

Calibrating primary crop parameters to capture undersown species impacts

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Bell, Q., Gerin, S., Douglas, N. ORCID: <https://orcid.org/0000-0002-3404-8761>, Quaife, T. ORCID: <https://orcid.org/0000-0001-6896-4613>, Liski, J. and Viskari, T. (2025) Calibrating primary crop parameters to capture undersown species impacts. *European Journal of Agronomy*, 169. 127676. ISSN 1873-7331 doi: [10.1016/j.eja.2025.127676](https://doi.org/10.1016/j.eja.2025.127676) Available at <https://centaur.reading.ac.uk/122793/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.eja.2025.127676>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



Calibrating primary crop parameters to capture undersown species impacts

Quentin Bell^{a,*}, Stephanie Gerin^a, Natalie Douglas^b, Tristan Quaife^b, Jari Liski^a,
Toni Viskari^{a,c}

^a Climate System Research, Finnish Meteorological Institute, P.O. BOX 503, Helsinki, Uusima 00101, Finland

^b National Centre for Earth Observation, Department of Meteorology, University of Reading, University of Reading, Earley Gate, Whiteknights Rd, Reading, Berkshire RG6 6ET, United Kingdom

^c European Commission, Joint Research Centre (JRC), Via Enrico Fermi, 2749, Ispra, Varese 21027, Italy

ARTICLE INFO

Keywords:

Crop model
STICS
Intercropping
Calibration
4DEnVar

ABSTRACT

Increasing plant diversity is seen as an important method for improving agricultural soil health, with benefits extending to increased soil carbon content and improved ecosystem health. However, including multi-species interactions into models is challenging. In this study, we represented multi-species interactions by producing different cash crop parameter values depending on the presence of undersown species. For this purpose, we used the STICS soil-crop model with observations of yield, green area index, and net ecosystem exchange chamber measurements from a number of plots in an extensive field experiment in southern Finland, where barley was cultivated with 0–8 undersown species. Calibration with the four-dimensional ensemble variational data assimilation method was found to be effective with up to three parameters of interest, beyond which issues of equifinality were present. Calibration had a positive effect on the performance of projected yields and other measurements, demonstrating the potential of this approach to capture unsimulated interactions. However, we found that the dominant parameter values change interannually to the degree that the calibration improvements did not translate, limiting the effectiveness of this approach to similar conditions as the calibration data. Additionally, large variance in secondary species abundance remains a challenge for modelling this style of intercropping.

1. Introduction

As the impacts of climate change on many facets of life become increasingly apparent, attempts to achieve net-zero carbon balances have become a central effort for various parties (Fankhauser et al., 2022; Black et al., 2021). With soils being the largest terrestrial carbon pool (Scharlemann et al., 2014; Friedlingstein et al., 2023), carbon farming practices are considered a potential method to increase ecosystem carbon allocation (Rumpel et al., 2020; Minasny et al., 2017). One of the more prominent suggestions of such methods has been a wider implementation of undersown species as increased agricultural biodiversity is theorised to improve the resilience of cash crops (Crystal-Ornelas et al., 2021), to enhance carbon accumulation into the system (Lal, 2015), as well as to more efficiently protect soil health (Adetunji et al., 2020). However, to widen the implementation of such methods and assess the impacts in a constructive manner requires both a better understanding of the various interactions as well as the capacity to reliably model the

system (Schmidt et al., 2011).

Crop models represent the current understanding of various dynamics affecting crop growth (Asseng et al., 2014; Boote et al., 2013) and are an instrumental tool both for scientific research and for practical applications such as harvest projections. Due to the modelled systems primarily being cultivated cropland, crop models need to be able to feature more processes than general ecosystem models, for example, various management methods (Brown et al., 2018) or the impact of nutrient availability to the system (Fatichi et al., 2014). By contrast, simplifications are made for general ecosystem models so that they broadly apply to many different ecosystems, often at far larger scales than crop models. Furthermore, due to agricultural systems themselves having been heavily specialised, it has been necessary to develop various crop models in order to properly capture the multitude of dynamics (Pimentel et al., 2012).

Not only do crop models still share some of the vulnerabilities of general ecosystem models, such as simplified radiation interception

* Corresponding author.

E-mail address: quentin.bell@fmi.fi (Q. Bell).

<https://doi.org/10.1016/j.eja.2025.127676>

Received 27 September 2024; Received in revised form 30 April 2025; Accepted 2 May 2025

Available online 10 May 2025

1161-0301/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

calculations or the use of plant functional types to represent average behaviour (Fisher and Koven, 2020), their very specialised nature further heightens those modelling challenges. Due to the number of baseline processes that have to be included in crop models and the subsequent computational cost, calibrating these models present genuine practical issues (Seidel et al., 2018; Guillaume et al., 2011). Including these various processes presents a central challenge as not only are their impacts intertwined, only a limited amount of system outputs can truly be continuously measured. As a consequence, both overparameterization and equifinality are common hurdles that have to be addressed in any efforts to calibrate these models (Beven and Freer, 2001).

Furthermore, attempting to include cover and undersown crop dynamics into these models brings additional fundamental issues. The common approach in vegetation dynamic modelling is to represent the development of each plant species independently of other species in the same plot (Ma et al., 2022; Martin Belda et al., 2022). Generally, the only interactive dynamics between different species that are currently implemented either relate to light transmittance through the canopy (Dai et al., 2004; Shiklomanov et al., 2021) or water availability (Gouttevin et al., 2012; Gerten et al., 2004). Some of the theories regarding the benefits of the secondary species concern soil health (Yang et al., 2020), symbiotic interactions (Callaway, 2007), or bacterial and fungal states (Baldrian, 2017), but there are no effective formulations representing these processes that could be included in models.

Since the primary interest in a lot of practical application of crop models is either in the productivity of the cash crop itself or the total performance of the plot when it comes to carbon sequestration, here we test an approach where the changes induced by the secondary species are included in the properties of the cash crop, thereby accounting for the interaction processes not currently simulated by crop models. In essence, we calibrate the traits of the cash crop in the presence of a given secondary species and ideally capture the impact of the plot synergies in that manner. To do this, a nimble parameter estimation method is necessary to be able to perform calibration at different sites with varying measurements.

In the four-dimensional ensemble variational (4DnVar; Pinnington et al. 2020) data assimilation method, the optimal set of state (or parameter) values is determined by how ensembles generated with different sets of values perform when compared to observations. Because it is calculated from a single collection of ensemble runs instead of from an iterative process, such as the more traditional Monte Carlo Markov Chain (MCMC; Chib 2001; Haario et al. 2001; Vrugt and Ter Braak 2011; Dumont et al. 2014), it requires far fewer computational resources and produces estimates much faster. Consequently, the 4DnVar approach has been used to calibrate, for example, global soil moisture models (Pinnington et al., 2021; Reichle et al., 2023) and a land surface model including a crop parametrisation (Pinnington et al., 2020).

To study the utility of this method, we chose the STICS model (Brisson et al., 2003), a widely used crop model that has a wide array of processes included and which allows for modelling of two crops in the same plot simultaneously (Brisson et al., 2004; Corre-Hellou et al., 2009; Kherif et al., 2022). STICS is conventionally calibrated using frequentist optimisation methods (Wallach et al., 2011, 2024; Guillaume et al., 2011; Ravelojaona et al., 2023). Although, Bayesian MCMC-methods have also featured in the literature, albeit more rarely (Dumont et al., 2014; Buis et al., 2023). However, there is an increasing movement to apply Bayesian methods to crop and other ecosystem models in order to better capture, represent, and reduce uncertainty from all sources (Fer et al., 2021; van Oijen, 2017; Hartig et al., 2012). Additionally, while STICS has been used in Finland and similar cold climates prior to this study (Korhonen et al., 2018; Ravelojaona et al., 2023; Jégo et al., 2010), we believe this to be the first application of the STICS intercrop methodology to such climates.

The aims of this study are twofold: (1) to evaluate the utility of 4DnVar as a calibration method for crop models, specifically STICS; (2)

to see if calibrating primary crop parameters on intercropped plots with different undersown crops produces improved predictions of primary crop yield compared with conventional single-crop calibration. The test was conducted with data from the TWINWIN experiment (Cappelli et al., 2024) which has both information from multiple crop setups with different undersown species as well as varied types of observations. An additional benefit of this setup is that it also allows us to evaluate how different datastreams impact the calibration.

2. Methods

2.1. Study site

The measurements were conducted on the experimental TWINWIN site (<https://www.bsag.fi/en/projects/twinwin-project/>), which was established in 2019. The main goal of the TWINWIN experiment was to study the effect of plant diversity on carbon sequestration. The site, owned by the University of Helsinki, is located in Viikki (60°13'29.4"N, 25°01'12.7"E), on a mineral agricultural field (plots varied between 4 main soil types according to USDA soil taxonomy: loam, clay loam, silty clay, and silty clay loam). At the start of the experiment, the field had on average (n = 60) a C:N ratio of 13.8 ± 0.7 , carbon content of $3.0 \pm 0.7\%$ and nitrogen content of $0.2 \pm 0.1\%$. In 2020 and 2021, the mean annual temperature and annual total precipitation were $8.4\text{ }^\circ\text{C}$ & 863 mm and $6.4\text{ }^\circ\text{C}$ & 709 mm , respectively. Monthly and long-term (1991–2020) averages can be found in Table 1.

