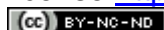


Article (refereed) - postprint

Persson, T.; Höglind, M.; Van Oijen, M.; Korhonen, P.; Palosuo, T.; Jégo, G.; Virkajärvi, P.; Bélanger, G.; Gustavsson, A.-M. 2019. **Simulation of timothy nutritive value: a comparison of three process-based models.**

© 2018 Elsevier B.V.

This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>



This version available <http://nora.nerc.ac.uk/521729/>

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at <http://nora.nerc.ac.uk/policies.html#access>

NOTICE: this is the authors' version of a work that was accepted for publication in *Field Crops Research*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Field Crops Research* (2019), 231. 81-92. <https://doi.org/10.1016/j.fcr.2018.11.008>

www.elsevier.com/

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

1 **[This paper has been published in: Field Crops Research 231 (2019): 81-92]**

2

3 **Simulation of timothy nutritive value: A comparison of three process-based models**

4 Persson, T.^{1*}, Höglind, M.¹, Van Oijen, M.², Korhonen P.³, Palosuo, T.³, Jégo, G.⁴, Virkajärvi, P.³,
5 Bélanger, G.⁴, Gustavsson, A.-M.⁵

6 ¹Norwegian Institute of Bioeconomy Research (NIBIO), Særheim, Norway

7 ²Centre for Ecology and Hydrology, UK

8 ³Natural Resources Institute Finland (Luke), Helsinki, Finland

9 ⁴Agriculture and Agri-Food Canada (AAFC), Québec, Canada

10 ⁵Swedish University of Agricultural Sciences (SLU), Umeå, Sweden

11 *Corresponding author

12 E-mail: tomas.persson@nibio.no

13

14

15

16

17

18 **Abstract**

19 Different forage grass models are used to simulate forage yield and nutritive attributes, but
20 these models are seldom compared, particularly those for timothy (*Phleum pratense* L.), a
21 widely grown forage grass species in agricultural regions with a cold temperate climate. We
22 compared the models BASGRA, CATIMO and STICS for their predictions of timothy crude
23 protein (CP) concentration, neutral detergent fibre (NDF) concentration and NDF digestibility
24 (dNDF), three important forage nutritive attributes. Data on CP and NDF concentrations, and
25 dNDF and the associated weather and soil data for seven cultivars, taken from eight field
26 experiments in Canada, Finland, Norway, and Sweden, were divided into calibration and
27 validation datasets. Model parameters were estimated for each cultivar separately (cultivar-
28 specific calibration) and for all cultivars together (generic calibration), using different methods
29 for the three models. Normalized root mean square error (RMSE) in prediction of CP
30 concentration varied between 16 and 26 % for BASGRA, 45 and 101 % for CATIMO and 27 and
31 45 % for STICS across the two calibration methods and the calibration and validation datasets.
32 Normalised RMSE in prediction of NDF concentration varied between 8 and 13 % for BASGRA,
33 14 and 21 % for CATIMO and 8 and 12 % for STICS, while for dNDF it varied between 7 and 22
34 % for BASGRA, 7 and 38 % for CATIMO and 5 and 6 % for STICS. Cultivar-specific calibration
35 improved the performance of CATIMO and STICS, but not BASGRA, compared with generic
36 calibration. The prediction accuracy for NDF concentration and dNDF with the three models
37 was within the same range or better than that for forage dry matter (DM) yield of timothy.
38 Overall, the three models performed well in predicting some nutritive attributes and yield in
39 Northern Europe and Canada, but improvements are required, particularly to increase the
40 prediction accuracy of CP concentration.

41 **Key words:** BASGRA, CATIMO, crude protein, fibre, forage grass, grassland, NDF, dNDF, STICS

42

43 1. Introduction

44 Forage grasses serve as the main source of energy and nutrients for ruminant livestock,
45 including dairy cows, beef cattle, sheep and goats, in many regions of the world (Fulkerson et
46 al., 2007; Thornton, 2010). Hence, management for optimal energy and nutrient content in
47 forage crops is crucial to these animals. Feed evaluation for ruminants usually takes into
48 account the digestibility and protein concentration of the forage (Bruinenberg et al., 2002).
49 Because the cell contents are almost totally digestible, the concentration of cell walls is crucial
50 to the nutritive value of forages (Buxton, 1996). Typically measured indicators of forage
51 nutritive value are the concentration and digestibility of neutral detergent fibre (NDF) and the
52 crude protein (CP) concentration. The development of stem and inflorescence on
53 reproductive tillers generally lowers the nutritive value of the forage, because these plant
54 parts are less digestible than leaves (Chapman et al., 2014; Elgersma and Sørensen, 2018).
55 However, as the forage grass sward grows and develops more reproductive tillers, the total
56 aboveground biomass also increases causing a negative relationship between nutritive value
57 and dry matter yield (Wilkinson and Rinne, 2018).

58 In Northern Europe and Canada, perennial forage grasses grown for intensive dairy production
59 are usually cut and harvested 2-3 times per growing season and conserved as silage (Höglind
60 et al., 2005; Casler and Kallenbach, 2007; Jing et al., 2012). Timothy (*Phleum pratense* L.) is
61 one of the most widely used forage grass species in cold-temperate regions of the world
62 (Wilkinson and Rinne, 2018), where it is grown under a wide range of climate, soil and
63 management conditions. This species exhibits slower development, expressed as growing
64 degree-days from the start of the growing season until the start of anthesis, than many other
65 cold temperate forage grasses (Pontes et al., 2007). Comparisons with other forage grasses in
66 the same environment also show that timothy has high (Pontes et al., 2007) to intermediate
67 (Jensen et al., 2016) CP concentration and digestibility. The DM yield and nutritive value of the
68 timothy vary with growing conditions and management practices, such as cutting and
69 fertilisation regimes (Bélanger et al., 2001). In addition, the relationship between
70 development, growth and nutritive value varies between timothy cultivars (Jokela et al.,
71 2015). Length of the growing season, temperature and precipitation patterns during the
72 growing season and conditions in the previous winter are particularly important for growth,
73 yield development and management of this species.

74 Process-based simulation models for forage grass (e.g. Bonesmo and Belanger, 2002a; Wu et
75 al., 2007; Köchy, 2008; Chang et al., 2013; Jégo et al., 2013; Vital et al., 2013; Höglind et al.,
76 2016) seek to represent the physiological processes behind sward growth and development.
77 However, the representation of processes such as water and nutrient uptake, carbon (C)
78 assimilation and carbohydrate allocation and transfer between plant compartments varies
79 between models (Kipling et al., 2016; Sándor et al., 2017). Previous studies showed different
80 responses in gross primary production (Sándor et al., 2016), biomass (Hurtado-Uria et al.,
81 2013; Sándor et al., 2017; Ehrhardt et al., 2018) and N₂O emissions (Ehrhardt et al., 2018) for
82 different grassland models when compared under various environmental conditions. As for
83 timothy, Korhonen et al. (2018) compared three models for their ability to predict DM yield
84 in Northern Europe and Canada. However, to our knowledge, there are no other published
85 comparisons of the ability of forage grass models to predict nutritive value.

86 The underlying processes explaining the yield and nutritive value in forage grasses are
87 arguably more complex than those explaining only DM production. In particular, as pointed
88 out by Virkajärvi et al. (manuscript under preparation), models of forage grasses differ
89 considerably in how they handle plant processes related to plant N requirements and cell wall
90 formation and content. A comparison of the ability of forage grass models to predict nutritive
91 value in field experiments could provide knowledge about the utility of these models under
92 different weather, soil, cutting and fertiliser management conditions. Such knowledge could
93 help select prediction models for different conditions, in quantifying uncertainty in model
94 predictions under different conditions, and in identifying potential improvements in the
95 representation of physiological processes in different models.

96 In this study, the ability of three simulation models (BASGRA, CATIMO and STICS) to predict
97 three key nutritive attributes [CP concentration, NDF concentration and the digestibility of
98 NDF (dNDF)] in timothy in a wide range of environments representing the main regions where
99 timothy is grown in the northern hemisphere was compared. In addition, we tested two
100 different calibration strategies: generic and cultivar-specific.

101

102 **2. Materials and methods**

103 **2.1. Model descriptions**

104 The BASGRA, CATIMO, and STICS models simulate the growth and the development of the
105 primary growth of timothy and its first regrowth as a function of the weather, soil conditions,
106 and management practices, with a daily time step. In all three models, accumulation of
107 biomass is based on the concept of radiation use efficiency where intercepted solar radiation
108 is converted into biomass.

109

110 **2.1.1. BASGRA**

111 The Basic Grassland (BASGRA) model (Höglind et al., 2016) is a further development of the
112 LINGRA model, which was initially developed to simulate perennial ryegrass (Schapendonk et
113 al., 1998) and later adapted to timothy (Höglind et al., 2001). In BASGRA, the plant is divided
114 into stem, leaf, stubble, root and reserve compartments. The model is based on the source-
115 sink concept. The source tissue, with net export of photosynthetic assimilates, consists of
116 developed leaves, other photosynthetic tissues and carbohydrate reserves. The sink tissue,
117 with net import of photosynthetic assimilates, comprises newly developed plant parts and
118 roots. Sward development is driven by air temperature and day length. Carbohydrate reserves
119 are used for producing new leaf tissue at the start of the growing season or after defoliation
120 when there is little source tissue. Equations to simulate soil and plant N and forage nutritive
121 value have recently been developed (Höglind et al., manuscript under preparation). The soil is
122 described as one single homogeneous layer. Plant-available water in the soil is set as the
123 difference between the water content at field capacity and the water content at wilting point.
124 The soil water content is affected by infiltration, soil surface evaporation and run-off, water
125 uptake by plants and percolation of water above field capacity, simulated using the tipping-
126 bucket method. Soil C is divided into three pools with different residence times, i.e. litter, soil
127 organic matter with a fast decomposition rate and soil organic matter with a slow
128 decomposition rate. Soil N is divided into four pools: similar litter and soil organic matter pools
129 as used for C, plus a pool of mineral N. Nitrogen can flow between these pools through
130 decomposition, mineralisation and immobilisation processes, which are all affected by soil
131 temperature. Nitrogen is added to the litter pool by shoot senescence, while root senescence
132 adds N to the fast-decomposing soil organic pool. Decomposition of organic N adds to the soil
133 mineral N pool, which is depleted by leaching, emissions of nitrous oxide and nitrogen oxide,
134 and plant N uptake. Nitrogen leaching is driven by the rate of water drainage which, in turn,

135 is affected by soil hydraulic properties and infiltration, transpiration and evaporation.
136 Nitrogen emissions increase with availability of mineral N. The soil N functions are obtained
137 from the forest model BASFOR (Van Oijen et al., 2005).

