the Nordic region of Europe. However, the generally higher observed yields used for calibration at

364   Særheim than at the evaluation sites might explain some of the overestimation of the yields at the latter

365   sites. Moreover, the earlier version of the LINGRA model did not simulate nitrogen functions, which

366   may explain some of the differences between the two model evaluations.

367   The simulation accuracy for the timothy models tested here was somewhat lower than that observed for

368   the cereal models in model intercomparison studies in Europe. Reported RMSE% of grain yield

369   estimates in those studies for winter wheat were 23-38 (Palosuo et al., 2011) and for spring barley 19-

370   33 (Rötter et al., 2012). This is not surprising considering the challenges of modelling multi-year

371   timothy swards related, for example, to the status of the crop after the previous growing season(s) and

372   over-wintering, multiple cuttings and regrowth processes. Accurate simulation of crop carbon and

373   nitrogen reserves plays an important role both for multi-annual simulations and regrowth. Additionally,

374   the establishment of the root system and root turnover, both of which are still poorly understood

375   processes, are more important for perennial crops than for annual crops.

376   Precisely simulated start of growth could be of high importance for projections of soil water status and

377   biomass development. It has particular importance in studies projecting the impacts of climate change,

378   as the start of growing seasons is projected to change due to increasing temperatures (Ruosteenoja et al.

379   2015). Here, we found differences in the simulated start of the growth of up to 55 days. However,

380   although the STICS model simulated the start of the spring growth earlier than BASGRA and CATIMO,

381   the early biomass accumulation was simulated relatively similarly in all models Further model

evaluations where more variable winter and spring precipitation, and soil water patterns are included could reveal information about how general this insensitivity of biomass prediction to spring initialization date is.

Model calibrations and input data

In contrast to the blind-test approach applied in some earlier crop model intercomparison studies (e.g. Palosuo et al., 2011; Rötter et al., 2012), here a detailed calibration of the models was allowed. Calibration was considered particularly important as the study involved sites and cultivars from two different continents including a large variation in conditions, such as the length of the growing season and the winter period, air temperature and precipitation patterns, day length and soil type. The evaluation treatments from sites with more than two treatments were used as an independent test of the model performance. Our hypothesis that cultivar-specific calibration would work better than generic calibration was confirmed for the first cut, but for the second cut the results were not so clear. This indicates that taking into account the genetic variability between cultivars could improve estimations of the production potential of this species, particularly for the first cut. Persson et al. (2014) showed that site-specific calibration of the BASGRA model provided more precise results on the site level, whereas for the whole Nordic region, wider calibration provided better results. In our case, for some sites, the small quantity of observed data for calibration reduced the information available for cultivar-specific calibration so much that the generic calibration provided more stable results. For example, for Québec there was only one DM observation for the summer growth period.

The calibration data used in our study were exceptionally comprehensive when compared to the calibration data used in recent crop model intercomparisons (e.g., Asseng et al., 2013). The data included several sequential data series for various treatments covering altogether more than 1,500 observations. The uncertainty of individual yield observations was assessed from replicates from Finnish sites (Maaninka and Rovaniemi) from where they were available. The standard deviation of those observations (27 g m^{-2}) was relatively close to that reported in the literature for yields in field trials (e.g. 10.5 to 14.8 g m^{-2} in Gustavsson et al. (2001)).

408 We did not apply any weighting of the data even in the generic calibration, which means that sites with

409 more observations had more weight than sites with fewer treatments and observations. In our case this

410 meant, for example, that the Saerheim site in Norway got more weight in generic calibrations than other

411 sites. However, a site by site analysis showed that no correlation was observed between model

412 performance with generic calibration and the amount of observed data from the site used for calibration

413 (Table S5). This result indicates that the simple approach used without weighting was reasonable.

414 The underestimation of the first cut yield and slight overestimation of the second cut yield may indicate

415 that the growth descriptions of the models and the parameters related to them are too stiff when

416 calibrated simultaneously using data from both the spring growth and the summer period. This might

417 reflect a need for improved understanding of the processes determining timothy growth during different

418 stages of the growing season. While using the current models for seasons with two or more cuts, it can

419 also be hypothesized that calibrating the models with data including third and fourth cuts could lead to

420 even higher compromise between the cuts.

Perspectives on model improvements

422 The STICS model gave the most accurate overall yield estimates over all treatments and this suggests

423 that a detailed description of soil processes improves yield simulations over variable sites and growing

424 seasons. STICS also better captured the effects of different N fertilisation rates on the first cut yield

425 (Fig. 6), although the CATIMO model uses the same approach for simulating N effects. This result

426 could be a consequence of a better simulation of soil N mineralisation and water balance in STICS.

427 Both of these processes affect the N availability for plants. Generally, the simulation of N cycling,

428 including residual N and mineralization, as well as N availability remains challenging for simulating

429 the growth of perennial forage grasses (Ehrhardt et al. 2018).

430 Our analysis only concentrated on the spring growth and the first summer growth. More complete

431 annual yield estimates would also require simulations of a second summer growth on sites and years

432 where more than two cuts are taken. This is particularly of relevance as the growing seasons in northern

433 latitudes are expected to get longer with changing climate (Ruosteenoja et al.2016). The proportion of

434 leaves to stems is usually very high in the autumn when a third cut would be taken and the nutritive

435 value seems to differ from that of the previous cuts.

436 Persson et al. (2014) noticed in their LINGRA model evaluation that the model performance also

437 depended on the ley year, since the establishment method affected the DM accumulation in the first

438 post-seeding ley year and these effects were not covered by the evaluated model. We observed a similar

439 trend. The first cut yield simulation of all models improved with increasing ley years at the Québec site

440 (Table S5). This indicates that processes affecting yield development, e.g., overwintering of plants, over

441 the life cycle of timothy swards are not yet fully covered by the evaluated models.