There, 60 plots were created, 4 bare fallows and 56 plots with barley (*Hordeum vulgare L.*) sown alone or with different undersown species. In total, there were 6 treatments: single-cropped barley with herbicide (4 plots) or without herbicide (8 plots), intercrops of barley with one undersown species (3 plots for each of 8 undersown species), with two undersown species (10 plots), four undersown species (6 plots) and with eight undersown species (4 plots). The different undersown species are listed in Table 2. Each species is characterised by two main functional traits: their rooting depth and their ability to fix nitrogen. More detailed information on the experimental design and its maintenance can be found in Cappelli et al. (2024).

2.1.1. Flux measurements

The manual chamber technique was used to measure Net Ecosystem Exchange (NEE) on 22 plots out of 60. The 22 plots included 4 different treatments: 4 plots with single-cropped barley with no herbicide, 8 plots with barley sown with one of each undersown species, 6 plots with barley plus 4 undersown species, and 4 plots with barley plus 8 undersown species. During the growing seasons, NEE measurements were conducted twice a month in 2020 and once a month in 2021. The set-up

Table 1

Monthly and yearly total precipitation, monthly and yearly mean air temperature from January 2020 to December 2021 (Kumpula station, Finnish Meteorological Institute 2024); in comparison to the long term average of 1991–2020 (Helsinki Kaisaniemi station, Jokinen et al. 2021).

Time period	Total precipitation (mm)			Mean air temperature (°C)		
	2020	2021	1991–2020	2020	2021	1991–2020
January	66.5	93.8	53	2.5	−3.9	−3.1
February	66.5	93.8	38	0.9	−7	−3.8
March	64.1	41.5	34	2.1	−0.3	−0.7
April	44	30.8	34	5	5	4.4
May	50.3	86.5	38	9.6	10.5	10.4
June	78.2	38.7	60	18.1	19.7	14.9
July	98.8	47.3	57	16.7	21.6	18.1
August	75.8	132.4	81	17	15.9	16.9
September	67.9	42.8	56	13.4	10.1	12.3
October	66.7	84.1	73	8.9	8.3	6.6
November	77.9	50.2	69	5	2.2	2.4
December	64	33.7	58	1.5	−5.9	−0.7
Year	863	709	653	8.4	6.4	6.5

Table 2
Undersown Species Used in the Experiment.

Common Name	Label	Species	Root Depth	Nitrogen Fixing
Alfalfa	AA	<i>Medicago sativa</i> L.	Deep	Yes
Alsike clover	AC	<i>Trifolium hybridum</i> L.	Shallow	Yes
Chicory	CI	<i>Cichorium intybus</i> L.	Deep	No
Tall fescue	FA	<i>Festuca arundinacea</i> Scrb.	Deep	No
Italian ryegrass	IR	<i>Lolium multiflorum</i> Lam.	Shallow	No
Red clover	RC	<i>Trifolium pratense</i> L.	Deep	Yes
Timothy-grass	TG	<i>Phleum pratense</i> L.	Shallow	No
White clover	WC	<i>Trifolium repens</i> L.	Shallow	Yes

consisted of a clear chamber connected by an inlet and outlet tube to a CO₂ and H₂O analyser (Li-140840A, LI-COR, Inc., Nebraska, USA) with a flow rate of 1 L min⁻¹. An air temperature and relative humidity sensor (BME280, Bosch Sensortec GmbH, Reutlingen, Germany) was placed inside the chamber while a photosynthetically active radiation (PAR) sensor (PQS1, Kipp & Zonen, Delft, Netherlands) was placed above the chamber. The chamber also had a small fan to ensure air mixing in the chamber headspace. A short clear chamber of 0.072 m³ was used when vegetation was low (beginning of June and after harvest), while a tall clear chamber of 0.288 m³ was used when vegetation was high (end of June until harvest). On each plot, a metal frame of 60 by 60 cm with a grooved edge was installed after sowing, including 4 rows of barley. During each measurement, the grooved edge was filled with water to ensure an air-tight seal between the chamber and the frame. To build the light response curve of each plot, NEE measurements were conducted in 5 different light conditions using netted or opaque fabrics (representing 100 %, 60 %, 36 %, 21.6 % and 0 % of the measurement day sunlight level). The closure time for each measurement was 3 min and the chamber was ventilated between each measurement. Similarly to Trémeau et al. (2024), NEE fluxes were calculated using an exponential fit:

$$F = \left(\frac{dC(t)}{dt} \right)_{t=0} \times \frac{M_{CO_2} \times P \times V}{R \times T \times A_b} \quad (1)$$

where $\left(\frac{dC(t)}{dt} \right)_{t=0}$ is the time derivative (ppm s⁻¹) of the exponential regression, M_{CO_2} is the molecular mass (44.01 g mol⁻¹), P is the air pressure (Pa), R is the universal gas constant (8.31446 J mol⁻¹ K⁻¹), T is the temperature inside the chamber headspace, V is the volume (m³), and A_b the base area (m²) of the chamber headspace. Missing air temperature and atmospheric pressure were retrieved from the FBES/OEB/SenPEP weather station located 230 m away from our site (Aphalo, 2023).

Normalised root mean squared error (NRMSE) of the exponential fit was used to filter the fluxes. When $\left(\frac{dC(t)}{dt} \right)_{t=0}$ was above 0.1 ppm s⁻¹ and NRMSE was above 0.01, NEE fluxes were discarded. The NRMSE filter was not applied for small fluxes with $\left(\frac{dC(t)}{dt} \right)_{t=0}$ below 0.1 ppm s⁻¹.

2.1.2. Vegetation index measurements

The Green Area Index (GAI) was measured within the same collar and week of NEE flux measurements. A protocol was created based on the Integrated Carbon Observation System (ICOS) ancillary vegetation measurements protocol for Croplands (Gielen et al., 2018).

Barley was sown in rows. Therefore, the GAI of barley (GAI_{barley}) was measured as follows:

$$GAI_{barley} = \frac{(n_{GLb} \times L_{Ab} + n_{GSb} \times S_{Ab}) \times n_{tot-row}}{A_c} \quad (2)$$

where n_{GLb} is the average number of green barley leaves per row, n_{GSb} is the average number of green barley stems per row, L_{Ab} is the average hemi-surface area per green barley leaf, S_{Ab} is the average hemi-surface area per green barley stem, $n_{tot-row}$ is the number of rows in the metal frame (4 in most cases, sometimes 3.5 due to unsuccessful germination), and A_c is the collar area (0.36 m²). n_{GLb} and n_{GSb} were determined by counting the total number of green leaves and green stems from the two middle rows and then averaged to one row. L_{Ab} and S_{Ab} were determined by harvesting 3–5 stems and 7–10 leaves from outside the collar and scanned with a planimeter (LI-3000C with LI-3050C, LI-COR, Inc., Nebraska, USA).

2.1.3. Yield measurements

The barley dry yield was estimated by harvesting the central strip (2 m wide) of the plot, where the area had reduced edge effects and was not affected by measurements.

2.2. STICS crop model

STICS (Scientific, Technical and Interdisciplinary simulator of soil-Crop System functioning) is a daily time-step process-based crop model that allows for variations in climate, soil, and crop systems to be simulated, covering plant growth, as well as fluxes of water, carbon, and nitrogen. STICS is able to model single species plots or intercrops of two species growing in the same plot. Crop growth is driven by radiation intercepted by the foliage, with biomass allocations varying according to the phase of development. Stresses reduce growth, with STICS accounting for nitrogen, water, waterlogging, and thermal stresses. Soils interact with the crops through the roots, which are specified as a root density distribution within the soil. Heterotrophic respiration is modelled from mineralisation of crop residues and humus.

We worked with JavaSTICS v1.41 and STICS v9.2 through the SticsOnR R package (Lecharpentier et al., 2023). Where STICS did not have a specific parametrisation for a species used in the field experiment, we substituted a similar plant's parameter set: the three clovers (*Trifolium* spp.) were all modelled with a crimson clover plant parameter set; chicory used the vetch parameter set.

Intercropping in STICS is built on a simplified approach in which the sole crop modelling process is extended to apply to two species. The modelled crops interact only through adaptations to the resource capture modules, simulating competition for light, water, and nitrogen (Brisson et al., 2004; Corre-Hellou et al., 2009). Radiation interception is differentiated for the dominant canopy and an understory canopy which is split into sunlit and shaded parts. The light partition and the associated microclimates drive the differences in water and nitrogen budgets as well as growth.

2.3. Experimental data usage

The collaborative nature of the experiment resulted in intense measurement campaigns occurring during the summer 2020 growing period (late May to mid September), while fewer measurements were taken in 2019, 2021, and 2022. Since STICS directly models systems including only one or two crops we chose to calibrate using observations from the single-cropped barley and intercrops of barley with one undersown species.

In order to calibrate STICS with the measurements from TWINWIN, we had to ensure that the observations from the field and the state variables available from STICS were comparable. Specifically, we sought to compare the model's yield (t ha⁻¹), NEE (t CO₂ ha⁻¹ d⁻¹), and LAI (m² m⁻²) estimates to the available observations from the field. We also restricted the observations used in calibration to only those taken between barley sowing and harvest since we are calibrating just the barley parameters.

While there were plenty of good observations in 2020, the other years presented a variety of problems. In 2019, the undersown crops

(that is, everything but barley) established poorly, and so the actual presence of an intercrop situation varied widely and modelling may have assumed much better crop establishment than occurred. 2021 had fewer observations, and there was a local drought in Helsinki (Ahongshangbam et al., 2023), leading to very low yields. In 2022, few measurements were taken in total and the final harvest was delayed, leading to improbably small yields (most being 10 % of the size or less than the poor harvest of the year before in 2021). Due to these circumstances, we have restricted the data in use to be solely 2020 for calibration and 2021 for validation.