138

139 Sub-optimal plant N status affects the shoot C sink strength and thus shoot growth. Tillering
140 rate also depends on the plant N status. Plant N availability is the sum of soil N that is available
141 for plant uptake and plant N that is available for remobilisation within the above-ground plant
142 parts. The latter is the amount of N above an optimal N concentration profile that follows the
143 light extinction profile from the top to the bottom of the canopy, which is allocated to growing
144 plant tissue. Consequently, the optimal N concentration decreases as more light is
145 extinguished through the canopy as it grows. The nitrogen-carbon ratio in the roots is
146 constant. The plant CP concentration is the N concentration multiplied by 6.25. The fraction
147 of cell walls in the biomass, as expressed by the NDF concentration, is allowed to differ
148 between leaves and stems, and increases with phenological stage (Bélanger and McQueen,
149 1999; Nordheim-Viken et al., 2009), but is not directly affected by temperature or N
150 concentration. The digestibility of the cell wall (dNDF) of both leaves and stems decreases with
151 phenological stage (Bélanger and McQueen, 1999; Nordheim-Viken et al., 2009). The
152 digestible fraction of the cell wall is assumed to be the same in all plant components. In
153 stubble, the cell wall fraction is set at 100%, whereas there is no cell wall fraction in the
154 reserves. The digestibility of the cell content is set at 100%.

155

156 **2.1.2. CATIMO**

157 The Canadian Timothy Model (CATIMO) was developed to simulate the growth of timothy,
158 including N processes (Bonesmo and Bélanger, 2002a) and fibre concentration and
159 digestibility (Bonesmo and Bélanger, 2002b). The model allocates biomass into green leaves,
160 stems and roots. Similarly to BASGRA, a portion of the biomass that is allocated to the above-
161 ground compartments is allocated to reserves, which is used to form new tissue after
162 defoliation or winter. The light-driven biomass growth is decreased under sub-optimal soil
163 water, plant N, photosynthetically active radiation (PAR) and temperature conditions. The
164 potential radiation use efficiency, which determines growth when there are no limiting

165 factors, is constant throughout the growing season. The soil hydraulic properties and C and N
166 content are simulated in one homogeneous layer. The N stress is estimated from an index of
167 N nutrition that is calculated as the ratio of N concentration to the critical N concentration for
168 a given biomass (Bélanger and Gastal, 2000). Plant N uptake is determined by crop demand
169 and soil N supply. The soil N supply is estimated from soil mineral N content and N
170 mineralisation. The N demand is defined as the difference between actual and maximum N
171 concentration, with the latter decreasing with increasing sward biomass using an N dilution
172 curve. The plant CP concentration is calculated by multiplying the N concentration by 6.25.
173 For simulation of forage digestibility, the sward is considered to consist of green leaves, dead
174 leaves and stems including leaf sheaths, each with their own NDF concentration and
175 digestibility. The dry matter (DM) digestibility of the sward is calculated by combining the DM
176 digestibility of green leaves, dead leaves, and stems with their respective weight. The DM
177 digestibility of the cellular content of green leaves and stems is set at $0.98 \text{ g g}^{-1} \text{ DM}$. Dead
178 leaves are assumed to have a NDF concentration of $1.0 \text{ g g}^{-1} \text{ DM}$, with a DM digestibility of
179 $0.70 \text{ g g}^{-1} \text{ DM}$. The NDF concentration of green leaves and stems is obtained by integrating the
180 proportion of the respective daily growth rates partitioned to cell wall, the daily rates of
181 conversion of cellular contents into cell wall and the daily death rate of leaves. The dNDF of
182 green leaves and stems is determined from an initial maximum value and a daily rate of
183 decrease related to daily mean temperature. Both temperature and N stress are taken into
184 account in simulation of the NDF concentration and dNDF of green leaves and stems.

185

186 **2.1.3. STICS**

187 The multidisciplinary simulator for standard crops (Simulateur multIdisciplinaire pour les
188 Cultures Standard, STICS) is a model for simulation of agricultural crops and cropping systems
189 (Brisson et al., 1998, 2008). It has an add-on module for timothy, including N and nutritive
190 value-related functions (Jégo et al., 2013). The potential radiation use efficiency, setting the
191 growth under non-limiting conditions, varies between juvenile, vegetative and reproductive
192 phenological phases. The model simulates soil water dynamics and C and N processes in a
193 multi-layer profile. Plant N demand is driven by the N dilution curve concept for calculating
194 the N requirements of the plants (Bélanger and Gastal, 2000). In the STICS model, the optimal

195 crop N uptake is described using the relationship between the critical N concentration and
196 total biomass. The critical N concentration (N_c , % N per DM unit) is defined as the lowest plant
197 N concentration required for maximum growth. As most crops can take up more N than is
198 needed for optimum growth, a maximum N concentration curve is also required in STICS, but
199 no additional biomass growth occurs for N uptake between the critical and maximum N
200 concentrations. The effective total N uptake rate is limited either by the crop N demand or by
201 the soil N availability. Plant metabolism is affected when the total N concentration is below
202 the critical concentration for a given biomass defined by the critical N concentration curve.
203 Functions to calculate NDF concentration and digestibility are from CATIMO (described briefly
204 above).

205

206 **2.2. Crop data**

207 Data on timothy from experimental sites at Fredericton (45°55'N; 66°32'W; 35 m asl),
208 Lacombe (52°28'N; 113°44'W; 860 m asl) and Québec (46°47'N; 71°07'W; 75 m asl) in Canada;
209 Maaninka (63°09'N; 27°17'E; 90 m asl), Rovaniemi (66°35'N; 26°01'E; 106 m asl) and Ruukki
210 (64°40'N; 25°06' E; 48 m asl) in Finland; Særheim (58°46'N; 5°39'E; 90 m asl) in Norway; and
211 Umeå (63°45'N; 20°17'E; 12 m asl) in Sweden were used for model calibration and validation
212 (Fig. 1). Data were from the spring growth before and during the first cut and the summer
213 regrowth after the first cut until the second cut. They covered forage dry matter (DM) yield,
214 DM yield of stems and leaves, leaf area index, tiller density, water-soluble carbohydrate
215 concentration and nutritive attributes (CP concentration, NDF concentration, dNDF, ash
216 concentration, digestible DM and digestible organic matter). Data were not available for all
217 experimental sites and years (see Table 1). The dataset was divided into combinations of sites,
218 years, cultivars and management regimes, with each unique combination called a “treatment”
219 according to Korhonen et al. (2018). In total, there were 101 treatments. Thirty-three of the
220 treatments were previously used in Korhonen et al. (2018) whereas the remaining 68
221 treatments have not been used in any other previous modelling study.

222 The methods used to measure nutritive value varied slightly between the locations. Nitrogen
223 concentration was analysed using a standard Kjeldahl method at Rovaniemi (Nissinen et al.,
224 2010) and Umeå (Gustavsson and Martinsson, 2004), near-infrared reflectance spectroscopy

225 (NIRS) at Maaninka, Ruukki and Særhiem (Marum, 1990). The NDF concentration at
226 Fredericton, Lacombe, and Québec was determined using a combination of chemical and NIRS
227 methods. At Fredericton, the NDF analyses were based on methods by Van Soest et al. (1991)
228 without using sodium sulphite, while at Lacombe and Québec the analyses were carried out
229 using an Ankom Fiber Analyzer. At Maaninka, Ruukki and Særheim, the NDF concentration
230 was analysed using NIRS and at Umeå using an ANKOM filter bag technique. The NDF
231 digestibility at Fredericton and Québec was analysed from rumen contents using a method
232 described by Van Soest et al. (1966).

233 Table 1.

234 Figure 1.

235 **2.3 Weather and soil data**

236 Daily weather data on minimum and maximum temperature, precipitation, global solar
237 radiation, wind speed and relative air humidity were obtained from weather stations near the
238 experimental sites. The data for Fredericton, Québec and Lacombe were obtained from
239 Environment Canada
240 (http://climat.meteo.gc.ca/historical_data/search_historic_data_e.html), those for
241 Maaninka, Rovaniemi and Ruukki from the Finnish Meteorological Institute, those for
242 Særheim from the Agrometeorology Norway network (<http://lmt.nibio.no/>) and those for
243 Umeå from the Swedish Meteorological and Hydrological Institute (SMHI) (www.smhi.se). The
244 soil input data comprised texture fractions, bulk density, soil organic material and pH. Soil
245 hydraulic characteristics, including water content at permanent wilting point, field capacity
246 and saturation, which are input to all three models, were either measured or estimated based
247 on available data on soil properties at each site.

248

249 **2.4. Model calibration and validation**

250 The dataset was divided into calibration and validation data by randomly selecting one
251 treatment for model evaluation from each treatment type from sites with more than two

252 treatments or years except for Rovaniemi, for which no nutritive value data were used in this
253 study (Table 2). Differences in nutritive attribute data availability between sites, geographical
254 location and differences in climate and soil conditions and management practices among sites
255 were taken into account in this division.

256 Table 2.

257 Two types of calibrations were conducted for each model. In one, parameters were calibrated
258 using data for each cultivar separately (cultivar-specific calibration). In the other, a common
259 set of parameter values representing all cultivars was obtained by using the data for all
260 cultivars together (generic calibration). The division between calibration and validation
261 datasets was the same for the two calibration types. In the two calibrations, each model was
262 calibrated using model-specific methods. BASGRA and CATIMO were calibrated using Bayesian
263 techniques (Van Oijen et al., 2005). For BASGRA, a prior probability distribution was first
264 defined for each parameter to be calibrated, which was then updated using the observed data,
265 which included nutritive value data as well as observations of biomass, and biomass-related
266 data such as leaf area index, specific leaf area and tiller density. For the BASGRA calibration,
267 beta prior distributions were used for all calibration parameters (Table S1). The prior
268 parameter distribution for most parameters was set using information from a previous
269 calibration for the cultivar Grindstad in the LINGRA model (the predecessor of BASGRA), in
270 which timothy data from the Nordic region of Europe were used (Persson et al., 2014). For
271 those parameters relating to nutritive value that were introduced into BASGRA later, the prior
272 probability distribution was set within a wide, yet plausible, range with the help of literature
273 information and preliminary calibrations. The BASGRA calibration was carried out by sampling
274 from the posterior distribution using the Metropolis algorithm and a chain length of 350 000.
275 A likelihood function by Sivia (2006), which is more robust to outliers than the Gaussian
276 distribution, was used in the calibration. For CATIMO, the prior probability distribution of
277 parameters was obtained from a previous calibration (Korhonen et al., 2018) for the same
278 cultivars as used in this study except for cv. Nuutti, for which the prior probability distribution
279 was set based on cv. Tammisto II. The posterior sampling chain length for the Metropolis
280 algorithm was 150 000 iterations for Grindstad, Champ, Climax, Jonatan and generic, 200 000
281 iterations for Nuutti and 250 000 iterations for Tammisto II and Iki. For both BASGRA and
282 CATIMO, the maximum *a posteriori* (MAP) vectors from the calibration were used to evaluate