442 Multi-year simulations for conditions where the winter conditions often play a major role in yield

443 formation require models able to simulate winter survival. The data used in this study were for hardy

444 winter cultivars in the given regions and therefore winter survival was expected to play only a minor

445 role in yield development. For example, the winter damage sustained by the cultivar Tammisto II in

446 northernmost location of the study, Rovaniemi, was observed to be less than 6.5 % in the official variety

447 trials (Laine et al., 2016). Although the effect of winter conditions on plant density is not simulated in

448 STICS, Jing et al. (2017) showed that the model performed well in predicting harvested biomass, soil

449 moisture and soil mineral nitrogen of a timothy field for three consecutive years in Eastern Canada. Of

450 the models used in this study, the BASGRA model is able to predict the winter survival of timothy

451 grass, but the winter module was not utilised in this study.

452 In addition to the yields, which were the focus of this study, nutritive value of forage crops is of

453 importance for an integrated assessment of the effect of climate change on farming systems. Our next

454 step will be to compare the performance of the three models for simulating the nutritive attributes of

455 timothy grass.

456 CONCLUSIONS

457 A comparison of simulation results for timothy growth in this study, using comprehensive data and

458 advanced calibration methods, revealed substantial differences in the responses of the compared models

to interannual climate and agro-management. Cultivar-specific calibration that was mostly done at site level yielded more accurate estimates of the first cut yields and preceding growth dynamics than the generic calibration applying data from all sites. Still, there was a high degree of variation among the simulated yield results and during some seasons at some sites, all three models systematically under- or overestimated the growth. This suggests that the current process descriptions in the models were not flexible enough to allow for the observed range of growth dynamics and related variables. There are obviously still needs for improvement in understanding the processes and their model descriptions, particularly regarding the summer growth and responses to nitrogen applications. These improvements are necessary to achieve reliable model applications across a wide range of current and future conditions.

ACKNOWLEDGEMENTS

REFERENCES

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D., Kimball,B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D.,Prasad, P.V.V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J.,De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G.,Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D.,

484   Kersebaum, K.C., Koehler,A.K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen,

485   J.E., Palosuo,T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov, M.A., Shcherbak,

486   I.,Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K.,Wang, E.,

487   Wallach, D., Wolf, J., Zhao, Z., Zhu, Y., 2015. Rising temperatures reduce global wheat

488   production. Nat. Clim. Change. 5, 143–147.

489 Asseng S., Ewert, F., Rosenzweig C., Jones J.W. Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn,

490   P. J., Rötter, R. P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P. K., Angulo,

491   C., Bertuzzi, P., Biernath, C., Challinor, A. J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng,

492   L., Hooker, J., Hunt, L. A., Ingwersen, J., Izaurralde, R. C., Kersebaum, K. C., Müller, C., Naresh

493   Kumar, S., Nendel, C., O'Leary, G., Olesen, J. E., Osborne, T. M., Palosuo, T., Priesack, E.,

494   Ripoche, D., Semenov, M. A., Shcherback, I., Steduto, P., Stöckle, C., Stratonovich, P., Streck,

495   T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J. W., Williams, J. R., Wolf, J.,

496   2013. Uncertainty in simulating wheat yields under climate change. Nat. Clim. Change 3 (9), 827-

497   832.

498 Bassu, S. et al., 2014. How do various maize crop models vary in their responses to climate

499   change factors? Glob. Change Biol., 20(7): 2301-2320.

500 Bélanger, G., Tremblay, G. F., and Mongrain, D., 2008. Yield and nutritive value of the spring growth

501   of an ageing timothy sward. Can. J. Plant Sci. 88(3), 457-464.

502 Bélanger, G. and Richards, J. E., 1997. Growth analysis of timothy grown with varying N nutrition.

503   Can. J. Plant Sci. 77(3), 373–380.

504 Bonesmo, H., and Bélanger, G., 2002a. Timothy yield and nutritive value by the CATIMO model. I.

505   Growth and nitrogen. Agron. J. 94(2), 337-345.

506 Bonesmo, H., and Bélanger, G., 2002b. Timothy yield and nutritive value by the CATIMO model. II.

507   Digestibility and fiber. Agron. J. 94(2), 345-350.

508   Bonesmo, H., Bélanger, G., Charmley, E., Drapeau, R., McKenzie, D. B., Michaud, R., Tremblay, G.
509   F., 2005. Timothy Yield and Nutritive Value by the CATIMO Model. III. Validation for Eastern
510   Canada. Agron. J. 97(1), 32–40.

511   Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-
512   Barret, F., Antonioletti, R., Durr, C., Richard, G., Beaudoin, N., Recous, S., Tayot, X., Plenet, D.,
513   Cellier, P., Machet, J.-M., Meynard, J.M., Delécolle, R., 1998. STICS: A generic model for the
514   simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied
515   to wheat and corn. Agronomie 18, 311–346.

516   Brisson, N., Launay, M., Mary, B., Beaudoin, N., 2008. Conceptual basis, formalisations and
517   parameterization of the STICS crop model. Éditions QUAE, Versailles, France.

518   Chenu K., Porter J.R., Martre P., Basso B., Chapman S.C., Ewert F., Bindi M., Asseng S. 2017.
519   Contribution of crop models to adaptation in wheat. Trends in Plant Science 22, 472–490.

520   Ehrhardt, F., J.-F. Soussana, G. Bellocchi, P. Grace, R. McAuliffe, S. Recous, R. Sándor, P. Smith, V.
521   Snow, M. De Antoni Migliorati, B. Basso, A. Bhatia, L. Brilli, J. Doltra, C. D. Dorich, N. Sciences,
522   L. Doro, N. Fitton, E. Sciences, S. Giacomini, B. Grant, M. Harrison, S. Jones, M. Kirschbaum,
523   K. Klumpp, P. Laville, J. leonard, M. Liebig, M. Lieffering, R. Martin, R. S. Massad, E. Meier, L.
524   Merbold, A. Moore, V. Myrgiotis, P. Newton, E. Pattey, S. Rolinski, J. Sharp, W. Smith, L. Wu
525   and Q. Zhang, 2018: Assessing uncertainties in crop and pasture ensemble model simulations of
526   productivity and N2O emissions. Global Change Biology 24: 603-616.

527   Gustavsson, A.-M., and Martinsson, K., 2001. Analysis of growth and nutrition value in timothy using
528   a dynamic model. Agr. Forest. Meteorol. 107(2), 83–101.