Several issues with the simulations of the intercrop of barley and timothy grass in STICS were noticed, where the model produced quite unrealistic estimates that were not usable in calibration. From discussion with the developers of STICS, these issues appeared to stem from multiple bugs in the model to be fixed in an upcoming model version which focuses on intercropping (Vezy et al., 2023). As a result, the intercrop of barley and timothy grass was excluded from simulation and calibration.

2.3.1. Yield

The measured dry yield was compared with `mafruit`, the dry biomass of harvested organs (t ha^{-1}). The mean of observed yields for each plot type was used as the observed value, with uncertainty given by the standard deviation of the measurements. There was no accounting for possible losses due to delayed harvests.

2.3.2. Net ecosystem exchange (NEE)

NEE estimates from STICS are derived from the difference of two output variables: daily growth rate (`d1_tams`, $\text{t ha}^{-1} \text{day}^{-1}$), corresponding to net primary production, and soil respiration (`CO2sol`, $\text{kg ha}^{-1} \text{day}^{-1}$), corresponding to heterotrophic respiration as shown in Equation (3). NEE is calculated as:

$$NEE = \frac{\text{CO2sol}}{1000} - 0.48 \cdot \frac{44}{12} \sum_{\text{plants}} \text{d1_tams}, \quad (3)$$

where the factor in front of the summation convert from total mass of daily growth to the corresponding mass of CO_2 absorbed from the atmosphere. The chamber observations from the plots needed to be scaled up to a daily time-step estimated NEE. The chamber measurements under varying levels of PAR were used to build the light response curve (Ruimy et al., 1995). The apparent quantum yield α was allowed to vary between -0.005 and $-0.00001 \text{ mg } \mu\text{mol}^{-1}$ and the asymptotic gross photosynthesis rate GP_{max} was allowed to vary between -5 and $-0.00001 \text{ mg m}^{-2} \text{ s}^{-1}$. These were used in Equation (4) with half-hourly average PAR measurements from the nearby Kumpula weather observation station (Finnish Meteorological Institute, 2024, $\text{N}60^{\circ}12'14.0''$, $\text{E}24^{\circ}57'38.9''$) to produce half-hourly estimated gross primary production and subsequently summed to daily gross primary production (GPP) for each plot. Note that the unit conversions are excluded from this formula for simplicity, but may be seen in the code appendix.

$$GPP = \sum_{\text{day}} \frac{\alpha \cdot GP_{\text{max}} \cdot \text{PAR}}{\alpha \cdot \text{PAR} + GP_{\text{max}}} \quad (4)$$

Finally, the fully shaded chamber measurement was used to create an approximate daily soil respiration measurement by simple conversion to a full-day value; the difference of these two terms then produced the NEE. The uncertainties associated with the chamber measurement-derived values (σ_{α} , $\text{mg } \mu\text{mol}^{-1}$; $\sigma_{GP_{\text{max}}}$, $\text{mg m}^{-2} \text{ s}^{-1}$) were used in corresponding uncertainty propagation formulae to produce uncertainties for the observed daily GPP values ($\sigma_{GPP, \text{days}}$, $\text{t CO}_2 \text{ ha}^{-1} \text{ d}^{-1}$). This naively assumes that the PAR measurements contain no variation or uncertainty for simplicity.

$$\sigma_{GPP, \text{day}} = \sqrt{\sum_{\text{day}} GPP^2 \cdot \left(\frac{\sigma_{\alpha}^2}{\alpha^2} + \frac{\sigma_{GP_{\text{max}}}^2}{GP_{\text{max}}^2} + \frac{\text{PAR}^2 \cdot \sigma_{\alpha} + \sigma_{GP_{\text{max}}}}{(\alpha \cdot \text{PAR} + GP_{\text{max}})^2} \right)} \quad (5)$$

The NEE uncertainty, σ_{NEE} , is then derived by combining the GPP and ecosystem respiration (σ_{Reco}) uncertainties:

$$\sigma_{NEE} = \sqrt{\sigma_{GPP, \text{day}}^2 + \sigma_{Reco}^2} \quad (6)$$

Our calibration method did not perform well with positive values of NEE in the middle of the growing season. Calibrating with such observations led to clear distortions in estimated NEE and decreases in estimation performance. Calibrating soil respiration terms may have alleviated these issues, however, calibrating soil-related parameters limits the applicability of the calibration and the transferability would be reduced. In our calculation of NEE from STICS (equation (3)) the only positive values come from `CO2sol`, which should not be directly affected by the parameters we calibrated. To mitigate this we removed calculated NEE observations that were above zero, and only calibrated with the remaining non-positive observations. This exclusion of positive NEE leads to between 2 and 6 observations of NEE being used to calibrate each plot type.

2.3.3. Leaf/green area index (LAI/GAI)

Finally, leaf area index (LAI) in STICS, `lai_n` ($\text{m}^2 \text{ m}^{-2}$), represents the green area, which is why we use the green area index as a direct comparison. However, the green area index measurements do not include uncertainty estimates, and for inter-cropped plots only one measurement is available for a given date. Hence, an uncertainty estimate must be created for the inter-cropped plots in order to calibrate STICS using 4DnVar. The single-cropped barley was measured in multiple plots for each date, thus we take the means of the single-cropped barley green area indices ($GAI_{SC, \text{Barley}}$) as the single-cropped barley measurements for each date, and the standard deviations ($\sigma_{GAI, SC, \text{Barley}}$) as the uncertainties. For the inter-cropped plots, the corresponding single-cropped barley measurement uncertainties were used to create proportional uncertainty estimates for the dual-crop plots.

$$\sigma_{GAI, IC, \text{Barley}} = GAI_{IC, \text{Barley}} \cdot \frac{\sigma_{GAI, SC, \text{Barley}}}{GAI_{SC, \text{Barley}}} \quad (7)$$

2.4. 4DnVar

The 4DnVar data assimilation technique can be used to estimate a set of model parameters as was seen in Pinnington et al. (2020). The 4DnVar method is centred on the minimisation of a cost function constructed from the two sources of information known prior to its use: (1) some prior knowledge of the parameters, and (2) the observational data, along with their respective uncertainty information. An ensemble of parameters is generated by sampling from parameter space and mapped to observation space via an observation operator encompassing model runs. The ensemble in observation space is used to approximate cost function gradient terms thus avoiding the calculation of an adjoint as would be required when using gradient descent methods with traditional 4DVar (Rawlins et al., 2007). Use of an ensemble also allows us to obtain posterior uncertainty information via a posterior error covariance matrix and a further benefit to implementing 4DnVar is that, once an ensemble of model runs is obtained, multiple experiments may be performed, say with new observational data, without the need to repeat the process of ensemble generation.

2.4.1. Nomenclature

In this paper, we use prior to mean the default calibration of STICS, and construct a prior ensemble of parameters based on this, which when given to STICS produces a prior ensemble in observation space; we use posterior to mean the set of outputs from 4DnVar (including both an

ensemble in parameter space and the optimal parameter set, also called the analysis vector) and the posterior ensemble in observation space.

In the context of parameter estimation, $\mathbf{x} \in \mathbb{R}^{N^p}$ represents a set of N^p parameters. The parameters are assumed to be normally distributed with known mean $\bar{\mathbf{x}}_b$ and prior error covariance matrix \mathbf{B} . An ensemble of size N^e is drawn: $\{\mathbf{x}_i, i = 1, \dots, N^e\}$ and perturbation matrix $\mathbf{X}' = (\mathbf{x}_1 - \bar{\mathbf{x}}_b | \mathbf{x}_2 - \bar{\mathbf{x}}_b | \dots | \mathbf{x}_{N^e} - \bar{\mathbf{x}}_b) \in \mathbb{R}^{N^p \times N^e}$ is constructed, where $\bar{\mathbf{x}}_b$ is the mean of the ensemble. The optimal parameter set $\bar{\mathbf{x}}_a$ is assumed to be an update to the ensemble mean in the form of a weighted combination of the perturbations:

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_b + \mathbf{X}'\mathbf{w} \quad (8)$$

where $\mathbf{w} \in \mathbb{R}^{N^e}$ is normally distributed with zero mean and unit covariance. The observations at time t are stored in a vector $\mathbf{y}_t \in \mathbb{R}^{N_t^o}$ where N_t^o is the number of observations available at time $t \in \{0, \dots, T\}$ and so $N^o = \sum_{t=0}^T N_t^o$ represents the total number of observations. The corresponding observation error covariance matrix at time t is given by $\mathbf{R}_t \in \mathbb{R}^{N_t^o \times N_t^o}$. An observation operator $\mathbf{h}_t : \mathbb{R}^{N^p} \mapsto \mathbb{R}^{N_t^o}$ is then required to map parameter space to observation space in order to compare the measured observations with those as simulated by the parameters at time t .