283 the models, not the whole posterior distribution, since uncertainty quantification was not
284 within the scope of this study. In the STICS and CATIMO calibrations, only the parameters
285 involved in calculation of NDF concentration and dNDF were calibrated. For STICS, parameters
286 calibrated in Korhonen et al. (2018) were used to simulate plant growth and N uptake except
287 for cv. Nuutti, for which a new calibration was required since this cultivar was not included in
288 the previous study. The parameters of the maximum and critical N dilution curves used in this
289 study were those defined by Jégo et al. (2013). These parameters were not calibrated, because
290 in STICS they are supposed to be common to all cultivars of the same species and because it
291 is not recommended to calibrate them directly in the model. If it is considered necessary to
292 define new parameters, then this should be done in a separate study following the
293 methodology proposed by Justes et al. (1994). All parameters used to calculate NDF
294 concentration were calibrated simultaneously by minimising the sum of squared differences
295 between measured and simulated NDF values. Two optimisation algorithms available in the
296 Flexible Modelling Environment (FME) package in R were used. The two-step method was
297 used, to avoid the problem of local minima. For both steps, calibrated values were constrained
298 in a calibration range. In the first step, the pseudo algorithm, a pseudo-random search
299 algorithm, was used with the maximum number of iterations (1000). A second algorithm (L-
300 BFGS-B; constrained quasi-Newton method) was then used to refine the calibration. The dNDF
301 parameters were calibrated using the same method.

302

303 2.5. Statistical analysis

304 The prediction accuracy of the observed CP concentration, NDF concentration and dNDF was
305 evaluated with the root mean square error (RMSE):

$$306 \quad RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (1)$$

307 where n is the number of observations and P_i and O_i are the predicted and observed values
308 for each data pair. The closer the RMSE is to 0, the better the agreement. The RMSE was
309 divided by the mean of the observed values (normalised RMSE) to allow comparison of the

310 prediction accuracy among different nutritive attributes. In addition, predictions were
311 evaluated with the relative mean bias error (rMBE):

$$312 \quad rMBE = \frac{\sum_{i=1}^n P_i - O_i}{\sum_{i=1}^n O_i} \quad (2)$$

313 The rMBE provides a measure of the relative magnitude of over- or under-estimation of the
314 nutritive attributes. Willmott's index of agreement (d-index) was also used to evaluate the
315 model predictions:

$$316 \quad d = 1 - \left[\frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (|P_i| + |O_i|)^2} \right] \quad (3)$$

317 where P_i' and O_i' are the means of the predicted and observed values and the closer d is to 1,
318 the better the agreement between observed and simulated values. According to Willmott
319 (1981), d-index should be used to show the agreement between observations and predictions
320 in a dimensionless way, as a complement to the RMSE. Observed and simulated pairs of
321 nutritive attributes were also plotted against the amount of N applied per cut, mean annual
322 temperature and accumulated annual precipitation, to identify any trends in prediction
323 accuracy across the environmental variability within the calibration and validation datasets.

324

325 **3. Results**

326 **3.1. Cultivar-specific calibration and validation**

327 Predictions of CP concentration with BASGRA had a lower normalised RMSE (19 %) than those
328 predicted by CATIMO and STICS (50 % and 40 %, respectively) in the cultivar-specific
329 calibration (Fig. 2; Table 3). Both STICS and BASGRA had a lower normalised RMSE (24 % and
330 26 %, respectively) than CATIMO (45 %) in the cultivar-specific validation (Table 4). BASGRA
331 under-estimated observed CP concentrations (rMBE = -6 %) in the cultivar-specific calibration
332 (Table 3; Fig. 2), due to under-estimation of high CP concentrations, whereas CATIMO and
333 STICS over-estimated the observed CP concentrations (rMBE = +19 % and +29 % respectively)
334 (Table 3), mostly because of over-estimation of high CP concentrations. In the cultivar-specific
335 calibration with data from several locations and cultivars, BASGRA and STICS predicted the

336 NDF concentration with lower normalised RMSE (13 % and 8 %, respectively) and greater d-
337 index (0.59 and 0.75, respectively) than CATIMO (21 % and 0.43, respectively) (Fig. 3; Table 3).
338 For the cultivar-specific validation, however, there were no clear differences between the
339 three models in their ability to predict NDF concentration (Fig. 3; Table 4). The NDF
340 concentration was slightly under-estimated by all three models in the cultivar-specific
341 calibration (Table 3) and validation (Table 4). This under-estimation tended to be greater for
342 BASGRA (rMBE = -9.0 %) than for CATIMO and STICS (rMBE = -0.2 % and -0.4 %, respectively)
343 in the cultivar-specific calibration. CATIMO and STICS predicted dNDF with lower normalised
344 RMSE than BASGRA in the cultivar-specific calibration (10 %, 6 % and 22 %, respectively) (Table
345 3) and the cultivar-specific validation (7 %, 5 % and 10 %, respectively) (Table 4). STICS under-
346 estimated and CATIMO over-estimated dNDF in both the cultivar-specific calibration and
347 validation, while BASGRA slightly over-estimated dNDF in the cultivar-specific calibration and
348 under-estimated it in the cultivar-specific validation. However, the over-estimation in the
349 cultivar-specific calibration with BASGRA was greatly influenced by a large error in one single
350 measurement (Fig. 4), so it can be assumed that BASGRA under-estimated dNDF in both
351 calibration and validation.

352 Figure 2.

353 Figure 3.

354 Figure 4.

355 Table 3.

356 Table 4.

357 When plotted against the amount of N applied per cut, there was a slight increase in both
358 observed and simulated CP concentrations with increasing N level. However, CATIMO and
359 STICS tended to over-estimate CP concentration. For CATIMO, this trend was more noticeable
360 at high than at low N fertiliser levels (Fig. 5; Fig. 6). There were no clear trends in the
361 predictability of NDF concentration and dNDF across N fertiliser levels for any of the three
362 models (not shown). Moreover, it was not possible to discern any trends in predictability
363 among climate conditions when the three observed and simulated nutritive attributes were

364 plotted against mean annual air temperature and mean annual accumulated precipitation
365 (data not shown).

366 Figure 5.

367 Figure 6.

368

369 There was no clear trend between N fertilizer level and DM yield, possibly because many of
370 the measurements of dry matter and nutritive value were taken in between normal cutting
371 times. To further analyse the underlying mechanisms for the differences in the prediction
372 accuracy of CP concentration, simulated CP concentrations with the three models were
373 plotted against simulated dry matter yield and, while observed CP concentrations were
374 plotted against observed dry matter yield for the cultivar specific and generic calibration (Fig.
375 7). The plotted relationships indicate that CATIMO, and in some cases the other two models,
376 simulated higher CP concentration than what was observed at a similar dry matter yield,
377 especially at low dry matter yields.

378

379 **3.2. Generic calibration and validation**

380 The prediction accuracy across the three models in the generic calibration and validation
381 followed the same pattern as the cultivar-specific calibration and validation. The prediction
382 accuracy of CP concentration in the generic calibration was higher for BASGRA (normalised
383 RMSE = 16 %, d-index = 0.89) and STICS (normalised RMSE = 38 %, d-index = 0.92) than for
384 CATIMO (normalised RMSE = 101 %, d-index = 0.36) (Table 5). Similar differences in prediction
385 accuracy between the three models were obtained with the validation dataset (Table 6). In
386 the calibration (Table 5) and validation datasets (Table 6), the observed CP concentration was
387 slightly under-estimated by BASGRA, over-estimated by STICS, and greatly over-estimated by
388 CATIMO, based on rMBE. The prediction accuracy for NDF concentration and dNDF was also
389 higher for BASGRA and STICS than for CATIMO in the calibration (Tables 5) and validation
390 (Table 6), as indicated by lower normalised RMSE. However, prediction of NDF concentration

391 had a lower d-index with STICS than with BASGRA (Tables 5 and 6). The NDF concentration
392 was slightly under-estimated by CATIMO in the calibration and validation datasets, whereas it
393 was slightly under-estimated in the calibration dataset and over-estimated in the validation
394 dataset by BASGRA. CATIMO and to a lesser degree BASGRA under-estimated dNDF in the
395 calibration and validation datasets. STICS slightly over-estimated NDF concentration and dNDF
396 in both datasets (Tables 5 and 6). The generic calibration of BASGRA resulted in slightly better
397 predictions of the three nutritive attributes than the cultivar-specific calibration across
398 locations and cultivars, as indicated by lower normalised RMSE (Table 5). CATIMO predicted
399 CP concentration and dNDF less accurately in the generic calibration than in the cultivar-
400 specific calibration, whereas it predicted NDF concentration better in the generic calibration
401 (Table 5). STICS predicted NDF concentration better in the cultivar-specific calibration than in
402 the generic calibration, whereas the predictions of dNDF and CP concentration differed only
403 slightly between the cultivar-specific and generic calibrations (Table 5).

404 The trends in prediction of CP concentration across N fertiliser levels for the three models
405 were similar to those in the cultivar-specific calibration and validation, but with a tendency
406 for larger over-estimations by CATIMO under conditions with high N-fertiliser levels (Fig. 6).
407 Similarly to the cultivar-specific calibration, simulated NDF concentration and dNDF did not
408 show any trends across N-fertiliser levels for any of the three models. Moreover, there were
409 no discernible trends in predictability of the three nutritive attributes across differences in
410 mean annual air temperature and accumulated precipitation (not shown).

411 Table 5.

412 Table 6.

413 Figure 7.

414 **4. Discussion**

415 **4. 1. Differences in prediction accuracy among nutritive attributes**

416 This study examined how models with different structures and calibration procedures affect
417 the prediction of dNDF and concentrations of CP and NDF, in timothy under a broad range of
418 environmental conditions in the northern hemisphere. The predictions of NDF concentration

419 and dNDF were generally better than those of CP concentration, as indicated by lower
420 normalised RMSE and relative MBE and higher d-index in the calibrations of the three models.
421 This indicates that fibre concentration and digestibility can be predicted with higher accuracy
422 than N or CP concentration. These patterns in prediction accuracy of nutritive attributes that
423 were previously reported from evaluations of CATIMO (Bonesmo et al., 2005; Jing et al., 2013)
424 and STICS (Jégo et al., 2013) against data from field experiments in Canada are confirmed and
425 extended to BASGRA. Even though the timothy CP concentration was less accurately
426 simulated than the NDF concentration and dNDF in the studies cited above, it was generally
427 predicted with better accuracy than in our study. In both CATIMO and STICS, crop N demand
428 is based on critical and maximum N dilution curves. The parameters of those curves
429 established for Canadian cultivars, which were not calibrated in our study, might not be
430 adequate for European cultivars. Our results indicate that existing forage grass models are
431 more efficient at predicting NDF concentration and dNDF than CP concentration.