529   Gustavsson, A-M. and Martinsson, K., 2004. Seasonal variation in biochemical composition of cell
530   walls, digestibility, morphology, growth and phenology in timothy. Eur. J. Agron. 20(3), 293-312.

531   Höglind, M., Schapendonk, A., Van Oijen, M., 2001. Timothy growth in Scandinavia: Combining

532   quantitative information and simulation modelling. New Phytol. 151(2), 355-367.

533   Höglind, M., Hanslin, H.M., Van Oijen, M., 2005. Timothy regrowth, tillering and leaf area dynamics

534   following spring harvest at two growth stages. Field Crops Res. 93(1), 51-63.

535   Höglind, M., Van Oijen, M., Cameron, D., and Persson, T. (2016). Process-based simulation of growth

536   and overwintering of grassland using the BASGRA model. Ecol. Model. 335, 1-15.

537   Hurtado-Uria, C., Hennessy, D., Shalloo, L., Schulte, R.P.O., Delaby, L., O'Connor, D., 2013.

538   Evaluation of three grass growth models to predict grass growth in Ireland. J. Agric. Sci. 151, 91–

539   104.

540   Jégo, G., Bélanger, G., Tremblay, G.F., Jing, Q., Baron, V.S., 2013. Calibration and performance

541   evaluation of the STICS crop model for simulating timothy growth and nutritive value. Field Crops

542   Res. 151, 65–77.

543   Jing, Q., Bélanger, G., Baron, V., Bonesmo, H., 2011. Modelling the biomass and harvest index

544   dynamics of timothy. Agron. J. 103, 1397-1404.

545   Jing, Q., Bélanger, G., Baron, V., Bonesmo, H., Virkajärvi, P., 2013. Simulating the nutritive value of

546   timothy summer regrowth. Agron. J. 105(3), 563-572.

547   Jing, Q., Bélanger, G., Baron, V., Bonesmo, H., Virkajärvi, P., Young, D., 2012. Regrowth simulation

548   of the perennial grass timothy. Ecol. Model. 232, 64-77.

549   Jing, Q., Jégo, G., Bélanger, G., Chantigny, M.H., Rochette, P., 2017. Simulation of water and nitrogen

550   balances in a perennial forage system using the STICS model. Field Crops Res. 201, 10-18.

551   Keating, B.A, Carberry, P.S, Hammer, G.L, Probert, M.E, Robertson, M.J, Holzworth, D, Huth, N.I,

552   Hargreaves, J.N.G, Meinke, H, Hochman, Z, McLean, G, Verburg, Snow, K V, Dimes, J.P,

553   Silburn, M, Wang, E, Brown, S, Bristow, K.L, Asseng, S, Chapman, S, McCown, R.L, Freebairn,

554   D.M, Smith, C.J, 2003. An overview of APSIM, a model designed for farming systems simulation.

555   Eur. J. Agron., 18, 267-288

556   Kipling, R. P., Bannink, A., Bellocchi, G., Dalgaard, T., Fox, N. J., Hutchings, N. J., Kjeldsen, C.,

557   Lacetera, N., Sinabell, F., Topp, C. F., Van Oijen, M., Virkajärvi, P., Scollan, N., 2016. Modeling

558   European ruminant production systems: Facing the challenges of climate change. Agric. Syst. 147,

559   24-37.

560   Kollas C., Kersebaum K.C., Nendel C., Manevski K., Müller C., Palosuo T., Armas-Herrera, C.M.,

561   Beaudoin, N., Bindi, M., Charfeddine M., Conradt, T., Constantin J., Eitzinger J., Ewert F, Ferrise

562   R., Gaiser T., Garcia de Contazar-Atauri, Giglio, L., Hlavinka P., Hoffman H., Launay M.,

563   Manderscheid R., MaryB., Mirchel W., Moriondo, M., Olesen J.E., Öztürk I., Pacholski A.,

564   Ripoche-Wachter D., Roggero P.P., Roncossek, S., Rötter R.P., Ruget, F., Sharif B., Trnka, M.,

565   Ventrella D., Waha, K., Wegehenkel, M., Weigel, H.J., Wu, L., 2015. Crop rotation modelling—

566   A European model intercomparison. Eur. J. Agron. 70, 98-111.

567   Laine, A., Högnäsbacka, M., Kujala, M., Niskanen, M., Jauhiainen, L., Nikander, H., 2016. Results of

568   the Official Variety Trials 2008-2015. (In Finnish) Luonnonvarakeskus (Luke). Luonnonvara- ja

569   biotalouden tutkimus 3/2016. 239p. http://jukuri.luke.fi/bitstream/handle/10024/531720/luke-

570   luobio_3_2016.pdf

571   Nelder, J.A., and Mead, R., 1965. A simplex method for function minimization. Comput. J. 7(4), 308–

572   313.

573   Nissinen, O., Kalliainen, P., Jauhiainen, L., 2010. Development of yield and nutritive value of timothy

574   in primary growth and regrowth in northern growing conditions. Agr. Food Sci. 19, 252–268.

575   Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.E., Patil, R.H., Ruget,

576   F., Rumbaur, C., Takac, J., Trnka, M., Bindi, M., Caldag, B., Ewert, F., Ferrise, R., Mirschel, W.,

577   Saylan, L., Siska, B., Rötter, R., 2011. Simulation of winter wheat yield and its variability in

578   different climates of Europe: a comparison of eight crop growth models. Eur. J. Agron. 35, 103–

579   114.

580   Persson, T., Höglind, M., Gustavsson, A., Halling, M., Jauhiainen, L., Niemeläinen, O., Thorvaldsson,

581   G., Virkajärvi, P., 2014. Evaluation of the LINGRA timothy model under Nordic conditions. Field

582   Crops Res. 161, 87-97.

583   Pirttioja, N., Carter, T.R., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F.,

584   Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D.,

585   Deligios, P., Destain, M.F., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka,

586   P., Jacquemin, I., Kersebaum, K.C., Kollas, C., Krzyszczak, J., Lorite, I.J., Minet, J., Minguez,

587   M.I., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A.,

588   Ruane, A.C., Ruget, F., Sanna, M., Semenov, M.A., Slawinski, C., Stratonovitch, P., Supit, I.,

589   Waha, K., Wang, E., Wu, L., Zhao, Z., Rötter, R.P., 2015. Temperature and precipitation effects

590   on wheat yield across a European transect: a crop model ensemble analysis using impact response

591   surfaces. Clim. Res. 65, 87–105.