2.4.2. 4DnEnVar cost function and gradient

The 4DnEnVar cost function to be optimised for the ensemble perturbation weighting \mathbf{w} and corresponding gradient function are given by

$$J(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{1}{2}(\hat{\mathbf{h}}(\bar{\mathbf{x}}_b) + \mathbf{Y}'\mathbf{w} - \hat{\mathbf{y}})^T \hat{\mathbf{R}}^{-1} (\hat{\mathbf{h}}(\bar{\mathbf{x}}_b) + \mathbf{Y}'\mathbf{w} - \hat{\mathbf{y}}) \quad (9)$$

and

$$\nabla J(\mathbf{w}) = \mathbf{w} + (\mathbf{Y}')^T \hat{\mathbf{R}}^{-1} (\hat{\mathbf{h}}(\bar{\mathbf{x}}_b) + \mathbf{Y}'\mathbf{w} - \hat{\mathbf{y}}), \quad (10)$$

respectively, where

$$\hat{\mathbf{y}} = \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix} \in \mathbb{R}^{N^o}, \quad \hat{\mathbf{h}}(\bar{\mathbf{x}}_b) = \begin{pmatrix} \mathbf{h}_0(\bar{\mathbf{x}}_b) \\ \mathbf{h}_1(\bar{\mathbf{x}}_b) \\ \vdots \\ \mathbf{h}_N(\bar{\mathbf{x}}_b) \end{pmatrix} \in \mathbb{R}^{N^o},$$

$$\hat{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_0 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_1 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}_N \end{pmatrix} \in \mathbb{R}^{N^o \times N^o},$$

and

$$\mathbf{Y}' = (\hat{\mathbf{h}}(\mathbf{x}_1) - \hat{\mathbf{h}}(\bar{\mathbf{x}}_b) | \hat{\mathbf{h}}(\mathbf{x}_2) - \hat{\mathbf{h}}(\bar{\mathbf{x}}_b) | \dots | \hat{\mathbf{h}}(\mathbf{x}_{N^e}) - \hat{\mathbf{h}}(\bar{\mathbf{x}}_b)) \in \mathbb{R}^{N^o \times N^e}$$

is the perturbation matrix in observation space approximating the derivative information.

2.4.3. Computing the posterior ensemble

Uncertainty information for the optimised parameter set can be acquired by updating the ensemble according to the following:

$$\mathbf{X}'_a = \mathbf{X}'_b \left(\mathbf{I} + [\mathbf{Y}'_b]^T \hat{\mathbf{R}}^{-1} [\mathbf{Y}'_b] \right)^{-1/2} \quad (11)$$

where \mathbf{X}'_a are the updated ensemble perturbations. This expression is obtained from an Ensemble Kalman Filter (Evensen, 2003) result and $\mathbf{X}'(\mathbf{X}')^T$ approximates the posterior error covariance matrix.

2.5. Calibration

The calibration is intended to capture inter-species dynamics not currently captured by STICS. To contrast with the conventional single-crop calibration approach and control for possible dynamics dependent on the location and characteristics of the field, we calibrate barley grown alone in addition to barley undersown with each of the different secondary species. We label calibration of the barley grown alone as "Barley", while calibrations specific to each undersown species, in which the measurements from the same plot type are used to calibrate, is labelled "Self". The aim is to produce individual parameter sets for both barley grown alone and barley in the presence of different undersown species that indirectly capture these dynamics. Here, we calibrate barley functional parameters and do not estimate the undersown plant parameters.

When discussing the model set up and associated plot type, we frequently refer to just the species undersown with barley as the barley's presence is constant across this study. As an example, FA LAI or FA yield refer to the LAI or yield of barley undersown with tall fescue (*F. arundinacea*, FA). FA LAI or yield does not include the LAI or yield of tall fescue, just barley. However, the NEE includes contributions from both barley and any undersown species, as well as soil respiration.

2.5.1. Parameter responses

From the set of plant parameters implemented in STICS v9.2, we removed all code-switching parameters, unused parameters, and parameters at the edge of the valid ranges specified in the STICS documentation. We further narrowed the set of parameters down to 37 parameters to be considered based on the expert opinion of the authors of the manuscript.

Using this reduced set, a one-at-a-time sensitivity analysis was conducted. To test sensitivity, the parameters were sampled around their corresponding preset barley (cork variety) values in STICS by increasing and decreasing the preset value by 5 %, 10 %, 15 %, 20 %, 30 %, and 40 %. Then, values of parameters that did not lie within the valid parameter range according to the JavaSTICS documentation were discarded. All the remaining parameter values were supplied to STICS one at a time with all other parameters held constant at their preset values. The model outputs of LAI, NEE, and yield were then used to assess sensitivity by separately considering final yield, mean LAI, and mean NEE. The means are calculated over the entire simulation period.

The parameter set under consideration was further reduced to the apparent most sensitive parameters, with the remaining parameters being individually assessed with an interactive tool (provided in the code appendix), comparing the model outputs of LAI, NEE, and yield with varied parameters to the baseline parameter outputs. By this method, parameters showing only minor influence or overly one-sided effects on the outputs were again discarded, until a parameter set that was feasible for assessing the data assimilation process was attained, consisting of the eight parameters: adens, efcroiveg, vitircarb, INNmin, stlevdrp, efcroirepro, vlaimax, and dlaimax.

2.5.2. Twin experiments for parameter selection

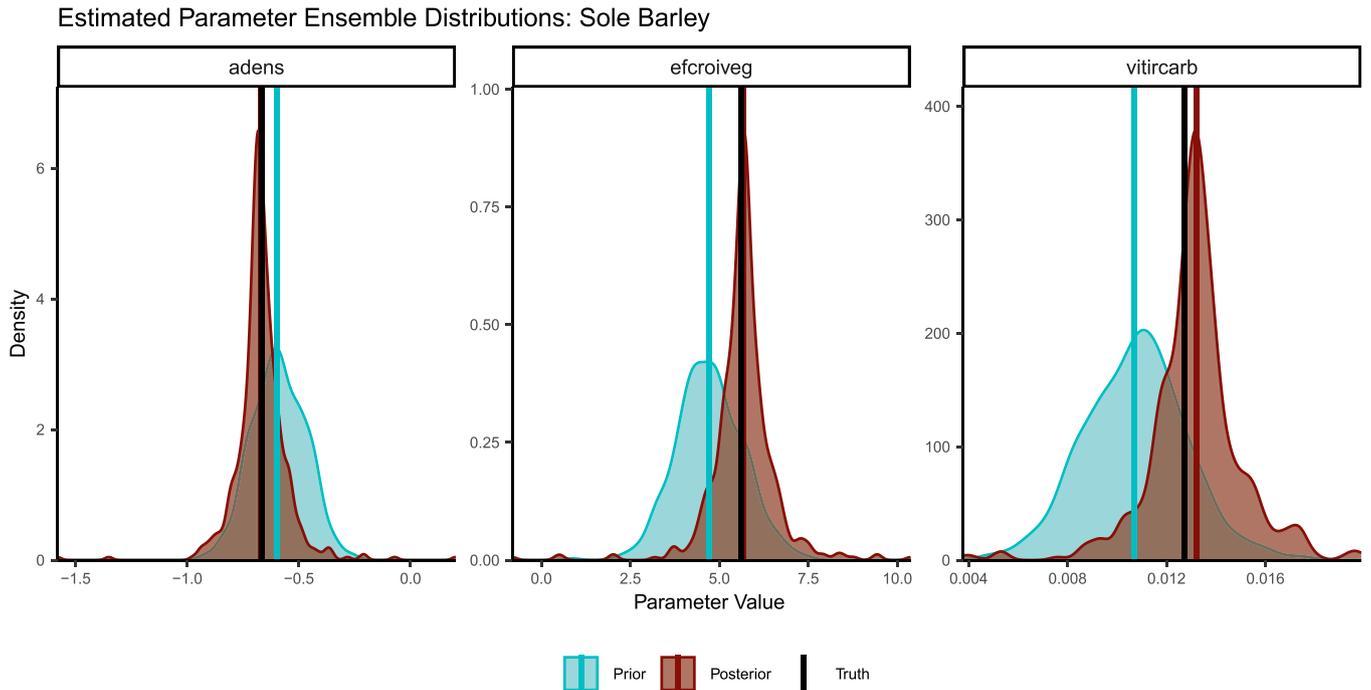
We examined the effectiveness of our data assimilation process with twin experiments similar to Pinnington et al. (2020). In particular, we evaluated how many parameters were able to be estimated while avoiding problems of equifinality. This involved using a synthetic truth, wherein a fixed set of parameter values was used to generate synthetic observations using STICS, which was then used in the data assimilation process to estimate the generating parameters. The function and the limitations of the process are thus tested, as STICS itself generated the observations there should not be any representation error affecting the calibration. For initial testing, eight parameters were selected, as described in the previous section.

The synthetic truth parameters were drawn from normal distributions with means given by the preset parameter values for barley (var.

cork) present in STICS and standard deviations of 10 % of the absolute value of the respective means. Model projections of LAI, NEE, and yield for each plot type were generated using combinations of varied parameters, this allowed comparison of both parameter-specific calibration performance and the number of parameters being calibrated together. The synthetic observations are then those projections with an associated uncertainty of 1 % of the mean of the observation type.