432

433 **4. 2. Differences in predictability between cultivar-specific and generic and calibration**

434 The variability in prediction accuracy between cultivar-specific and generic calibrations
435 provides information on the required calibration of forage grass models used to predict
436 nutritive value. The fact that CATIMO and STICS tended to have better prediction accuracy
437 with the cultivar-specific calibration than with the generic calibration suggests that separate
438 calibrations for different cultivars could improve their predictive capacity. The overall slightly
439 better prediction accuracy of BASGRA in the generic calibration than in the cultivar-specific
440 calibration is, however, surprising. One reason could be that the larger dataset in the generic
441 calibration than in the cultivar-specific calibration limited the influence of outliers and resulted
442 in more accurate predictions. Van Oijen et al. (2013) found that a generic calibration of models
443 for Scots pine trees did not result in less accurate growth predictions than calibrations using
444 country-specific data. It should be noted, however, that the cultivar-specific datasets in our
445 study were obtained from experiments under different environmental conditions. Hence,
446 differences in prediction accuracy between the cultivar-specific and generic calibrations could
447 be at least partly the result of non-cultivar differences between experimental sites, including
448 differences in climate, soils and crop management. However, this was not confirmed by the

449 analyses of observed and simulated nutritive attributes against N fertiliser levels, mean annual
450 air temperature and accumulated annual precipitation, which revealed little information
451 about the impact of environmental variability on model prediction ability. Nevertheless, a
452 previous study in which LINGRA, the predecessor of BASGRA, simulated only one timothy
453 cultivar (cv. Grindstad) in a number of field experiments in northern Europe showed better
454 prediction of aboveground DM biomass when the model was calibrated specifically for one
455 experimental site than when it was calibrated using data from several sites (Persson et al.,
456 2014). To single out the effects of cultivars on calibration accuracy without any possible
457 confounding effects from weather, soil or other environmental factors, comparisons of
458 cultivar-specific and generic calibrations could be performed against data from one single site
459 should there be any such datasets available. Moreover, further knowledge on cultivar-specific
460 traits that are important to the prediction accuracy for nutritive attributes could possibly be
461 obtained by grouping cultivars with similar traits together in the same calibration.

462

463 **4. 3. Comparisons with dry matter yield predictability**

464 The prediction accuracy of nutritive attributes was generally within the same range or better
465 than the prediction accuracy of the forage DM yield for the same three models and partly the
466 same experimental data (Korhonen et al., 2018). The normalised RMSE for the forage DM yield
467 predictions reported from the study by Korhonen et al. (2018), which varied between 24 and
468 93 % across calibrations and validations, was generally greater than that for the nutritive
469 attributes in both generic calibrations and validations. Sixty-eight out of the 101 treatments
470 that were used in our study were not included in the study of Korhonen et al. (2018). The
471 calibration techniques applied for CATIMO and STICS meant that dry matter related
472 parameters calibrated in the study of Korhonen et al. (2018) for the other cultivars and the
473 generic calibration did not change. The new Grindstad treatments added here can hence be
474 regarded as an additional validation of the Grindstad and generic calibrations. The normalized
475 RMSE for the calibration treatments of the Nuutti (CATIMO 62 %, STICS 27 %) and the newly
476 added Grindstad treatments (CATIMO 66 %, STICS 25 %) from Maaninka and Ruukki 2015 and
477 2016, and the normalised RMSE of the generic calibration as evaluated against the same
478 treatments (CATIMO 129 %, STICS 26 %) were mostly within the same range as the normalised

479 RMSE of the DM yield predictions in Korhonen et al. (2018). For comparing models, we
480 calculated the RMSE of the DM predictions for the same treatment also for the BASGRA
481 calibrations in which, unlike the CATIMO and STICS calibrations, the values of all parameters
482 changed during the cultivar-specific calibration of this model. The normalised RMSE for the
483 calibration treatments of the Nuutti and Grindstad cultivars from Maaninka 2015 and 2016
484 was 15 and 20 % respectively. For the generic calibration of BASGRA, the normalised RMSE for
485 the same treatments was 32 %. In total, these results illustrate that regardless of the
486 calibration technique the inclusion of nutritive value here was not at the expense of the
487 predictability of the DM yield.

488 In previous evaluations of STICS (Jégo et al., 2013) and CATIMO (Bonesmo et al., 2005; Jing et
489 al., 2013) for Canadian locations and timothy cultivars, the normalised RMSE for forage DM
490 yield predictions was between 70 and 120 % greater than for NDF concentration, and between
491 220 and 440 % greater than for dNDF. Our results confirm that nutritive value predictions can
492 be as accurate as DM yield predictions in forage grasses.

493

494

495 **4. 4. Possible explanations for the differences in prediction accuracy**

496 Crude protein concentration in plants results from rather complex soil and plant N processes,
497 which are all affected by soil conditions, weather and crop management. Besides possible
498 errors in the input data, errors in the descriptions of processes could have affected the CP
499 concentration predictions. That those parameters, which were related to CP concentration
500 were calibrated in BASGRA, but not in CATIMO and STICS, may have been a reason for the
501 difference in prediction accuracy of this attribute among the three models. The higher
502 simulated CP concentration at low simulated dry matter yield than the observed CP
503 concentration at similar observed dry matter yield in CATIMO and to a lesser extent in STICS
504 (Figure 7) indicates that the assumption of N dilution with biomass that was taken from
505 previous model development against field trial data in Canada was not applicable to the
506 cultivars and environmental conditions in northern Europe that were investigated here, at
507 least not at low dry matter yield. Further experimental studies are needed to demonstrate

508 whether there are differences in N demand and uptake between timothy cultivars. However,
509 the greater difference in CP prediction accuracy between CATIMO and STICS than between
510 the latter and BASGRA indicates that there are other underlying reasons than the
511 representation of plant N dilution with biomass or light extinction for the differences in CP
512 prediction accuracy. One reason could be differences in leaf/stem ratio prediction accuracy,
513 which would affect growth and hence N uptake and concentration. Unfortunately, there were
514 insufficient data available to thoroughly analyse correlations between leaf/stem ratio and
515 predictions of CP concentration. Forage NDF concentration increases and dNDF decreases
516 with phenological development, and these variables also directly affected by temperature in
517 CATIMO and STICS. Although severe N stress affects NDF concentration and dNDF in CATIMO
518 and STICS, there is no effect of soil and plant N on NDF concentration and dNDF under normal
519 N conditions in any of the three models which, given the complexity of soil-plant N processes,
520 could explain why they are better predicted than CP concentration. Differences between
521 models in calculation of leaf/stem ratios could also explain some of the variation in predicted
522 NDF concentration and dNDF among the three models. However, the effect of the leaf/stem
523 ratio on CP concentration is probably larger, due to the complex interaction between N
524 distribution in the plant and growth.

525

526 **4. 5. Uncertainty in input data**

527 Because the methods used for analysis of the three nutritive attributes were not always the
528 same at all sites, there is some uncertainty in the values (Huhtanen et al., 2006). Of the three
529 nutritive attributes included in our study, dNDF most likely has the largest uncertainty
530 associated with the analysis methods and CP concentration the smallest. Different dNDF
531 values for the same forage sample analysed *in vitro* in different laboratories may stem from
532 differences in the pore size of the nylon bags in which the samples are incubated and from
533 differences in the incubation time and the rumen liquid used. Similarly, differences in
534 methodology between laboratories, such as the use of different extraction chemicals, may
535 result in laboratory differences in NDF concentration estimates (Tavares da Silva et al., 2018).
536 It should also be noted that NIRS often has poorer prediction accuracy for NDF concentration
537 and dNDF than for CP concentration, although high accuracy can also be achieved for the

538 former attributes if the method is carefully calibrated with an adequate number of
539 representative reference samples and suitable reference analysis methods (Huhtanen et al.,
540 2006). Nevertheless, the better prediction accuracy of NDF concentration and dNDF than of
541 CP concentration indicates that other reasons than the uncertainty in nutritive value
542 measurements were more important to the prediction accuracy. Errors related to the weather
543 input data, mainly due to the distance between weather stations and observations in the field,
544 could also have affected our results. For most sites, there were no direct measurements of
545 soil hydraulic properties available as input to the soil modules of the grass models and instead
546 these variables were estimated from data on soil texture fractions. For BASGRA and STICS, the
547 soil water contents at wilting point and at field capacity were therefore treated as parameters.
548 However, that was not the case for CATIMO. Possible within-field variation in soil texture may
549 also have caused differences between the actual soil properties and those that were input to
550 the simulations.

551

552

553 **4. 6. Suggested further studies**

554 The low prediction accuracy of CP concentration, the importance of CP concentration for
555 nutritive value and the general importance of N for crop performance and for its
556 environmental impact emphasise the need for improved descriptions of soil and plant N in the
557 three models. Moreover, studies with synchronised calibration procedures could help assess
558 whether the differences in prediction accuracy between the models were due to differences
559 in calibration methods or the model structure. Such information could increase the
560 performance of models and thus their prospects of being applied in analysis of forage grass
561 performance under various existing or hypothetical environmental conditions. Further
562 calibrations with experimental data from other areas, such as Russia, northern Japan and
563 mountainous regions at lower latitudes in Europe, could provide more information about the
564 general applicability of the models. To place the performance of the three models in a broader
565 context, validation of the performance of other grassland models in predicting forage nutritive
566 value could also be valuable.

567

568 **5. Conclusions**

569 Three models with different structures (BASGRA, CATIMO and STICS) predicted NDF
570 concentration and digestibility in timothy with similar accuracy to previous predictions of
571 forage DM yield of timothy across a wide range of climate and soil conditions in Canada and
572 northern Europe. However, prediction of CP concentration was rather poor compared with
573 the other nutritive attributes. Cultivar-specific calibrations improved the performance of
574 CATIMO and STICS, but not of BASGRA, compared with calibrations where data on all cultivars
575 were used together.

576

577 **Acknowledgements**

578 This research was conducted in the context of MACSUR (Modelling European Agriculture with
579 Climate Change for Food Security), a project launched by the Joint Research Programming
580 Initiative (JPI) on Agriculture, Food Security and Climate Change. The authors gratefully
581 acknowledge financial support from the following sources: Academy of Finland (project
582 PLUMES, decision nos. 277403 and 292836), the European Agricultural Fund for Rural
583 Development (project NuRa) and the Research Council of Norway.