592   Ruiz-Ramos, M., Ferrise, R., Rodríguez, A., Lorite, I.J., Bindi, M., Carter, T.R., Fronzek, S., Palosuo,

593   T., Pirttioja, N., Baranowski, P., Buis, B., Cammarano, D., Chen, Y., Dumont, B., Ewert, F.,

594   Gaiser, T., Hlavinka, P., Hoffmann, H., Höhn, J.G., Jurecka, F., Kersebaum, K.C., Krzyszczak, J.,

595   Lana, M., Mechiche-Alami, A., Minet, J., Montesino, M., Nendel, C., Porter, J.R.,Ruget, F.,

596   Semenov, M.A., Steinmetz, Z., Stratonovitch, P., Supit, I., Tao, F., Trnka, M., De Wit, A., Rötter,

597   R.P. 2017. Adaptation response surfaces for managing wheat under perturbed climate and CO 2 in

598   a Mediterranean environment. Agr. Syst 159: 260-274.

599   Ruosteenoja, K., Räisänen, J., Venäläinen A., Kämäräinen, M., 2016. Projections for the duration and

600   degree days of the thermal growing season in Europe derived from CMIP5 model output.

601   Int.J.Climatol. 36(8): 2029-2055.

602   Rötter, R.P., Palosuo, T., Kersebaum, K.C., Angulo, C., Bindi, M., Ewert, F., Ferrise, R., Hlavinka, P.,

603   Moriondo, M., Nendel, C., Olesen, J.E., Patil, R.H., Ruget, F., Takác, J., Trnka, M.,

604   2012. Simulation of spring barley yield in different climatic zones of Northern and Central Europe:

605   a comparison of nine crop models. Field Crops Res. 133, 23–36.

606   van Oijen, M., Höglind, M., Hanslin, H. M., Caldwell, N., 2005. Process-based modeling of timothy

607   regrowth. Agron. J. 97(5), 1295-1303.

608   Virkajärvi, P., Pakarinen, K., Hyrkäs, M., Savolainen, J., Isolahti, M., Schnyder, H., Schellberg, J.

609   (2010). Does tiller type distribution explain the differences in yield and nutritive value of timothy

610   genotypes. Grassland Sci. Eur. 15, 572-574.

611   Virkajärvi, P., Hyrkäs, M., Pakarinen, K., Rinne, M., 2012. Timotein ja ruokonadan erot

612   sadontuottoprosessissa. In: Hyrkäs, M., and Virkajärvi, P. (Eds.) Nurmen kasvu- ja

613   kehitysprosessit. MTT Report 56 (In Finnish), 22-46.

Supplementary Material

Table S1

Table S1. Treatments used for model calibration and evaluation (marked with asterisks) with their years, fertilisation regimes, and dates of first cut.

Treatment #	Location	Year	Fertilisation regime spring/after 1st cut kg N ha^{-1}	Day number for first cut	Ley year
1	Maaninka	2006	90/90	167	1st
2	Maaninka	2007	90/90	169	2nd
3	Rovaniemi	1999	80/80	179	1st
4	Rovaniemi	1999	100/100	179	1st
5*	Rovaniemi	2000	80/80	181	2nd
6*	Rovaniemi	2000	100/100	181	2nd
7	Rovaniemi	2001	80/80	177	3rd
8	Rovaniemi	2001	100/100	177	3rd
9	Særheim	2000	140/80	150	1st
10	Særheim	2000	140/80	173	1st
11*	Særheim	2001	140/80	164	1st
12*	Særheim	2001	140/80	186	1st
13	Særheim	2002	140/80	150	2nd
14	Særheim	2002	140/80	178	2nd
15	Quebec	1999	-/-	167	1st
16	Quebec	1999	60/-	167	1st
17*	Quebec	1999	120/-	167	1st
18	Quebec	2000	-/-	172	2nd
19*	Quebec	2000	60/-	172	2nd
20	Quebec	2000	120/-	172	2nd
21*	Quebec	2001	-/-	171	3rd
22	Quebec	2001	60/-	171	3rd
23	Quebec	2001	120/-	171	3rd
24	Lacombe	2004	100/-	201	1st
25	Lacombe	2005	100/-	185	2nd
26	Umeå	1995	90/90	171	2nd
27	Umeå	1996	90/90	178	3rd
28	Fredericton	1991	168/-	169	1st
29	Fredericton	1992	200/-	173	1st
30*	Fredericton	1993	-/-	173	1st
31	Fredericton	1993	70/-	173	1st
32	Fredericton	1993	140/-	173	1st
33*	Fredericton	1993	210/-	173	1st

Table S2

Table S2. BASGRA model parameters: prior ranges and maximum a posteriori (MAP) values as gained from cultivar-specific and non-cultivar specific

(generic) calibrations.