The prior parameter ensemble was generated by drawing parameter values similarly to the synthetic truth parameter vector, however the normal distributions had a standard deviation of 20 % of the absolute value of their respective means; draws were repeated up to the size of the ensemble. The posterior parameter ensemble generated by 4DnVar was compared with the prior parameter ensemble, and a posterior ensemble of estimates in observation space was generated for comparison with

(a) Parameter distributions, three parameter calibration



(b) Parameter distributions, four parameter calibration

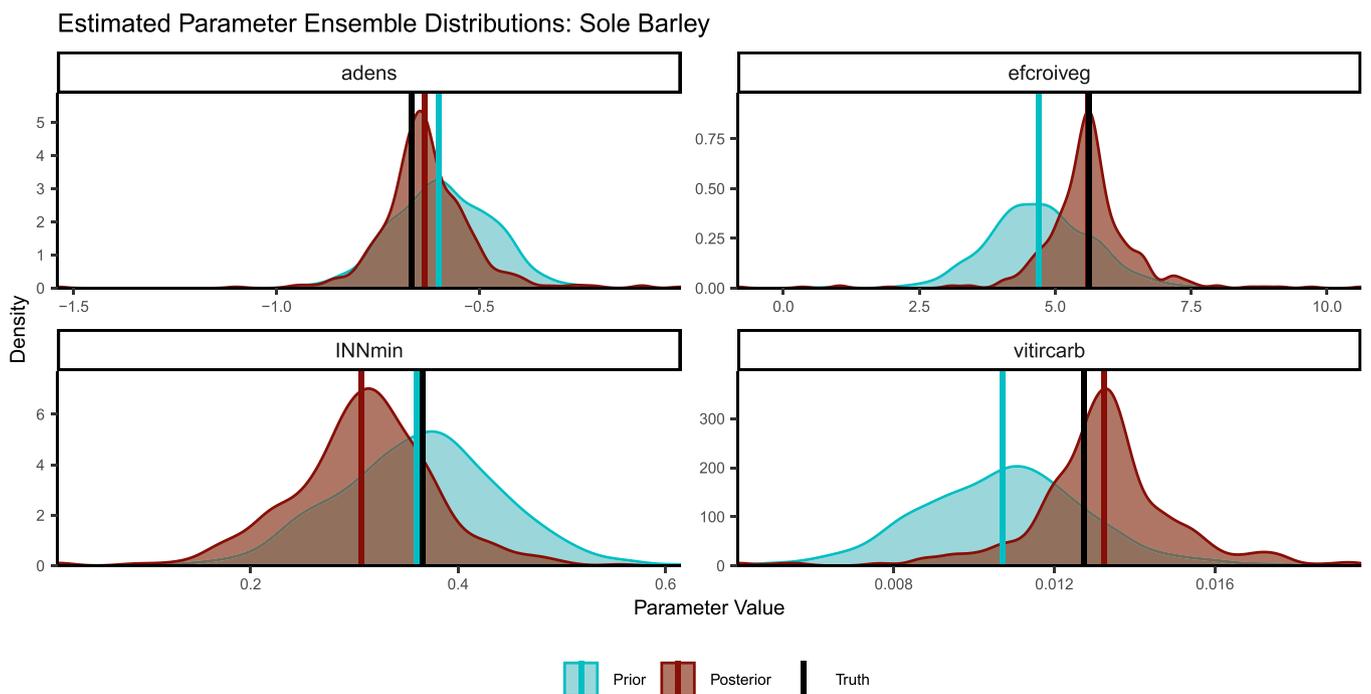


Fig. 1. Parameter distributions when calibrating sole barley with the shown parameters using a synthetic truth. Due to uncertainty about the reliability of retrieval of more than three parameters we chose to calibrate only the three parameter set shown in 1a.

both the prior ensemble estimates and the synthetic observations.

Initial twin experiments and variations were conducted using a single fixed synthetic truth, before testing alternate synthetic truths to ensure the robustness of the observed behaviours. Similarly, the number of ensemble members was varied to assess the effect of ensemble size on the results and to establish a suitable compromise between speed of estimation and coverage of the parameter space. The twin experiment was run with ensemble sizes of 50, 100, 200, 250, 500, and 1000. The initial exploratory selection was conducted with sole barley, and other plot types were checked for similar behaviour.

2.5.3. Calibrating for the TWINWIN site

Finally, the calibration was performed using the observations from the TWINWIN plots. A reduced parameter set and appropriate ensemble size were chosen based on the twin experiments, and a prior ensemble of parameter values were generated in the same way as the twin experiments.

3. Results

3.1. Twin experiments for parameter selection

The twin experiment showed a lack of parameter convergence to the synthetic truth when varying all eight parameters chosen by the sensitivity tests, even with larger ensemble sizes. Closer examination of the parameter responses, and trials excluding groups of parameters at a time led to a final parameter set containing *adens*, *efcroiveg*, and *vitircarb* that consistently retrieved the synthetic parameter values in calibration. If one of the other parameters was included in the calibration then there were issues – that parameter was often not retrieved and the posterior distributions of the other parameters were broader and the mean further from the true parameter value. In Fig. 1, the posterior parameter estimate of *INNmin* had moved away from the synthetic truth, and the posterior distributions of the other parameters in this calibration were somewhat broader than those in Fig. 1a. Similarly, if we replaced one of the three chosen parameters with another, the performance worsened. Some parameters were excluded at this stage based on weaker responses in the observed variables, or one-sidedness in the responses. Other parameters appeared to have comparable effects on the observed variables to another parameter used in calibration. For example, *vitircarb* and *efcroirepro* had very similar effects, largely affecting the yield, thus calibrating them together demonstrated issues of equifinality. The choice then came down to relative performance, with *vitircarb* being better retrieved by the data assimilation process and those calibrated estimates better fitting the observations.

When varying the ensemble size used for calibration, we found that the prior distributions did not appear normally distributed for those ensembles containing 250 members or fewer, despite being sampled from normal distributions. For 500 and 1000 ensemble members the results were quite similar, thus we chose to proceed with 500 ensemble members when calibrating with the actual observations.

Table 3

Percentage error and improvement in parameter estimates in the twin experiment. The upper section is the 3 parameter case, the lower has 8 parameters varying.

Parameter	Truth	Prior	Posterior	% Error Prior	% Error Posterior	Percentage Improvement
<i>adens</i>	− 0.65	− 0.59	− 0.66	− 9.8	1.1	111
<i>efcroiveg</i>	5.7	4.7	5.8	− 16.8	1.6	110
<i>vitircarb</i>	0.01	0.011	0.01	4.5	1.6	64
<i>adens</i>	− 0.65	− 0.59	− 0.69	− 9.8	5.7	158
<i>dlaimax</i>	0.00041	0.00044	0.0004	6.7	− 2.5	137
<i>efcroirepro</i>	5.2	4.7	5.2	− 8.9	0.8	109
<i>efcroiveg</i>	5.7	4.7	5.6	− 16.8	− 0.9	95
<i>INNmin</i>	0.3	0.36	0.34	22.7	15	34
<i>stlevdrp</i>	710	660	690	− 7.3	− 3.1	58
<i>vitircarb</i>	0.01	0.011	0.011	4.5	8.8	− 98
<i>vlaimax</i>	1.9	1.8	1.9	− 5.9	− 2.5	57

With three parameters being varied, the performance was excellent, and with all eight being varied the performance was still very good – one parameter estimate was further from the truth in the posterior compared to the prior, but all other posterior parameter estimates were much closer to the true parameter values than the corresponding priors (Table 3). Unfortunately, this good performance in the eight parameter case was not robust to alternate synthetic truths, it was more common to see multiple parameters moving away from the truth in the posterior estimates.

The three parameter calibration reproduced the synthetic observations well, showing improvement relative to their priors (Table 3, and Figures A.6-A.9 and Table A.5 in the supplementary material). Calibrations involving more parameters, including all eight considered in this section, also reproduced the synthetic observations well, though not quite as well as the three parameter calibrations. However, their failure to retrieve the synthetic parameters led to the choice of the reduced set of three parameters for calibration with real observations. This maximises the ability to transfer our calibrated barley parametrisation to other locations by reducing the likelihood of equifinality having a significant effect on the calibration results.

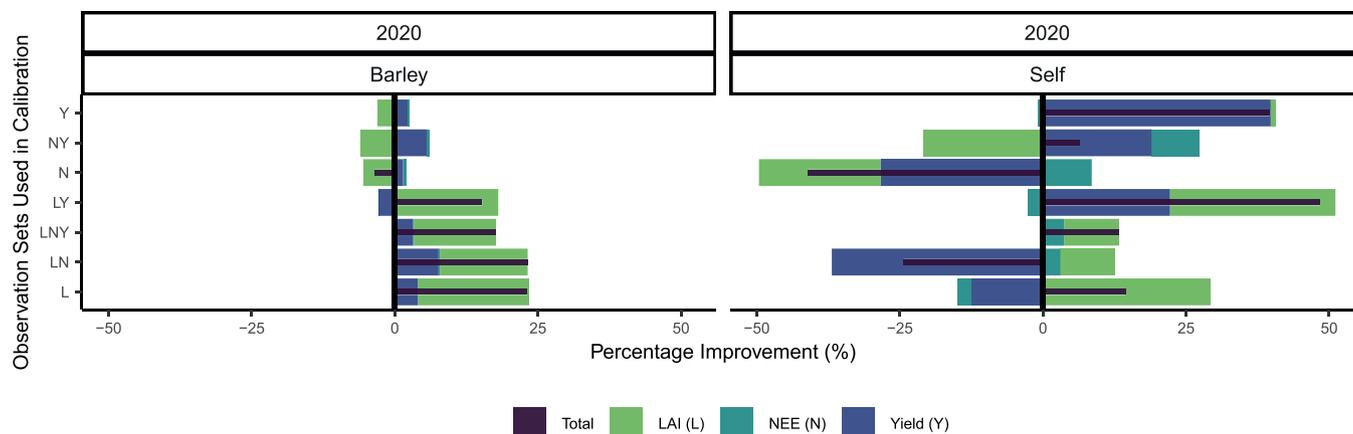
3.2. Calibrating for the TWINWIN site

Self-calibration performed better than using the sole barley calibration (Fig. 2; Table 4). Unsurprisingly, when considering the calibration year 2020, calibrating only with yield produced the best yield results, followed by calibrating with yield and LAI or yield and NEE, then calibrating with all three observation streams. However, the calibration with LAI and yield predicted 2021 yields worse than the prior. Calibrating with yield and NEE showed similar improvement relative to calibrating with yield alone.