584

585

586 **References**

- 587 Bélanger, G., and Gastal, F. 2000, Nitrogen utilization by forage grasses. *Can. J. Plant Sci.*
588 80:11-20.
- 589 Bélanger, G., McQueen, R.E., 1999. Leaf and stem nutritive value of timothy grown with
590 varying N nutrition in spring and summer. *Can. J. Plant Sci.* 79, 223-229.
- 591 Bélanger, G., Michaud, R., Jefferson, P.G., Tremblay, G.F., Bréard, A., 2001. Improving the
592 nutritive value of timothy through management and breeding. *Can. J. Plant Sci.* 81,
593 577-585.
- 594 Bonesmo, H., Bélanger, G., 2002a. Timothy yield and nutritive value by the CATIMO model: I.
595 Growth and nitrogen. *Agron. J.* 94, 337-345.

596 Bonesmo, H., Bélanger, G., 2002b. Timothy yield and nutritive value by the CATIMO model:
597 II. Digestibility and fiber. *Agron. J.* 94, 345-350.

598 Bonesmo, H., Bélanger, G., Charmley, E., Drapeau, R., McKenzie, D.B., Michaud, R., Tremblay,
599 G.F., 2005. Timothy yield and nutritive value by the CATIMO model: III. Validation for
600 eastern Canada. *Agron. J.* 97, 32-40.

601 Brisson, N., Launay, M., Mary, B., Beaudoin, N., 2008, Conceptual basis, formalisations and
602 parameterization of the STICS crop model. Editions QUAE (Versailles).

603 Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.H., Ruget, F., Nicoullaud, B., Gate, P.,
604 Devienne-Barret, F., Antonioletti, R., Durr, C., Richard, G., Beaudoin, N., Recous, S.,
605 Tayot, X., Plenet, D., Cellier, P., Machet, J.M., Meynard, J.M., and Delécolle, R. 1998.
606 STICS: A generic model for the simulation of crops and their water and nitrogen
607 balances. I. Theory and parameterization applied to wheat and corn. *Agronomie* 18,
608 311-346.

609 Bruinenberg, M.H., Valk, H., Korevaar, H., Struik, P.C., 2002. Factors affecting digestibility of
610 temperate forages from seminatural grasslands: a review. *Grass Forage Sci.* 57, 292-
611 301.

612 Buxton, D.R., 1996. Quality-related characteristics of forages as influenced by plant
613 environment and agronomic factors. *Anim. Feed Sci. Tech.* 59, 37-49.

614 Casler, M.D., Kallenbach, R.L., 2007. Cool-season grasses for humid areas. *Forages. The*
615 *Science of Grassland Agriculture* 2, 211-220.

616 Chang, J., Viovy, N., Vuichard, N., Ciais, P., Wang, T., Cozic, A., Lardy, R., Graux, A.-I., Klumpp,
617 K., Martin, R., 2013. Incorporating grassland management in ORCHIDEE: model
618 description and evaluation at 11 eddy-covariance sites in Europe. *Geosci. Model Dev.*
619 6, 2165-2181.

620 Chapman, D.F., Lee, J.M., Waghorn, G.C., 2014. Interaction between plant physiology and
621 pasture feeding value: a review. *Crop Pasture Sci.* 65, 721-734.

622 Ehrhardt, F., Soussana, J.F., Bellocchi, G., Grace, P., McAuliffe, R., Recous, S., Sándor, R.,
623 Smith, P., Snow, V., de Antoni Migliorati, M., 2018. Assessing uncertainties in crop
624 and pasture ensemble model simulations of productivity and N₂O emissions. *Global*
625 *Change Biol.* 24, e603-e616.

626 Elgersma, A., Sjøgaard, K., 2018. Changes in nutritive value and herbage yield during
627 extended growth intervals in grass–legume mixtures: effects of species, maturity at
628 harvest, and relationships between productivity and components of feed quality.
629 *Grass Forage Sci.* 73, 78-93.

630 Fulkerson, W.J., Neal, J.S., Clark, C.F., Horadagoda, A., Nandra, K.S., Barchia, I., 2007.
631 Nutritive value of forage species grown in the warm temperate climate of Australia
632 for dairy cows: Grasses and legumes. *Livest. Sci.* 107, 253-264.

633 Gustavsson, A.-M., Martinsson, K., 2004. Seasonal variation in biochemical composition of
634 cell walls, digestibility, morphology, growth and phenology in timothy. *Eur. J. Agron.*
635 20, 293-312.

636 Huhtanen, P., Nousiainen, J., Rinne, M., 2006. Recent developments in forage evaluation
637 with special reference to practical applications. *Agr. Food Sci.* 15, 293-323.

638 Hurtado-Uria, C., Hennessy, D., Shalloo, L., Schulte, R.P.O., Delaby, L., O'Connor, D., 2013.
639 Evaluation of three grass growth models to predict grass growth in Ireland. *J. Agr. Sci-*
640 *Cambridge* 151, 91-104.

641 Höglind, M., Schapendonk, A., Van Oijen, M., 2001. Timothy growth in Scandinavia:
642 combining quantitative information and simulation modelling. *New Phytol.* 151, 355-
643 367.

644 Höglind, M., Hanslin, H.M., Van Oijen, M., 2005. Timothy regrowth, tillering and leaf area
645 dynamics following spring harvest at two growth stages. *Field Crops Res.* 93, 51-63.

646 Höglind, M., Van Oijen, M., Cameron, D., Persson, T., 2016. Process-based simulation of
647 growth and overwintering of grassland using the BASGRA model. *Ecol. Model.* 335, 1-
648 15.

649 Jégo, G., Bélanger, G., Tremblay, G.F., Jing, Q., Baron, V.S., 2013. Calibration and
650 performance evaluation of the STICS crop model for simulating timothy growth and
651 nutritive value. *Field Crops Res.* 151, 65-77.

652 Jensen, K.B., Robins, J.G., Rigby, C., Waldron, B.L., 2016. Comparative trends in forage
653 nutritional quality across the growing season in 13 grasses. *Can. J. Plant Sci.* 97, 72-
654 82.

655 Jing, Q., Bélanger, G., Baron, V., Bonesmo, H., Virkajärvi, P., Young, D., 2012. Regrowth
656 simulation of the perennial grass timothy. *Ecol. Model.* 232, 64-77.

657 Jing, Q., Bélanger, G., Baron, V., Bonesmo, H., Virkajärvi, P. 2013. Simulating the nutritive
658 value of timothy summer regrowth. *Agron. J.* 105, 563-572.

659 Jokela, V., Trevaskis, B., Seppänen, M.M., 2015. Genetic variation in the flowering and yield
660 formation of timothy (*Phleum pratense* L.) accessions after different photoperiod and
661 vernalization treatments. *Front. Plant Sci.* 6.

662 Justes, E., Mary, B., Meynard, J.M., Machet, J.M., and Thelier-Huche, L. 1994. Determination
663 of a critical nitrogen dilution curve for winter wheat crops. *Ann. Bot-London* 74, 397-
664 407.

665 Kipling, R.P., Virkajärvi, P., Breitsameter, L., Curnel, Y., De Swaef, T., Gustavsson, A.-M.,
666 Hennart, S., Höglind, M., Järvenranta, K., Minet, J., Nendel, C., Persson, T., Picon-
667 Cochard, C., Rolinski, S., Sandars, D.L., Scollan, N.D., Sebek, L., Seddaiu, G., Topp,
668 C.F.E., Twardy, S., Van Middelkoop, J., Wu, L., Bellocchi, G., 2016. Key challenges and
669 priorities for modelling European grasslands under climate change. *Sci. Total Environ.*
670 566–567, 851-864.

671 Köchy, M., 2008. Effects of simulated daily precipitation patterns on annual plant
672 populations depend on life stage and climatic region. *BMC Ecol.* 8, 4.

673 Köppen, W., 1936: *Das Geographische System der Klimate*. In: W. Köppen and R. Geiger eds.
674 *Handbuch der Klimatologie*, vol I, Part C. Gebrüder Borntraeger, Berlin. Germany.

675 Korhonen, P., Palosuo, T. Persson, T., Höglind, M., Jégo, G., Van Oijen, M., Gustavsson, A.-M.,
676 Bélanger, G., Perttu Virkajärvi, P., 2018. Modelling grass yields in northern climates -
677 a comparison of three growth models for timothy. *Field Crops Res.* 224,37-47.

678 Marum, P., 1990. Bestemmelse av kvalitet i forvekster ved hjelp av NIRS og metodens
679 muligheter i engvekstforedling. *Norsk Landbruksforskning (Norway)*. Norwegian
680 *Agricultural Research*. no. 9/90 (suppl.).(NLH accession no. NOR).

681 Nissinen, O., Kalliainen, P., Jauhiainen, L., 2010. Development of yield and nutritive value of
682 timothy in primary growth and regrowth in northern growing conditions. *Agr. Food*
683 *Sci.* 19, 252-268.

684 Nordheim-Viken, H., Volden, H., Jorgensen, M., 2009. Effects of maturity stage, temperature
685 and photoperiod on growth and nutritive value of timothy (*Phleum pratense* L.).
686 *Anim. Feed Sci. Tech.* 152, 204-218.

687 Persson, T., Höglind, M., Gustavsson, A.-M., Halling, M., Jauhiainen, L., Niemeläinen, O.,
688 Thorvaldsson, G., Virkajärvi, P., 2014. Evaluation of the LINGRA timothy model under
689 Nordic conditions. *Field Crops Res.* 161, 87-97.

690 Pontes, L.S., Carrère, P., Andueza, D., Louault, F., Soussana, J.F., 2007. Seasonal productivity
691 and nutritive value of temperate grasses found in semi-natural pastures in Europe:
692 responses to cutting frequency and N supply. *Grass Forage Sci.* 62, 485-496.

693 Sándor, R., Barcza, Z., Acutis, M., Doro, L., Hidy, D., Köchy, M., Minet, J., Lellei-Kovács, E., Ma,
694 S., Perego, A., 2017. Multi-model simulation of soil temperature, soil water content
695 and biomass in Euro-Mediterranean grasslands: Uncertainties and ensemble
696 performance. *Eur. J. Agron.* 88, 22-40.

697 Sándor, R., Barcza, Z., Hidy, D., Lellei-Kovács, E., Ma, S., Bellocchi, G., 2016. Modelling of
698 grassland fluxes in Europe: evaluation of two biogeochemical models. *Agr. Ecosyst.*
699 *Environ.* 215, 1-19.

700 Schapendonk, A., Stol, W., van Kraalingen, D.W.G., Bouman, B.A.M., 1998. LINGRA, a
701 sink/source model to simulate grassland productivity in Europe. *Eur. J. Agron.* 9, 87-
702 100.