Parameter name	Description	Unit	Prior parameter range	MAP cultivar specific						MAP non-cultivar specific (generic)
				Tammisto II	Iki	Grindstad	Champ	Climax	Jonatan	
LOG10CLVI	Initial weight of leaves	log(gC m^{-2})	-0.5–2.5	1.76	0.72	1.43	0.754	1.51	1.71	1.24
LOG10CRESI	Initial weight of reserves	log(gC m^{-2})	-1.5–1.5	0.962	-0.586	1.21	1.22	0.212	0.871	0.375
LOG10CRTI	Initial weight of roots	log(gC m^{-2})	-1–2	1.42	0.993	0.538	0.103	0.983	0.897	0.214
LOG10LAII	Initial LAI	log(m^2 leaf m^{-2})	-2–1	-0.0392	-0.289	0.0325	-1.1	-0.224	0.205	-0.319
TILTOTI	Initial tiller density	m^{-2}	1000–3000	2320	1830	2410	1630	2110	2090	2390
FRTILGI	Initial fraction of elongating tillers	-	0–0.5	0.117	0.242	0.4	0.0268	0.159	0.195	0.399
CLAIV	Maximum LAI remaining after harvest, when no tillers elongate	m^2 leaf m^{-2}	0.25–1	0.438	0.656	0.371	0.645	0.651	0.515	0.569
COCRESMX	Maximum concentration of reserves in aboveground biomass	-	0.0706–0.282	0.136	0.154	0.142	0.151	0.148	0.151	0.145
CSTAVM	Maximum size of elongating tillers	gC tiller^{-1}	0.1–1.9	1.11	0.995	1.24	1.26	1.26	1.03	1.11
DAYLB	Day length below which phenological stage is reset to zero	d d^{-1}	0–0.784	0.363	0.331	0.55	0.558	0.402	0.48	0.533
DAYLP	Day length below which phenological development slows down	d d^{-1}	0.316–1	0.581	0.662	0.687	0.662	0.635	0.699	0.71
DLMXGE	Day length below which DAYLGE becomes less than 1	d d^{-1}	0–1	0.943	0.837	0.985	0.574	0.988	0.898	0.985
FSLAMIN	Minimum SLA of new leaves as a fraction of maximum possible SLA	-	0–0.933	0.47	0.471	0.797	0.711	0.62	0.534	0.862
HAGERE	Fraction of reserves in elongating tillers that is harvested	-	0.7–1	0.818	0.799	0.77	0.741	0.803	0.799	0.834
K	PAR extinction coefficient	m^2 m^{-2} leaf	0.3–0.9	0.45	0.543	0.386	0.591	0.518	0.41	0.467
LAICR	LAI above which shading induces leaf senescence	m^2 leaf m^{-2}	1.9–7.59	4.42	5.44	4.45	3.39	4.51	4.15	3.91
LAIEFT	Decrease in tillering with leaf area index	m^2 leaf m^{-2}	0.1–0.4	0.272	0.186	0.242	0.182	0.192	0.174	0.122
LAITIL	Maximum ratio of tiller and leaf appearance at low leaf area index	-	0.283–1.13	0.666	0.529	0.789	0.815	0.542	0.487	0.966
LFWIDG	Leaf width on elongating tillers	m	0.00426–170000	0.0096	0.00643	0.00952	0.0103	0.00757	0.00801	0.0104
LFWIDV	Leaf width on non-elongating tiller	m	0.00246–0.00984	0.00603	0.00389	0.00683	0.00375	0.00433	0.0042	0.007
NELLVM	Number of growing leaves per non-elongating tiller	tiller^{-1}	1.05–3.5	2.37	1.24	2.13	1.64	2.1	1.81	2.17

PHENCR	Phenological stage above which elongation and appearance of leaves on elongating tillers decreases	-	0.248–0.99	0.523	0.53	0.319	0.614	0.431	0.474	0.499
PHY	Phyllochron	°C d	31.5–126	54.3	59.2	46.4	48.9	63.6	61.1	45.1
RDRSCO	Relative death rate of leaves and non-elongating tillers due to shading	d^{-1}	0.0356–1420000	0.0797	0.0964	0.0888	0.0942	0.0612	0.0618	0.0956
RDRSMX	Max. rel. death rate of leaves and non-elongating tillers due to shading	d^{-1}	0.03–0.12	0.0738	0.0583	0.0629	0.064	0.0522	0.0465	0.0634
RDRTEM	Proportionality of leaf senescence with temperature	d^{-1} °C^{-1}	0.000513–0.00205	0.00114	0.0014	0.00119	0.000795	0.00115	0.00109	0.000937
RGENMX	Maximum relative rate of tillers becoming elongating tillers	d^{-1}	0.00544–218000	0.0105	0.0117	0.0194	0.0166	0.011	0.0105	0.0172
RRDMAX	Maximum root depth growth rate	m d^{-1}	0.006–0.024	0.0132	0.0102	0.0088	0.0156	0.0114	0.0121	0.0155
RUBISC	Rubisco content of upper leaves	g m^{-2} leaf	2.89–11.6	6.36	4.43	4.82	5.39	4.41	5.27	4.55
SHAPE	Area of a leaf relative to a rectangle of same length and width	-	0.269–1	0.585	0.456	0.561	0.602	0.563	0.544	0.844
SIMAX1T	Sink strength of small elongating tillers	gC tiller^{-1} d^{-1}	0.00225–0.00901	0.00505	0.00666	0.0055	0.00683	0.00312	0.00504	0.00547
SLAMAX	Maximum SLA of new leaves	m^2 leaf gC^{-1}	0.03–0.09	0.0658	0.0748	0.0676	0.0724	0.0574	0.0552	0.0781
TBASE	Minimum value of effective temperature for leaf elongation	°C	1.81–6	3.24	3.98	4.17	4.51	4.26	3.89	2.77
TCRES	Time constant of mobilisation of reserves	d	0.945–3.78	2.21	1.55	2.37	1.45	2.2	1.58	2.64
TOPTGE	Opt. temperature for vegetative tillers to become generative	°C	7–25.2	4.45	12.1	13	8.8	9.47	14.7	21.4
TRANCO	Transpiration coefficient	mm d^{-1}	4–16	9.04	12.1	5.81	6.73	6.83	6.62	6.44
YG	Growth yield per unit expended carbohydrate	gC gC^{-1}	0.65–0.9	0.866	0.863	0.836	0.864	0.821	0.836	0.876
Dparam	Constant in the calculation of de-hardening rate	°C^{-1} d^{-1}	0.00125–0.00375	0.00307	0.00267	0.00291	0.00324	0.00269	0.00306	0.0019
FGAS	Soil pore space	m^3 m^{-3}	0.2–0.6	0.394	0.465	0.347	0.333	0.456	0.399	0.418
FO2MX	Maximum oxygen fraction of soil gas	mol O$_2$ mol^{-1} gas	0.105–0.315	0.24	0.174	0.233	0.194	0.242	0.206	0.201
gamma	Temperature extinction coeff. snow	m^{-1}	32.5–97.5	66.8	71.5	59.9	69.5	60.2	66.8	68.9
Hparam	Hardening parameter	°C^{-1} d^{-1}	0.004–0.012	0.00524	0.00714	0.0078	0.00625	0.00553	0.00453	0.00826
KRDRANAER	Maximum relative death rate due to anaerobic conditions	d^{-1}	0–1	0.157	0.264	0.227	0.307	0.268	0.181	0.64
KRESPHARD	Carbohydrate requirement of hardening	gC gC^{-1} °C^{-1}	0.001–0.05	0.0178	0.00523	0.012	0.0234	0.00485	0.00733	0.0224
KRSR3H	Constant in the logistic curve for frost survival	°C^{-1}	0–1	0.896	0.702	0.986	0.739	0.973	0.969	0.962
KRTOTAER	Ratio of total to aerobic respiration	-	1–2	2.04	1.69	1.98	2.34	2.36	1.69	1.78
KSNOW	Light extinction coefficient of snow	mm^{-1}	0.0175–0.0525	0.0373	0.0263	0.0413	0.0327	0.0304	0.0379	0.0387
LAMBDAsoil		J m^{-1} °C^{-1} d^{-1}	86400–259000	171000	162000	128000	167000	153000	185000	199000
LDT50A	Intercept of linear dependence of LD50 on LT50	d	0.67–2.01	1.35	1.24	0.946	1.35	1.59	1.33	1.56
LDT50B	Slope of linear dependence of LD50 on LT50	d °C^{-1}	-3.2–(-1.1)	-2.16	-2.51	-1.98	-1.57	-2.42	-2.19	-2.17