Looking at the yield estimate performance in the validation year, 2021, we see that the self-calibrations had a mixed performance, while the sole barley calibrations did better (Fig. 2b, Table 4b).

Yields were generally well predicted (Fig. 3), with much lower yields in 2021 reflecting a harsher year captured in the estimates. In 2020, all LAI+Yield self-calibrations, except AC and CI, improved yield predictions as expected for estimates of observations in the calibration dataset. However, the poor estimation of AA and AC portray some underlying issues. That being said, the AA and CI plot types had the largest observed variance in yields and the estimated yields lied within a 95 % confidence interval of the observations. In 2021, four of posterior estimates were further from the observed yield than their respective priors. Furthermore, whilst the AA yield estimates improved, in both years they were large overestimates of the yield. However, of the other posterior estimates across both years, all but two clovers and AA in 2021 included their corresponding observations within the respective 95 % confidence intervals. Comparing the barley and self-calibration, the barley calibration had a noticeable, albeit not significant, flattening effect, bringing all the posterior estimates closer together, whereas the self-calibration appeared to capture some of the differences between plot

(a) Percentage improvement in RMSE in calibration dataset



(b) Percentage improvement in RMSE in validation dataset

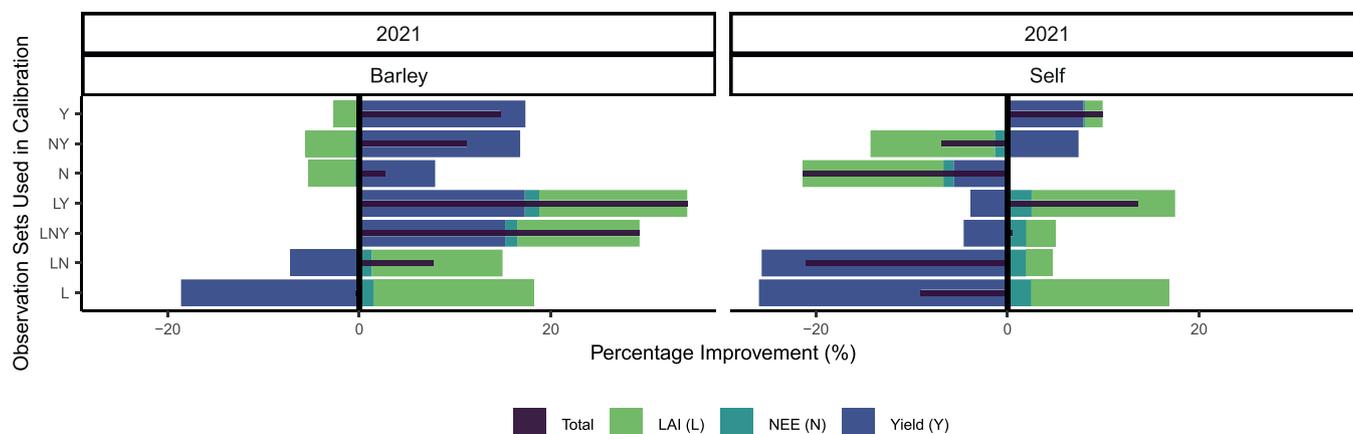


Fig. 2. Percentage improvement in RMSE of the posterior relative to the prior under different calibration scenarios, namely calibration specific to each plot type (Self, right) or calibration with sole barley applied to all plot types (Barley, left). All rows are averages over the full set of plot types. 0 implies no improvement, 100 implies no error (perfect prediction, that is, posterior RMSE = 0), and negative values imply worsened predictions.

Table 4

Relative root mean squared error (RRMSE) for calibrations using LAI+yield observations from TWINWIN. In the calibration column Prior refers to STICS default parameter values, Barley refers to the posterior parameter set estimated by barley single-crop calibration, and Self refers to the posterior parameter set estimated by intercrop calibration.

Observation	Calibration	RRMSE for the calibration data.							
		Barley	AA	AC	CI	FA	IR	RC	WC
GAI	Prior	0.59	0.68	0.68	0.68	0.59	0.92	0.8	0.81
GAI	Barley	0.3	0.72	0.87	0.5	0.48	0.78	0.6	0.65
GAI	Self	NA	0.67	0.61	0.43	0.51	0.66	0.47	0.51
NEE	Prior	-111	-2.17	-1.43	-1.42	-31.46	-6.95	17.9	-141
NEE	Barley	-117	-2.02	-1.19	-1.47	-33.9	-7.37	17.4	-136
NEE	Self	NA	-2.2	-1.04	-1.48	-35.1	-7.97	16.7	-138
yield	Prior	0.11	0.99	0.16	0.12	0.29	0.1	0.05	0.21
yield	Barley	0.02	0.76	0.32	0.04	0.36	0.22	0.16	0.29
yield	Self	NA	0.8	0.42	0.15	0.06	0.01	0.01	0.07
Observation	Calibration	RRMSE for the validation data.							
		Barley	AA	AC	CI	FA	IR	RC	WC
GAI	Prior	0.42	1.03	0.77	1.07	0.62	0.78	1.06	1.21
GAI	Barley	0.49	0.82	0.59	0.87	0.39	0.52	0.88	1.04
GAI	Self	NA	1.02	0.89	0.79	0.44	0.3	0.78	0.94
NEE	Prior	-10.43	1.46	-1.04	3.52	-24.9	-1.54	-25.3	3.79
NEE	Barley	-9.2	1.47	-1.06	3.72	-24.1	-1.47	-24	3.82
NEE	Self	NA	1.46	-1.04	3.79	-24.3	-1.41	-23.4	3.71
yield	Prior	0.18	2.23	0.12	0.04	0.16	0.15	0.5	0.56
yield	Barley	0.03	1.89	0.21	0.1	0.26	0.01	0.28	0.42
yield	Self	NA	1.96	0.06	0.3	0.05	0.25	0.56	1.21

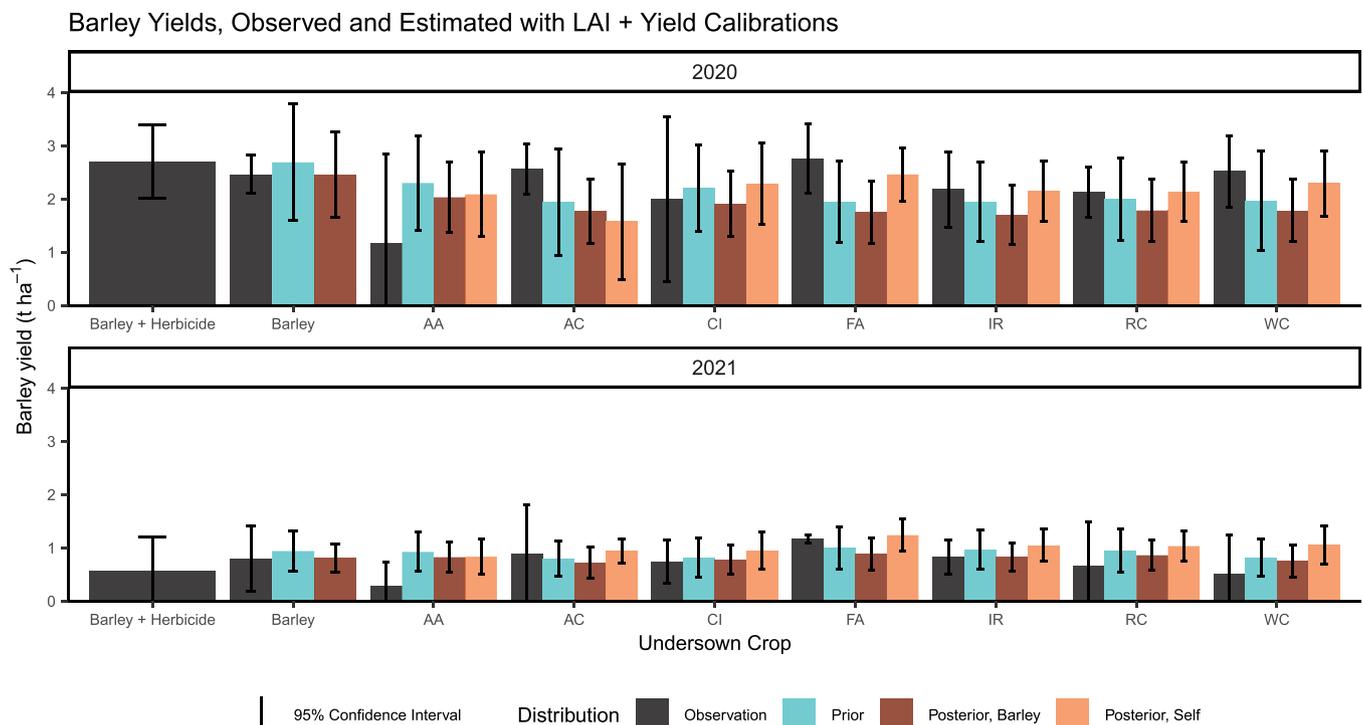


Fig. 3. Yield observations, prior estimates, and posterior estimates using the LAI+Yield calibrations. 2020 is the calibration data set, 2021 is the validation dataset. The x-axis label “Barley” refers to the monocropped barley plot type. The species acronyms on the x-axis refer to the undersown species listed in Table 2. For example, the values at AA are barley yields from plots of barley sown with alfalfa.

types with different undersown species.