703 Sivia, D.S., 2006. *Data Analysis: A Bayesian Tutorial*. Oxford University Press, Oxford, UK.

704 Tavares da Silva, R.S., Fernandes A.M., dos Santos Gomez, R., Ramos Bendia, L.C., da Costa e
705 Silva, L., Mendonca Viera, R.A. 2018. On the specificity of different methods for
706 neutral fiber and related problems. *Anim. Feed Sci. Tech.* 240, 128-144.

707 Thornton, P.K., 2010. Livestock production: recent trends, future prospects. *Philos. T. Roy.*
708 *Soc. B.* 365, 2853-2867.

709 Van Oijen, M., Reyer, C., Bohn, F., Cameron, D., Deckmyn, G., Flechsig, M., Härkönen, S.,
710 Hartig, F., Huth, A., Kiviste, A., 2013. Bayesian calibration, comparison and averaging
711 of six forest models, using data from Scots pine stands across Europe. *Forest Ecol.*
712 *Manag.* 289, 255-268.

713 Van Oijen, M., Rougier, J., Smith, R., 2005. Bayesian calibration of process-based forest
714 models: bridging the gap between models and data. *Tree Physiol.* 25, 915-927.

715 Van Soest, P.J., Robertson, J.B., Lewis, B.A., 1991. Methods for dietary fiber,
716 neutraldetergent fiber, and non starch polysaccharides in relation to animal nutrition.
717 *J. Dairy Sci.* 74, 3583-3597.

718 Van Soest, P.J., Wine, R.H., Moore, L.A., 1966. Estimation of the true digestibility of forages
719 by the in vitro digestion of cell walls. *Proc. Int. Grassland Congress, 10th, Helsinki,*
720 *Finland, 7-16 July 1966. Finnish Grassl. Assoc., Helsinki (1966), pp. 438-441.*

721 Vital, J.-A., Gaurut, M., Lardy, R., Viovy, N., Soussana, J.-F., Bellocchi, G., Martin, R., 2013.
722 High-performance computing for climate change impact studies with the Pasture
723 Simulation model. *Comput. Electron. Agr.* 98, 131-135.

724 Wilkinson, J.M., Rinne, M., 2018. Highlights of progress in silage conservation and future
725 perspectives. *Grass Forage Sci.* 73, 40-52.

726 Willmott, C.J., 1981. On the validation of models. *Phys. Geogr.* 2, 184-194.

727 Wu, L., McGechan, M., McRoberts, N., Baddeley, J., Watson, C., 2007. SPACSYS: integration
728 of a 3D root architecture component to carbon, nitrogen and water cycling—model
729 description. *Ecol. Model.* 200, 343-359.

731

732 Table 1. Cultivars, nutritive value (crude protein (CP) concentration, neutral detergent fibre
 733 (NDF) concentration, digestibility of NDF (dNDF)) and N fertilizer treatments at the
 734 experimental sites

Location	Mean annual temp (°C) /annual acc. precipitatio n (mm)	Köppen climate classification ¹	Soil type	Datase t year	Cultivar	Nutritiv e values	N fert. regim e (kg ha ⁻¹ yr ⁻¹)
Fredericto n (45°55'N; 66°32'W; 35 m asl)	5.7/1108	Dfb (Warm- summer humid continental climate)	Loam	1991- 1993	Champ	NDF, dNDF	0, 70, 140, 168, 200, 210
Lacombe (52°28'N; 113°44'W; 860 m asl)	3.5/429	Dfb (Warm- summer continental climate)	Silty clay loam	2004- 2005	Climax	NDF	100
Québec (46°47'N; 71°07'W; 75 m asl)	5.3/1009	Dfb (Warm- summer humid continental climate)	Loam	1999- 2001	Champ	NDF, dNDF	0, 60, 120

Maaninka (63°09'N; 27°17'E; 90 m asl)	4.2/560	Dfc (Continental Subarctic Climate).	Silt loam / loam 2	2006- 2007, 2015- 2016	Tammist o II, Nuutti, Grindsta d	NDF, CP	0, 150, 180, 190, 200, 250, 300, 350, 400, 450
Rovaniemi (66°35'N; 26°01'E; 106 m asl)	1.0/610	Dfc (Continental Subarctic Climate).	Silt loam	1999- 2001	Iki -	-	160, 200
Ruukki (64°40'N; 25°06' E; 48 m asl)	2.6/513	Dfc (Continental Subarctic Climate).	Silt loam	2015- 2016	Nuutti, Grindsta d	NDF, CP	0, 150, 200, 250, 300, 350, 400, 450
Særheim	8.0/1392	Cfc (Cool oceanic climate)	Sand y loam	2000- 2002	Grindsta d	NDF, CP	220

(58°46'N;

5°39'E; 90

m asl)

Umeå 3.3/595 Dfc Silt 1995- Jonatan NDF, CP 180

(63°45'N; (Continental loam 1996

20°17'E; subarctic

12 m asl) climate).

735 ¹ Köppen 1936

736 ² Treatments 1-3, 33-37: Silt loam soil. Treatments 38-68: Loam soil

737

738 Table 2. Division between calibration and validation data within the dataset

Treatment number	Location	Cultivar	N fertiliser application (kg ha ⁻¹ yr ⁻¹)	Calibration/validation
1-2	Maaninka	Tammisto II	180	Calibration
3-8	Rovaniemi	Iki	160, 200	Calibration
9, 11-12, 14	Særheim	Grindstad	220	Calibration
10, 13	Særheim	Grindstad	220	Validation
15-21,23	Québec	Champ	0,60,120	Calibration
22	Québec	Champ	60	Validation
24-25	Lacombe	Climax	100	Calibration
26-27	Umeå	Jonatan	180	Calibration
28, 30-32	Fredericton	Champ	0, 70, 140, 168, 210	Calibration
29, 33	Fredericton	Champ	200, 210	Validation
34-35, 37, 46- 50, 52-53, 62- 68	Maaninka	Nuutti	0, 150, 190, 200, 250, 300, 350, 400, 450	Calibration
36, 51, 69	Maaninka	Nuutti	190, 350, 450	Validation
38-40, 42-45, 54-59, 61	Maaninka	Grindstad	0, 150, 200, 250, 300,	Calibration

			350,	400,	
			450		
41,60	Maaninka	Grindstad	250, 400		Validation
71-77,	86-91,	Ruukki	Grindstad	0, 150, 200,	Calibration
93			250,	300,	
			350,	400,	
			450		
70, 92	Ruukki	Grindstad	0, 400		Validation
78-82,	84-85,	Ruukki	Nuutti	0, 150, 200,	Calibration
94-97, 99-101			250,	300,	
			350,	400,	
			450		
83, 98	Ruukki	Nuutti	300, 350		Validation

739

740

741 Table 3. Statistics on the cultivar-specific calibration: Observed and simulated means, root
 742 mean squared error (RMSE), normalised RMSE, relative mean bias error (rMBE) and Willmott's
 743 index of agreement (d-index) for crude protein (CP) concentration, neutral detergent fibre
 744 (NDF) concentration and digestibility of NDF (dNDF)

	Number of observations	Mean of observation	Mean of simulation	RMSE	Normalised RMSE (%)	rMBE (%)	d-index
CP concentration (g g ⁻¹ DM)							
BASGRA			0.13	0.027	19	-6.0	0.82
CATIMO	173	0.14	0.17	0.070	50	19.0	0.57
STICS			0.18	0.055	40	29.0	0.88
NDF concentration (g g ⁻¹ DM)							
BASGRA			0.51	0.072	13	-9.0	0.59
CATIMO	252	0.56	0.57	0.120	21	-0.2	0.43
STICS			0.56	0.045	8	-0.4	0.75
dNDF (g g ⁻¹ NDF)							
BASGRA			0.71	0.170	22	0.7	0.72
CATIMO	28	0.78	0.82	0.077	10	5.0	0.64
STICS			0.78	0.046	6	-3.0	0.82

745

746

747

748 Table 4. Statistics of the cultivar-specific validation: Observed and simulated means, root
 749 mean squared error (RMSE), normalised RMSE, relative mean bias error (rMBE) and Willmott's
 750 index of agreement (d-index) for crude protein (CP) concentration, neutral detergent fibre
 751 (NDF) concentration and digestibility of NDF (dNDF)

	Number of observations	Mean of observation	Mean of simulation	RMSE	Normalized RMSE (%)	rMBE (%)	d-index
CP concentration (g g ⁻¹ DM)							
BASGRA			0.13	0.037	26	-5.0	0.72
CATIMO	48	0.14	0.17	0.063	45	18.0	0.67
STICS			0.16	0.034	24	11.0	0.93
NDF concentration (g g ⁻¹ DM)							
BASGRA			0.52	0.063	11	-0.1	0.63
CATIMO	62	0.55	0.55	0.077	14	-0.7	0.64
STICS			0.55	0.047	9	-0.5	0.71
dNDF (g g ⁻¹ NDF)							
BASGRA			0.70	0.081	10	-0.1	0.70
CATIMO	14	0.78	0.81	0.053	7	0.5	0.81
STICS			0.75	0.041	5	-3.0	0.87

752

753

754

755 Table 5. Statistics of the generic calibration: Observed and simulated means, root mean
 756 squared error (RMSE), normalised RMSE, relative mean bias error (rMBE) and Willmott's index
 757 of agreement (d-index) for crude protein (CP) concentration, neutral detergent fibre (NDF)
 758 concentration and digestibility of NDF (dNDF)

	Number of observations	Mean of observation	Mean of simulation	RMSE	Normalised RMSE (%)	rMBE (%)	d-index
CP concentration (g g ⁻¹ DM)							
BASGRA			0.15	0.022	16	-0.8	0.89
CATIMO	173	0.14	0.26	0.14	101	87	0.36
STICS			0.17	0.052	38	25	0.92
NDF concentration (g g ⁻¹ DM)							
BASGRA			0.56	0.050	8.8	-0.4	0.72
CATIMO	252	0.56	0.55	0.095	17	-3	0.49
STICS			0.56	0.066	12	0.8	0.46
dNDF (g g ⁻¹ NDF)							
BASGRA			0.75	0.072	9.3	-4	0.59
CATIMO	28	0.78	0.51	0.29	38	-34	0.34
STICS			0.79	0.050	6.4	0.23	0.82

759

760

761 Table 6. Statistics of the generic validation: Observed and simulated means, root mean
 762 squared error (RMSE), normalised RMSE, relative mean bias error (rMBE) and Willmott's index
 763 of agreement (d-index) for crude protein (CP) concentration, neutral detergent fibre (NDF)
 764 concentration and digestibility of NDF (dNDF)