Parameter	Description	Units	Range							
LT50MN	Minimum LT50	°C	-40–20	-25.9	-25.1	-24.9	-25.2	-30.2	29	-25.5
LT50MX	Maximum LT50	°C	-7–(-3)	-4.88	-5.13	-5.85	-5.03	-5.1	-5.3	-4.54
RATEDMX	Maximum de-hardening rate	°C d^{-1}	1–3	1.7	1.73	1.82	2.37	1.8	1.71	1.81
reHardRedDay	Duration of period over which re-hardening capability disappears	d	5–160	144	138	135	133	150	139	134
RHOnewSnow	Density of newly fallen snow	kg SWE m^{-3}	50–150	97.3	102	89.4	91.6	105	109	85.9
RHOpack	Relative packing rate of snow	d^{-1}	0.01–0.03	0.0197	0.0212	0.0205	0.0137	0.0185	0.0194	0.0219
SWret	Liq. water storage capacity of snow	mm mm^{-1}d^{-1}	0.05–0.15	0.11	0.0928	0.104	0.0983	0.103	0.0977	0.0842
SWrf	Maximum refreezing rate per degree below 'TmeltFreeze'	mm d^{-1} °C^{-1}	0.005–0.015	0.00782	0.0121	0.00856	0.00673	0.00914	0.0101	0.0104
THARDMX	Maximum surface temperature at which hardening is possible	°C	5–15	14.7	13.2	13.2	14.2	14.3	14.5	14.8
TrainSnow	Temperature below which precipitation is snow	°C	0.005–0.015	0.0103	0.01	0.00938	0.00906	0.00917	0.0102	0.0106
TsurfDiff	Constant in the calculation of de-hardening rate	°C	0–10	0.192	2.32	1.19	0.328	1.82	0.675	3.35
KLUETILG	LUE-increase with increasing fraction elongating tillers	-	0–1	0.36	0.46	0.12	0.317	0.391	0.653	0.579
ROOTDM	Initial and max. value rooting depth	m	0–1.52	0.866	0.559	0.626	1.14	0.484	0.788	1.18
FWCWP	Water conc. at wilting point	m^3 m^{-3}	0.02–0.3	0.0456	0.159	0.063	0.0517	0.0837	0.0698	0.0774
FWCFC	Water conc. at field capacity	m^3 m^{-3}	0.3–0.9	0.813	0.797	0.734	0.619	0.63	0.648	0.65
FRTILGG1I	Initial fraction of non-elongating generative tillers	-	0–1	0.131	0.446	0.127	0.255	0.0402	0.142	0.565
DAYLG1G2	Day length above witch non-elongating generative tillers starts to become elongating tillers	d d^{-1}	0–1	0.683	0.879	0.551	0.568	0.688	0.697	0.63
RGRTG1G2	Relative rate of non-elongating tillers becoming elongating tillers	d^{-1}	0–1	0.907	0.955	0.64	0.505	0.899	0.85	0.824
RDRTMIN	Minimum root death rate	d^{-1}	0–0.02	0.00969	0.00705	0.0129	0.00854	0.00881	0.00961	0.0121
NCSHMAX	Maximum N-C ratio of shoots	g N g^{-1} C	0.02–0.08	0.0412	0.0578	0.0657	0.0326	0.0445	0.0404	0.0693
NCR	N-C ratio of roots	g N g^{-1} C	0.015–0.06	0.0364	0.0217	0.0238	0.0214	0.0245	0.0247	0.022
RDRSTUB	Relative death rate of stubble	d^{-1}	0.1–0.4	0.196	0.217	0.22	0.192	0.242	0.205	0.209
FNCGSHMIN	Shoot N-sink at N-deficiency	-	0–0.5	0.13	0.0554	0.173	0.0474	0.166	0.126	0.327
TCNSHMOB	Time constant for remobilisation of shoot N	d	1–64	4.58	16.6	9.1	6.6	5.66	14.8	7.64
TCNUPT	Time constant of soil mineral N supply	d	1–64	10.1	28.7	24.9	11.8	7.01	17.6	27.6

Table S3

Table S3. CATIMO model parameters: prior ranges and maximum a posteriori (MAP) values gained from cultivar-specific and non-cultivar specific (generic) calibrations.