The single-crop barley with herbicide was included in the yield comparison of Fig. 3 as a more conventional approach, as none of the plots with undersown species, nor the “Barley” plot type, were treated with herbicides. Barley grown with herbicide yielded marginally more than almost all other plot types in 2020, but had a much lower yield in 2021, slightly lower than most other plot types. However, in neither case were there statistically significant differences, as the largest changes in mean observed value were accompanied by larger uncertainty.

Overall, the yearly estimated NEE (not shown) was highest for the barley inter-cropped with clovers or italian ryegrass, and lowest for barley inter-cropped with alfalfa, followed by solo barley. However, calibration did not show notable changes in these yearly estimates.

Yields remained underestimated in several of the inter-cropped barley plot types in 2020; this was contrary to expectation as the harvest itself was late, delayed by other measurement campaigns in the project, so we had expected to overestimate yields.

However, the LAI performance when calibrating with yield and LAI struggled to capture the level in the early growing season, as seen in Figure A.10, where the second and third observation was notably higher than the estimated values. Similar issues of notably poorer NEE estimates were not observed in our calibrations, as indicated by the relatively small changes in NEE RMSE in Fig. 2 or visible in Fig. 4.

The parameter distributions when calibrating sole barley (Fig. 5) exhibited a clear dominance of LAI measurements in determining the posterior estimate of $adens$, NEE data in determining $efcroiveg$, and a yield dependence influenced by both LAI and NEE data streams determining the posterior distributions of $vitircarb$. This last remark is particularly interesting, as the LAI and NEE data streams barely affected $vitircarb$ when used without yield measurements, but showed distinct differences when the yield data stream was included.

4. Discussion

4.1. 4DnVar as a calibration method for crop models

Our twin experiment results highlight aspects that should be addressed in an implementation expanding the temporal breadth of observations. When doing the test with synthetic data representing an observation frequency matching the real situation, we see that, at least for this system, there is a hard limit on the number of dominant parameters that could be calibrated before the estimated values begin to diverge from the known true parameters. Thus, there are questions around whether these new additional parameters could be simply added to the calibration, especially considering the imbalanced representation of the different years in the data. A potential solution is, if the dominant parameters are sufficiently isolated for the different conditions, in other words, if their impact on the system is minor outside the specified condition, then that would support doing a step-wise calibration (Tarantola, 2005; Guillaume et al., 2011). There, instead of calibrating the system as a whole over different conditions, a first set of parameters is calibrated under conditions where they are dominant, then set as constant for calibrating another set of parameters under conditions where these new parameters are dominant. This approach, however, operates under the assumption that the parameter sets can be considered independent. Regardless of the chosen approach, what these results argue is that while the sensitivity analysis is often used to select the parameters to be calibrated (Mathers et al., 2023; Ruget et al., 2002), it is also important to then do a twin experiment to ascertain how many parameters can actually be calibrated reliably.

Another interesting feature of the results is the response to the different combinations of data streams. When looking at the average improvements in Figure 2, it is evident that data selection for calibration depends completely on the output of interest as there aren't any cross-variable benefits shown there, as found by Guillaume et al. (2011). For example, when using both yield and LAI data in calibration, the results match the observed values better than if calibrated specifically to

Average estimated NEE, barley growing season 2020, Self calibrated

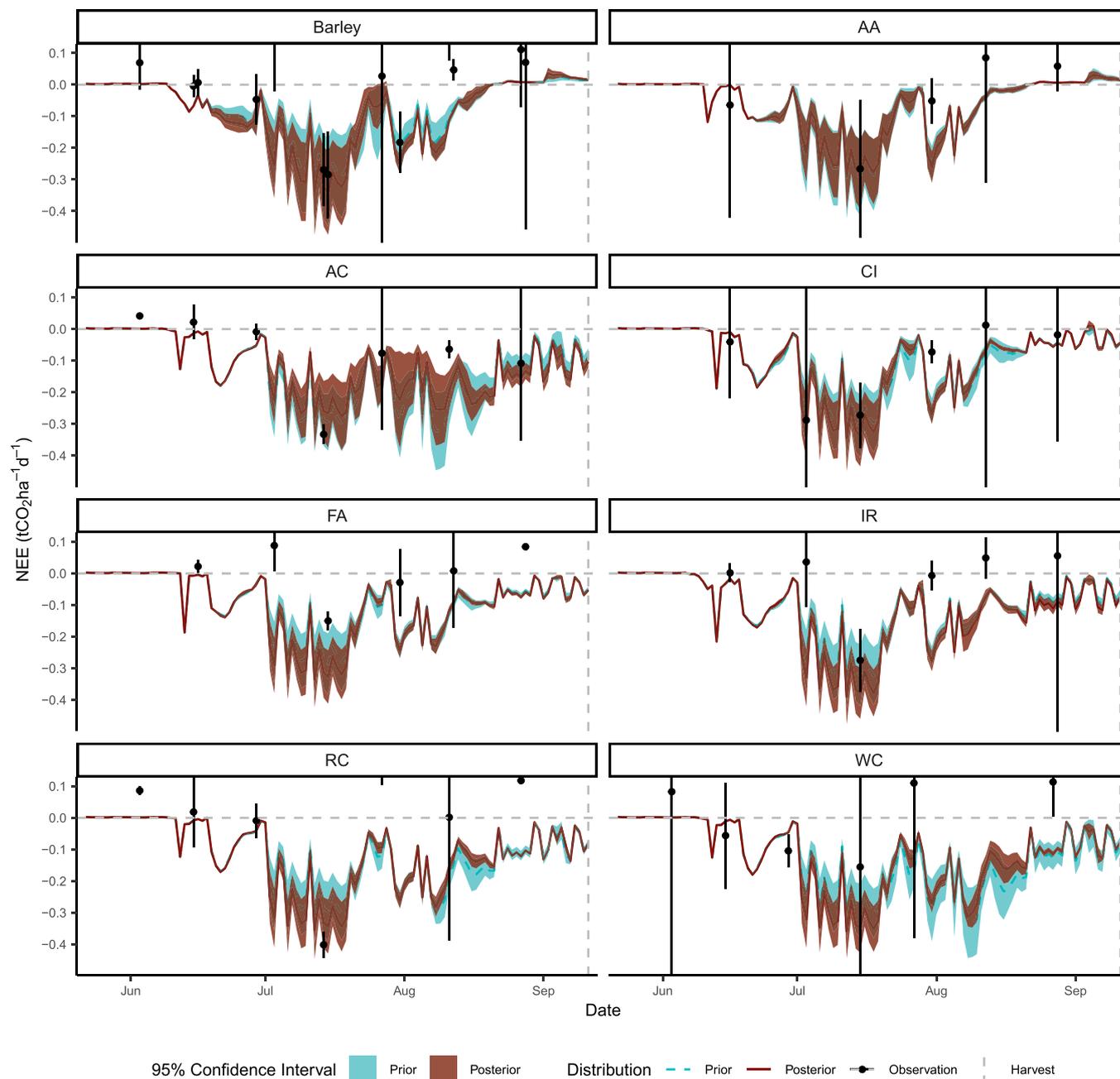


Fig. 4. Estimated NEE over the barley growing season of plots containing barley and the indicated undersown species before and after self-calibration of 3 parameters using TWINWIN yield and GAI measurements. The species acronyms are clarified in Table 2.

those variables individually. This is not necessarily beneficial, as the LAI observations are often used as indicators of other variables (Franklin et al., 1997; Demarty et al., 2007) and calibrating with LAI, or here GAI, alone causes a decrease in the yield and NEE projection performances. This directly implies that the different observation streams have conflicting constraints on the parameter values, as found by Ruget et al. (2002), which is most likely due to the number of associated parameters and equations in the model. In STICS, each of the examined outputs is determined by complex dynamics (Brisson et al., 2003; Kherif et al., 2022) and, consequently, changing the value of a tuned parameter will not have a straight-forward effect. For this reason, the issue of equifinality arises for a relatively low number of parameters.

The issue is even more pronounced with NEE outputs – there is no

meaningful improvement when it comes to the NEE observations themselves. Furthermore, examining how well the yield+NEE self-calibration performed with different secondary crops (Figures A.12–A.14), when the calibration with CI produces a large change in posterior NEE, there is also a large decrease in how well the yield posterior matches the measured yield. This outcome is assumed to be due to, not just the issues discussed before, but also partially because the NEE measurement time series is noisy and indirectly calculated from a group of shaded measurements, a radiation measurement from a nearby tower, and a single respiration measurement taken over a few minutes. A critical limitation in making NEE calibrations is that our calibration does not affect potential respiration – STICS does not calculate plant respiration, thus the calibration could only alter the NEE through varying the