	Number of observations	Mean of observation	Mean of simulation	RMSE	Normalized RMSE (%)	rMBE (%)	d-index
CP concentration (g g ⁻¹ DM)							
BASGRA			0.15	0.025	18.0	-0.2	0.91
CATIMO	48	0.14	0.26	0.130	92.0	86.0	0.47
STICS			0.16	0.032	23.0	11.0	0.95
NDF concentration (g g ⁻¹ DM)							
BASGRA			0.56	0.043	7.8	2.0	0.77
CATIMO	62	0.56	0.55	0.095	17.1	-2.0	0.51
STICS			0.56	0.069	12.4	0.9	0.49
dNDF (g g ⁻¹ NDF)							
BASGRA			0.75	0.050	6.5	-4.0	0.99
CATIMO	14	0.78	0.51	0.290	37.0	-36.0	0.34
STICS			0.79	0.047	6.0	0.2	0.99

765

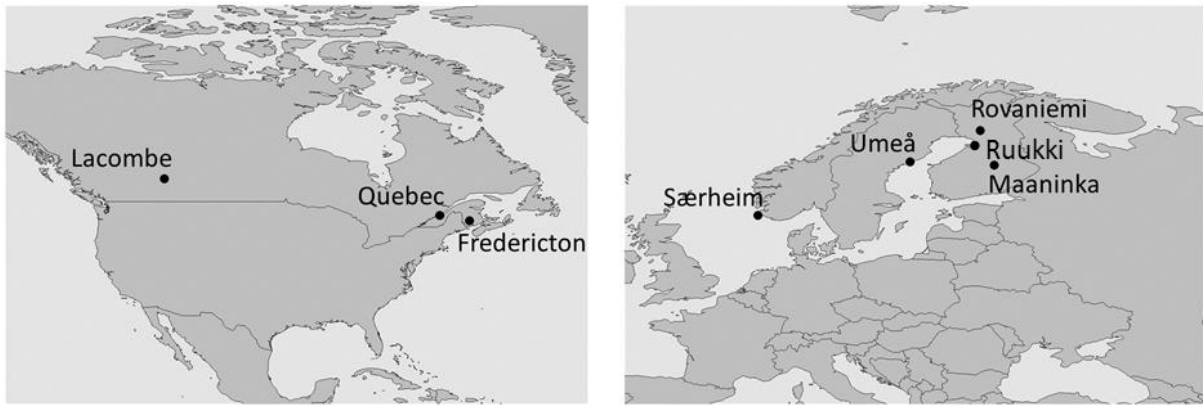


Figure 1. Geographical location of the eight experimental sites in Canada and Northern Europe.

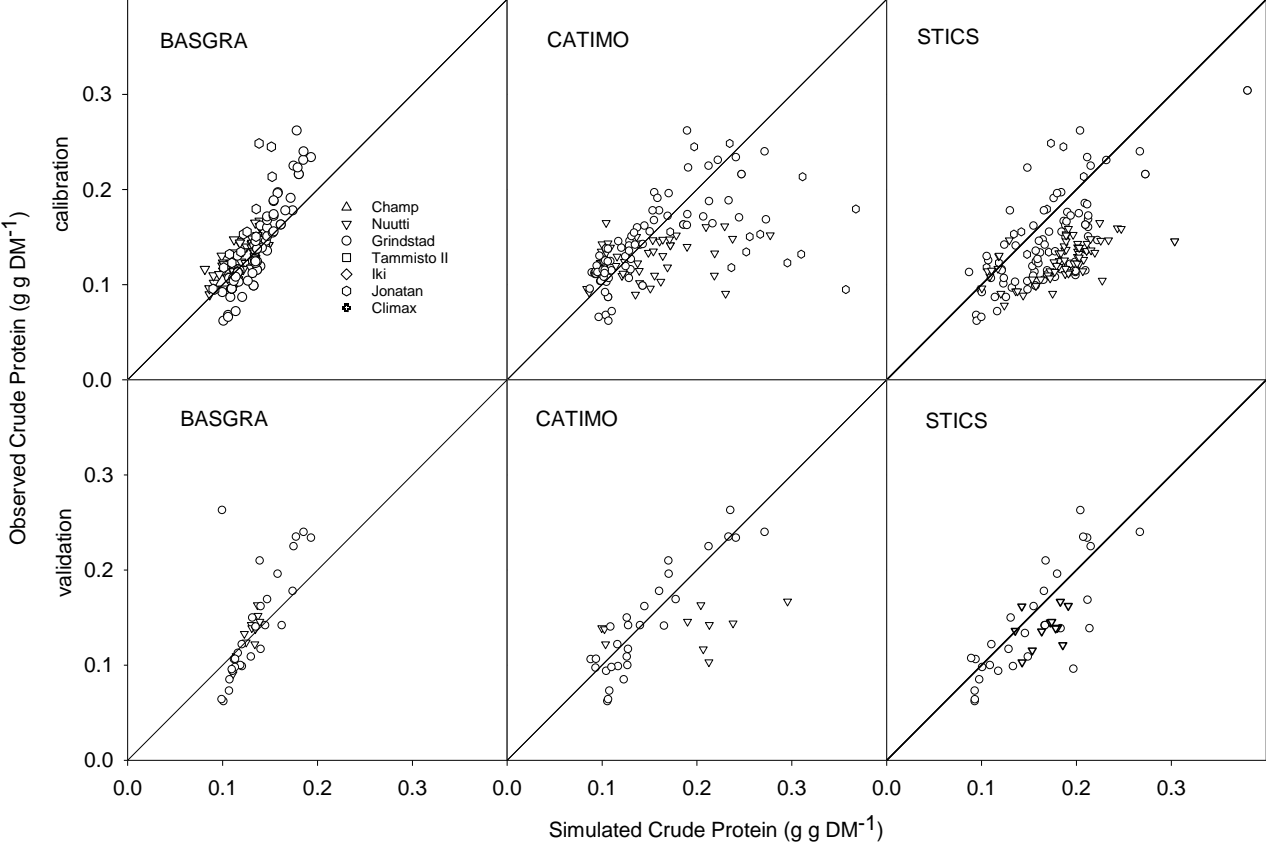


Figure 2. Observed crude protein concentration plotted against the simulated concentrations produced by BASGRA, CATIMO and STICS in the cultivar-specific calibration (upper row) and validation (lower row). Statistics on the cultivar-specific calibration and validation are presented in Tables 3 and 4.

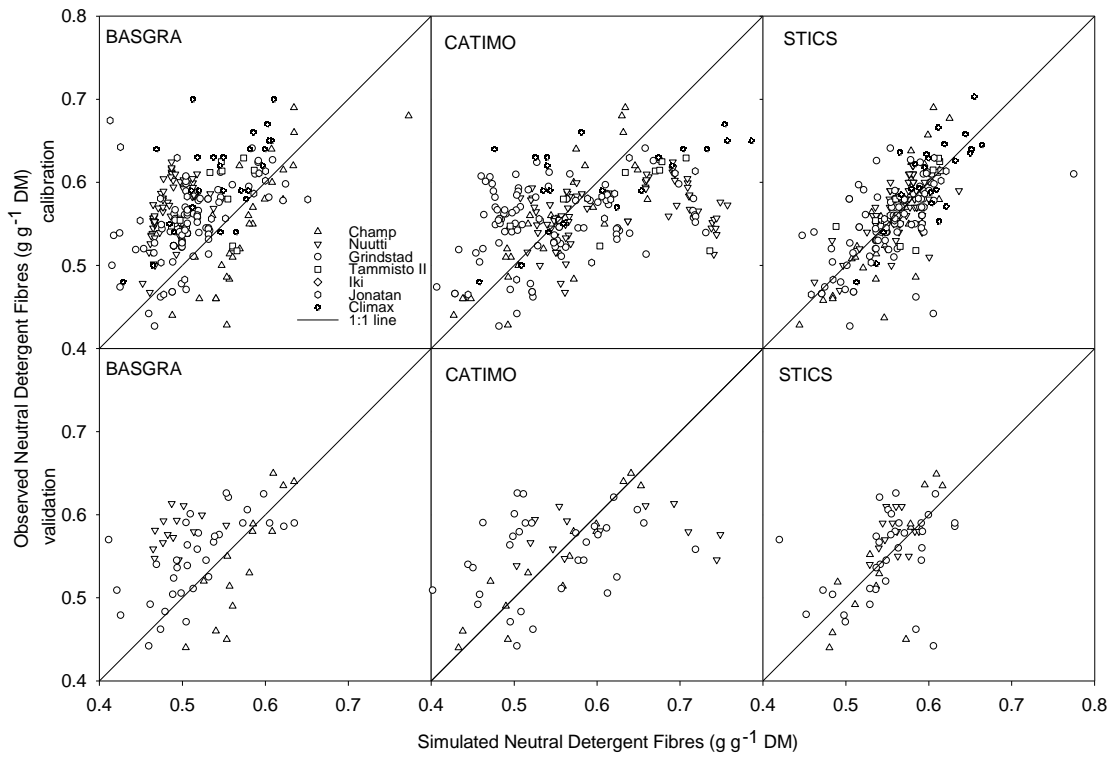


Figure 3. Observed neutral detergent fibre concentration plotted against the simulated concentrations produced by BASGRA, CATIMO and STICS in the cultivar-specific calibration (upper row) and validation (lower row). Statistics on the cultivar-specific calibration and validation are presented in Tables 3 and 4.