Parameter name	Description	Unit	Prior parameter range	MAP cultivar specific						MAP non-cultivar specific (generic)
				Tammisto II	Iki	Grindstad	Champ	Climax	Jonatan	
Fropt	Partitioning fraction of biomass into roots under optimal conditions	-	0.10-0.20	0.16	0.17	0.19	0.17	0.19	0.15	0.19
Hvl0	Vertical height of leaves above shoot apex	cm	1-10	6	6	7	7	6	6	5
Initial_LAI	LAI at the beginning of simulation	m^2 leaf m^{-2} soil	0.01-0.25	0.12	0.04	0.07	0.23	0.16	0.11	0.06
RUEpot	Potential radiation use efficiency	g DM MJ^{-1} PAR	2.0-5.0	3.4	2.4	3.4	3.1	2.3	3.4	2.5
SLAadd	Maximum increase of specific leaf area of new leaves due to temperature	m^2 leaf g^{-1} leaf DM	0.00-0.040	0.006	0.020	0.002	0.011	0.002	0.022	0.008
To	Optimum temperature for radiation use efficiency	°C	11-17	12	16	12	12	14	14	12
Tsumcrib	Thermal time for bud burst	°C d	0-50	28	13	26	26	45	13	18
Tbase.spring	Base temperature for start of growing season	°C	0-10	3.5	1.2	2.0	6.6	9.7	2.3	1.9
Rooting.depth	Rooting depth.	mm	500-1500	1242	1020	705	842	688	570	*
soil_N_init	Amount of N in soil at the beginning of simulation period.	kg ha^{-1}	1-20	5	9	4	6	2	4	*
soil_water_init	Amount of water in soil in the beginning of simulation period	cm^3 cm^{-3}	100-400	203	130	129	188	211	195	*
MINmax	Maximum N mineralization rate	g N m^{-2} d^{-1}	0.200-0.500	0.369	0.355	0.290	0.263	0.221	0.374	*
FNAmax	Maximum fraction of available N in soil	-	0.100-1.000	0.351	0.325	0.208	0.188	0.621	0.293	*

*Values of cultivar-specific calibration used in generic calibration.

Table S4

Table S4. STICS model parameters gained from cultivar-specific and non-cultivar specific (generic) calibrations.

Parameter name*	Description	Unit	Prior parameter range	Cultivar specific calibration						Non-cultivar specific (generic) calibration
				Tammisto II	Iki	Grindstad	Champ	Climax	Jonatan	
dlaimaxbrut	maximum rate of daily increase of leaf area index	m^2 leaf plant^{-1} degree-day^{-1}	2.10^{-4}-4.10^{-4}	0.00032	0.00032	0.00032	0.00032	0.00032	0.00032	0.00032
slamin	minimum specific leaf area of green leaves	cm^2 g^{-1} DM	200-300	246	246	246	246	246	246	246
slamax	maximum specific leaf area of green leaves	cm^2 g^{-1} DM	300-500	377	377	377	377	377	377	377
durvief	maximum lifespan of an adult leaf	Q10	0-100	51	36	40	46	36	26	45
stamflax	cumulative growing degree-days between the maximum acceleration of leaf growth and the maximum LAI	Degree-days	200-1000	676	467	994	510	916	991	821
stlevamf	cumulative growing degree-days between beginning of growth and maximum acceleration of leaf growth	Degree-days	10-300	10	290	159	197	110	127	148
teopt	beginning of the thermal optimum plateau for net photosynthesis	°C	8-15	10	10	10	10	10	10	10
teoptbis	end of the thermal optimum plateau for net photosynthesis	°C	12-20	15	15	15	15	15	15	15
temin	minimum threshold temperature for net photosynthesis	°C	0-5	1	1	1	1	1	1	1
temax	maximum threshold temperature for net photosynthesis	°C	20-40	20	20	20	20	20	20	20
efcroijuv	maximum radiation use efficiency during the juvenile phase	g DM MJ^{-1}	1.5-3	2.6	2.6	2.6	2.6	2.6	2.6	2.6
efcroiveg	maximum radiation use efficiency during the vegetative phase	g DM MJ^{-1}	1.5-3	2.7	2.7	2.7	2.7	2.7	2.7	2.7
efcroirepro	maximum radiation use efficiency during the reproductive phase	g DM MJ^{-1}	2-5	4.5	4.5	4.5	4.5	4.5	4.5	4.5
croirac	growth rate of the root front	cm degree^{-1} day^{-1}	0-0.5	0.043	0.086	0.026	0.048	0.117	0.06	0.027
extin	extinction coefficient of photosynthetic active radiation	-	0.1-1.5	0.67	0.67	0.67	0.67	0.67	0.67	0.67
adens	interplant competition parameter	-	-2-0	-0.5	-0.81	-0.7	-0.55	-0.58	-1.83	-0.5

*Cultivar-specific parameters in bold

Table S5

Table S5. Root mean squared errors (RMSE) of the simulated above-ground biomass (g DM m^{-2}) from the calibration and model evaluation (marked with asterisks) of three timothy models during the spring and summer growth for all 33 treatments.

Location	Treatment	BASGRA				CATIMO				STICS			
		cultivar-specific calibration		generic		cultivar-specific		generic		cultivar-specific		generic	
		Spring growth	Summer growth	Spring growth	Summer growth	Spring growth	Summer growth	Spring growth	Summer growth	Spring growth	Summer growth	Spring growth	Summer growth
Maaninka (Finland)	1	28	85	175	49	235	147	338	224	78	32	180	37
	2	94	30	107	30	134	62	273	83	31	24	192	22
Rovaniemi (Finland)	3	44	17	61	47	61	28	58	46	36	39	90	57
	4	33	31	64	22	82	48	68	31	51	42	68	46
	5*	175	57	189	104	164	108	100	82	78	59	250	48
	6*	116	51	113	40	151	114	83	87	52	48	272	45
	7	158	10	142	37	122	78	92	118	64	81	82	192
	8	131	45	100	88	121	69	92	94	64	92	84	182
Særheim (Norway)	9	109	208	133	246	86	192	197	157	88	223	132	174
	10	257	288	295	334	235	257	399	235	156	314	264	250
	11*	286	260	294	276	114	193	214	140	53	231	53	231
	12*	474	102	493	123	256	90	412	53	187	61	187	61
	13	164	67	168	99	168	167	156	84	205	191	174	127
	14	74	134	65	169	34	32	163	92	132	82	45	156
Québec (Canada)	15	154	32	416	22	175	280	195	156	129	102	117	163
	16	150	6	317	17	83	335	151	200	62	195	84	212
	17*	254	83	250	5	174	262	246	122	172	132	189	154
	18	24	56	256	15	19	326	38	199	69	14	89	3
	19*	99	4	333	48	98	297	133	166	50	23	88	11
	20	14	121	225	74	153	198	212	56	89	77	139	105
	21*	73	23	240	52	82	403	60	274	81	13	133	31
	22	72	60	264	75	87	459	86	314	72	95	103	103
	23	92	3	155	70	46	422	131	270	62	61	115	78
Lacombe (Canada)	24	125	34	61	50	74	87	155	124	62	38	86	28
	25	67	95	104	121	87	98	232	119	83	56	134	76
Umeå (Sweden)	26	62	37	214	83	34	157	168	184	111	84	72	114
	27	68	33	99	62	25	223	104	127	74	62	119	102
Fredericton (Canada)	28	52	NA	133	NA	72	NA	149	NA	58	NA	54	NA
	29	110	NA	86	NA	301	NA	321	NA	70	NA	97	NA
	30*	244	NA	47	NA	121	NA	168	NA	83	NA	83	NA
	31	25	NA	111	NA	29	NA	106	NA	97	NA	137	NA
	32	56	NA	76	NA	53	NA	116	NA	85	NA	124	NA
	33*	65	NA	81	NA	59	NA	120	NA	91	NA	123	NA

Figure S1

Figure S1. Observed and simulated above-ground biomass dynamics for all sites and treatments. Evaluation treatments are marked with an asterisk. BASGRA (Blue), CATIMO (Red) and STICS (Grey). Simulation results with cultivar-specific calibrations are shown with solid lines and results with generic calibrations with dashed lines. Observations are marked with black dots with the standard deviation as error bars.
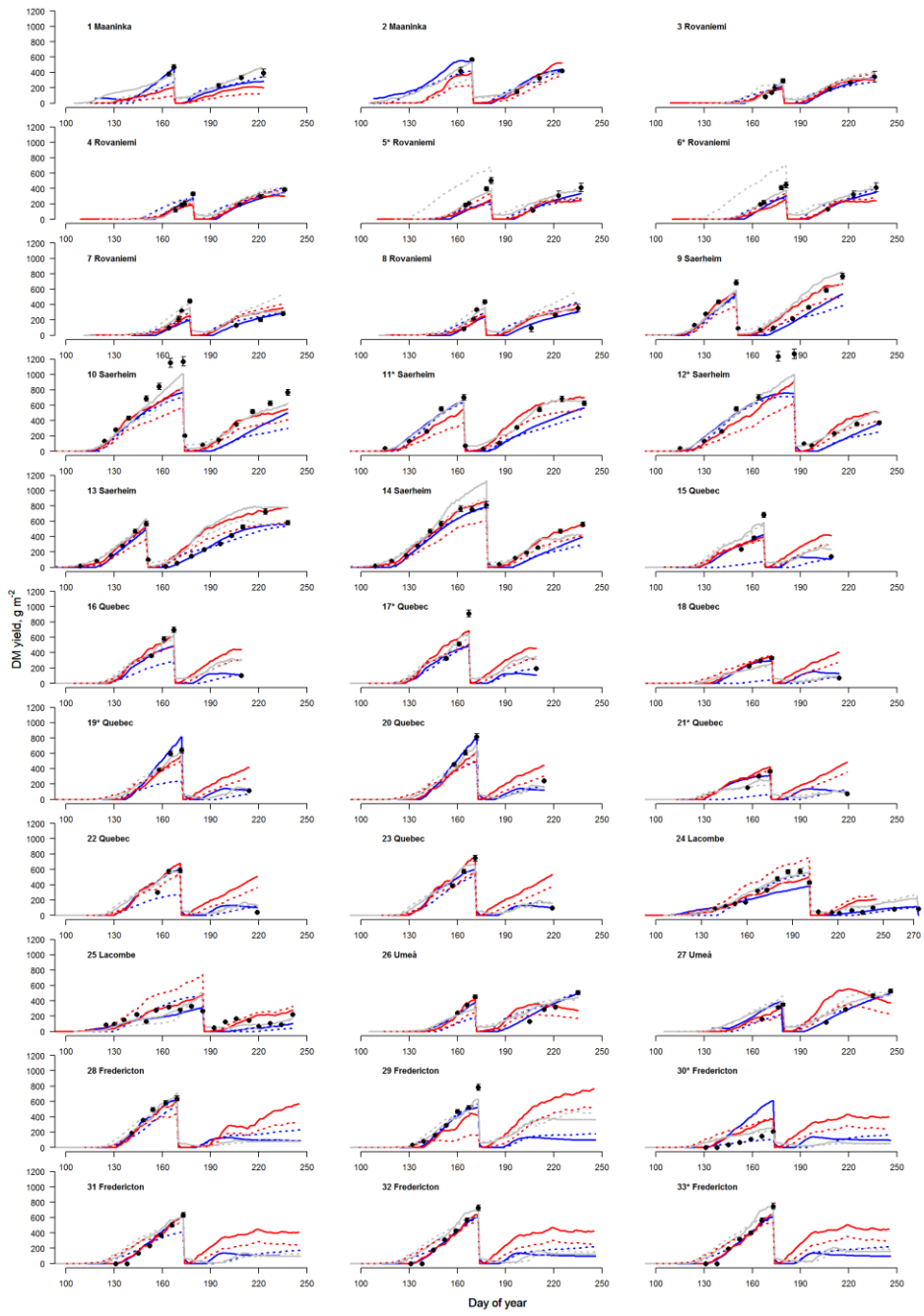
Figure S2. Leaf area index (LAI) dynamics (cultivar-specific calibration) for all sites and treatments. Evaluation treatments are marked with an asterisk. BASGRA (Blue), CATIMO (Red) and STICS (Grey). Observations are marked with black dots with the standard deviation as error bars.