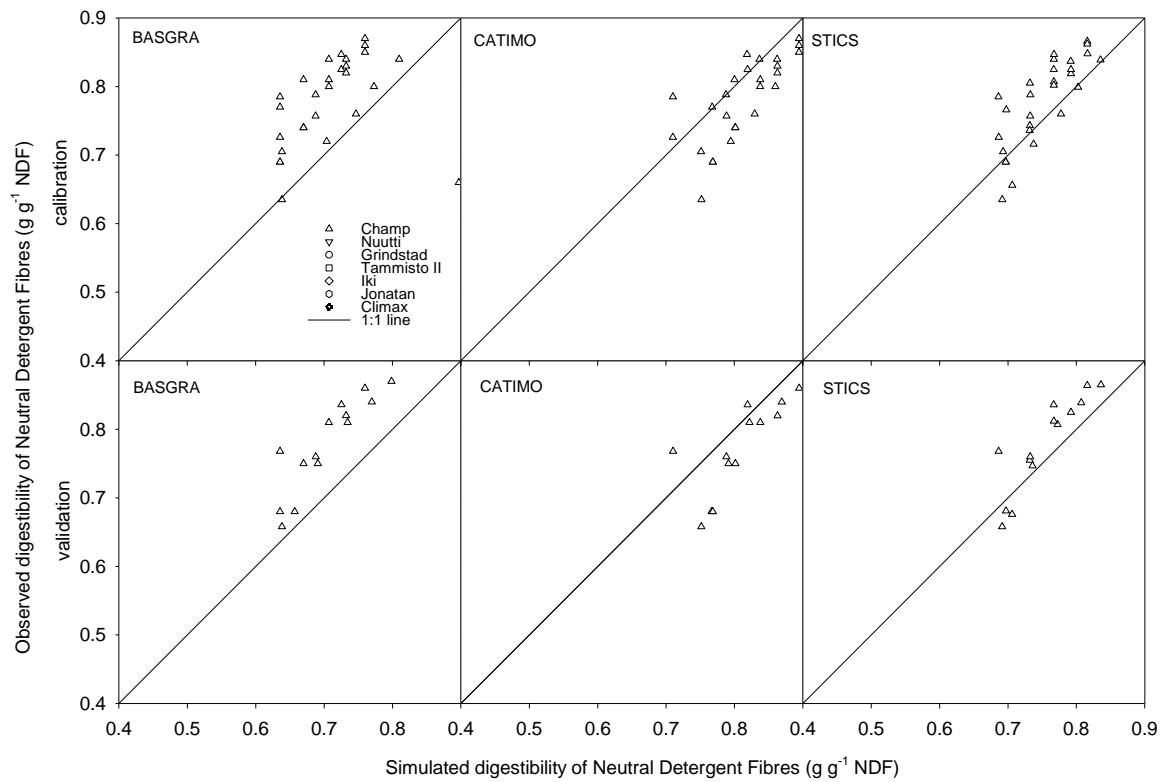


Figure 4. Observed digestibility of neutral detergent fibre (dNDF) plotted against the simulated values produced by BASGRA, CATIMO and STICS in the cultivar-specific calibration (upper row) and validation (lower row). Statistics on the cultivar-specific calibration and validation are presented in Tables 3 and 4.

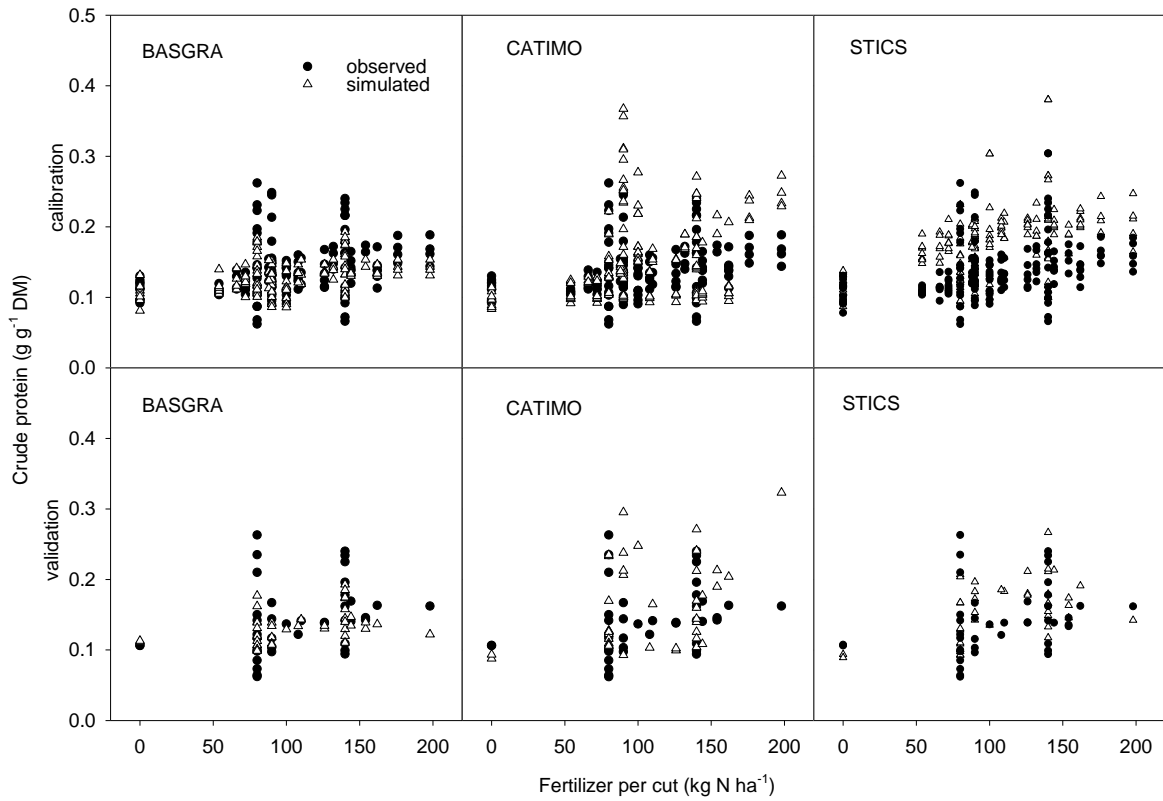


Figure 5. Observed crude protein concentration and concentration simulated by BASGRA, CATIMO, and STICS with cultivar-specific parameters as a function of amount of N applied per cut. Upper row: calibration dataset, lower row: validation dataset.

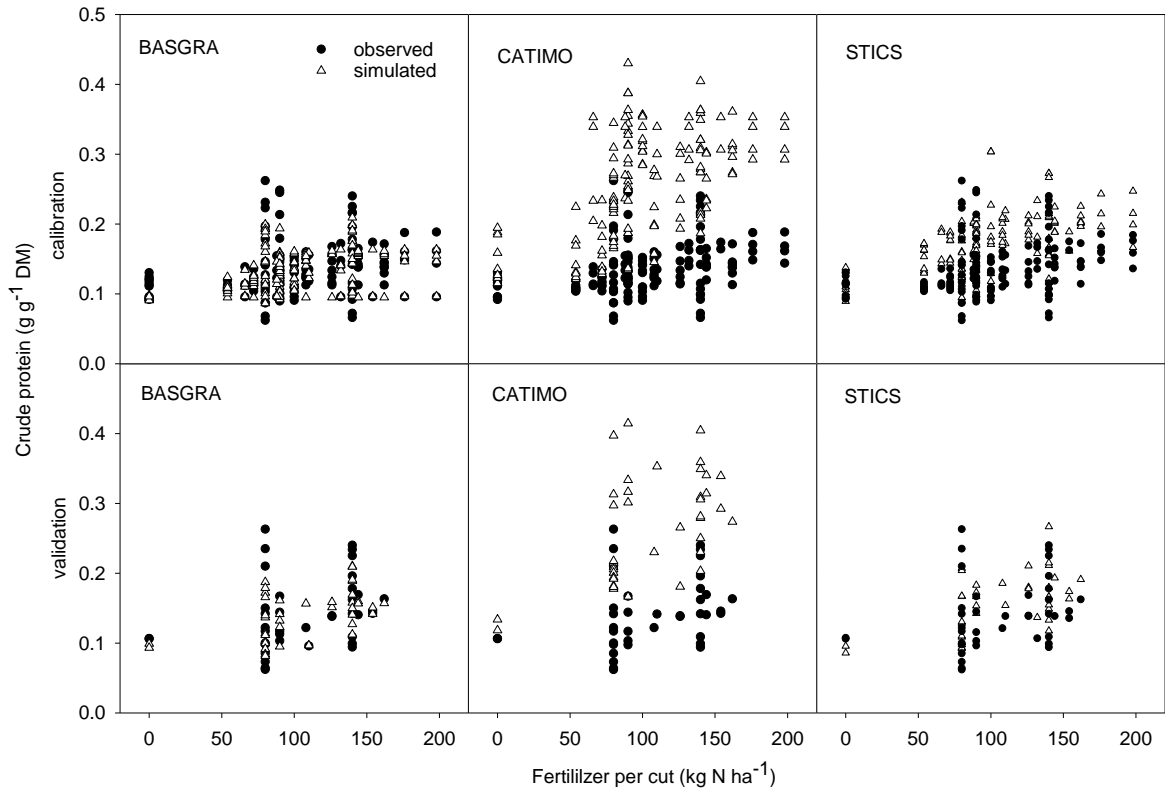


Figure 6. Observed crude protein concentration and concentration simulated by BASGRA, CATIMO, and STICS with generic parameters as a function of the amount of N applied per cut. Upper row: calibration dataset, lower row: validation dataset.

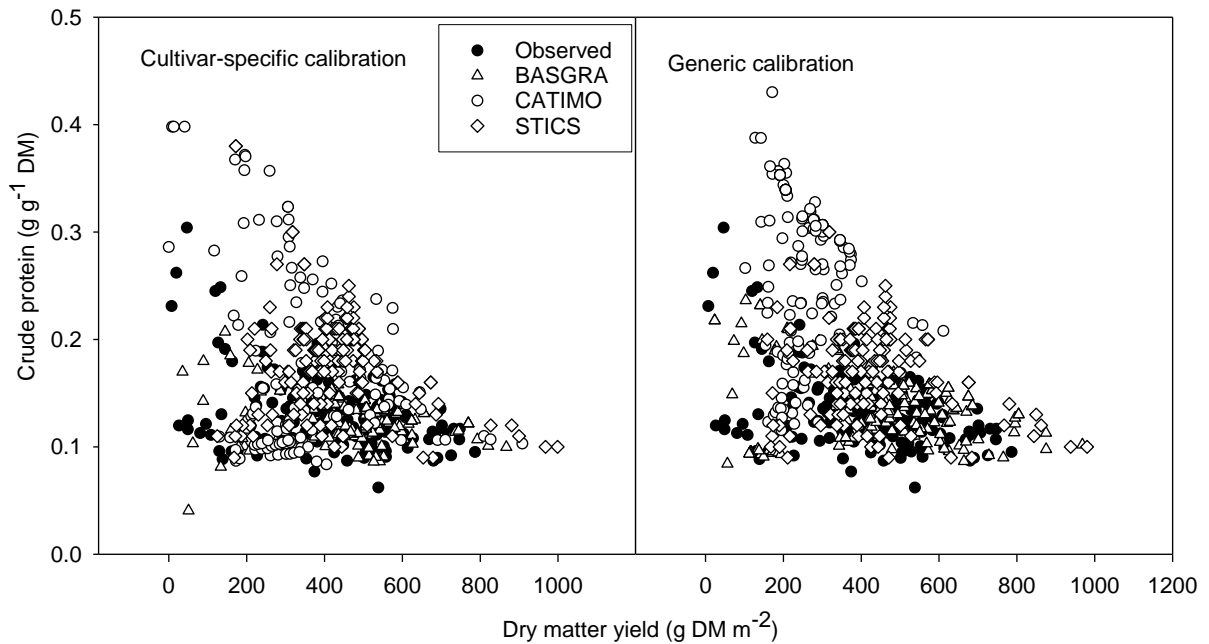


Figure 7. Observed crude protein concentration vs observed dry matter yield, and simulated crude protein vs simulated dry matter yield for the cultivar specific (left) and generic (right) calibrations.