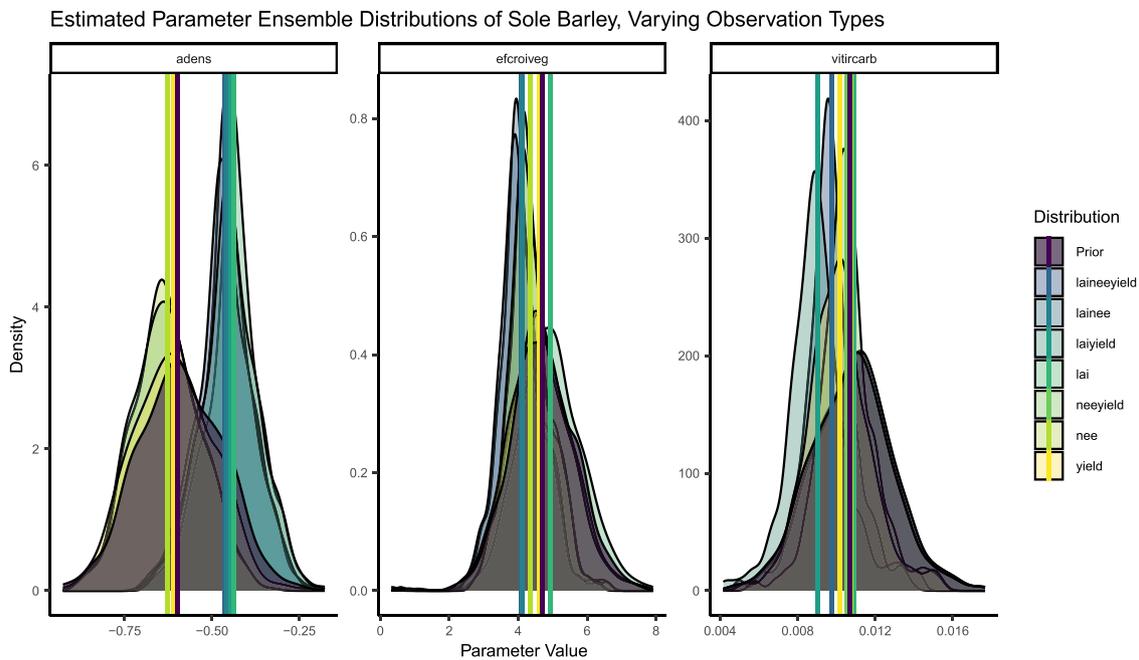


Fig. 5. Parameter distributions for sole barley before and after calibration using TWINWIN measurements. The label describes which observation data streams were used in calibration.

NPP. This restriction led us to exclude positive NEE values, as the model would not be able to reproduce them with the choices made concerning calibration. It may be that NEE could be better assimilated by this method if we include parameters affecting all contributing elements. It should be noted, however, that the calibration is still altering the parameter values as otherwise the other outputs wouldn't see such changes.

Overall, we found applying the 4DEnVar data assimilation technique useful for evaluating the effect of choices such as which and how many parameters to calibrate, or which data streams are used as observations in the calibration process. Of particular benefit was the ability to use an ensemble of prior model runs for the twin experiment as well as calibration with varying streams of real observations, thereby reducing computation time compared with alternative Bayesian calibration methods.

4.2. Calibrating for intercrop effects vs single-crop calibration

On average, the yield+LAI self-calibrations improved performance for yield and LAI for 2020 (the calibration dataset). However, those calibrations have little impact on the projected values for 2021 (the validation dataset), and are outperformed in 2021 by the monocropped barley calibration. What is crucial here, though, is that the changed parameter values do not have that much of a positive or a negative effect in 2021, which strongly implies that they are unimportant for that year. This system behaviour makes sense when considering that the environmental conditions limiting crop growth can change from year to year (Dzotsi et al., 2013), affecting the relative importance of the intercrop interactions, which the overall change in both observed and modelled results strongly suggests happened between 2020 and 2021. Indeed, June and July of 2021 had significantly lower precipitation and higher average temperatures compared to 2020 (Table 1), causing water stress during the majority of the growth period of the barley, which was then followed by a very wet August. Consequently, the results here still support the use of cash crop parameter calibration to reflect the secondary crop they are grown with, but the projections will only be improved in conditions similar to when the model was calibrated.

Expanding the calibration time window to include more varied annual dynamics is a logical suggestion to further improve the model

performance in different conditions. The primary reason we only used data from 2020 for calibration was due to the number of observations available compared to 2021. However, our twin experiment results discussed above highlight aspects that should be addressed in such an implementation.

Finally, when looking at how well the calibration performs with specific secondary crops, the variation across the species is reasonable considering how differently they are expected to impact the system (Valkama et al., 2015). Thus, the calibration should produce varying results for systems with different cover crops. What is even more important to note here is how much more variation there is within the measured values for the undersown crops compared to plots with just the cash crop. This is not solely due to the sample size being smaller, but actually due to the undersown species abundance varying across the plots, which in turn, causes the plot outputs to change accordingly. This is a fundamental challenge in modelling secondary crops. While the cash crop can reasonably be assumed to be near-uniform across plots, there is more variation across the plots with undersown species due to the dominant nature of the primary crop. However, the standard deviation of the observed yields was only slightly smaller for sole barley than some intercropped plots in our data. Our speculation is that this variation in abundance is a reason why the posterior uncertainty did not improve significantly during the calibration, with the exception of the yield and LAI calibrations. This internal variation is also a key reason why we have not analysed the impact of the species on the calibration as we could not reliably state if, for example, plant species with deeper roots showed more change in parameter values.

We see potential in this direct intercrop calibration method compared to the more conventional approach of calibrating in single crops and validating in intercrop situations (Wallach et al., 2011; Vezy et al., 2023). The self-calibration captured more of the variation between plot types than the single crop barley calibration, although both remained quite uncertain. In capturing this variation, the direct intercrop calibration may be better suited to answering questions concerning which undersown crop to grow in a particular situation, but only if conditions remain consistent, while the conventional single crop calibration remains more generically applicable. However, the data requirements are larger, and the issue of relative crop abundance poses a significant problem for the reliability of calibrations using this approach.

4.3. Future work

The results of our experiment are largely positive, regarding both the calibration method and how the calibration affects the projection performance, but they also highlight a number of aspects to further explore. Expanding the observations used in both calibration and validation to ensure that similar limiting conditions are present in both datasets would help clarify the utility of this intercrop calibration method. Additionally, considering the nitrogen cycle would be an important aspect for future work, as undersown crops affect fertiliser needs and nitrogen leaching (Valkama et al., 2015; Bedoussac et al., 2015; Corre-Hellou et al., 2009). However, a study by Barton et al. (2015) showed the importance of measuring daily N₂O fluxes in order to remain within 10 % of the best estimate of the annual N₂O budget. Therefore, high frequency measurements would be necessary in order to best capture the effect of cover crops on the nitrogen cycle and ensure these effects are accurately represented by the model.

5. Conclusion

In this work, we were able to calibrate cash crop parameters in STICS for simulated plots containing barley sown with other species to reflect the influence of the undersown species. The projections, though, only improve when the conditions are similar to those used for calibration. Additionally, the twin experiments proved to be a crucial tool in evaluating the reliability of particular calibrations. When analysing the calibration results, we saw enhancement in the yield projections, but NEE estimates turned out to be difficult to improve. Furthermore, simultaneous calibration of multiple different data streams resulted in worse individual output performances compared to just calibrating with the relevant data. This is the first pass at the approach and the results highlight the value of twin experiments for future experiments that extend the range of conditions and data streams used in calibration.

CRedit authorship contribution statement

Quentin Bell: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Investigation. **Stephanie Gerin:** Writing – review & editing, Writing – original draft, Data curation. **Natalie Douglas:** Writing – review & editing, Writing – original draft, Methodology. **Tristan Quaife:** Writing – review & editing, Software, Methodology. **Jari Liski:** Resources, Project administration, Funding acquisition. **Toni Viskari:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Acknowledgements

This work was supported by the TWINWIN project funded by the Maj and Tor Nessling Foundation, the MULTA grant (grant number 352431) funded by the Strategic Research Council (SRC) at the Research Council of Finland, and funded by the Research Council of Finland (grant number 337552). Natalie Douglas and Tristan Quaife were funded under the International Programme of the UKRI National Centre for Earth Observation (NE/X006328/1). The authors would like to thank Liisa Kulmala, Pinja Lindgren, Jussi Heinonsalo, and Sonja Repetti for their involvement and contributions.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Stephanie Gerin and Toni Viskari report financial support was provided by Maj and Tor Nessling Foundation. Natalie Douglas and Tristan Quaife report financial support was provided by National Centre for Earth Observation. Quentin Bell, Stephanie Gerin, Jari Liski, and Toni Viskari

report financial support was provided by Research Council of Finland. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Code for the analysis presented in this article, including the generation of all figures and most tables, is available at <https://github.com/qdbell/TWINWIN-Intercrop> alongside instructions for replicating these results. Data to replicate the results is available from <https://doi.org/10.57707/fmi-b2share.5045c445712b4086840ac199e4b5d2b2>. In order to replicate the results both 4DnVar (https://github.com/tquaife/4DnVar_engine/tree/7bb23a2 for the specific version used), and STICS v9.2 <https://stics.inrae.fr/eng/download> should also be downloaded.

Data for the manuscript "Calibrating primary crop parameters to capture undersown species impacts" by Q. Bell et al. (Original data) (METIS - FMI's Research Data repository) [TWINWIN-Intercrop \(Original data\) \(Github\)](#)

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.eja.2025.127676](https://doi.org/10.1016/j.eja.2025.127676).

References

- Adetunji, A.T., Ncube, B., Mulidzi, R., Lewu, F.B., 2020. Management impact and benefit of cover crops on soil quality: a review. *Soil Tillage Res.* 204, 104717. <https://doi.org/10.1016/j.still.2020.104717>.
- Ahongshangbam, J., Kulmala, L., Soininen, J., Frühauf, Y., Karvinen, E., Salmon, Y., Lintunen, A., Karvonen, A., Järvi, L., 2023. Sap flow and leaf gas exchange response to a drought and heatwave in urban green spaces in a Nordic city. *Biogeosciences* 20, 4455–4475. <https://doi.org/10.5194/bg-20-4455-2023>.
- Aphalo, P.J., 2023. High frequency weather data for viikki, helsinki, finland. *osf.io/e4vau*, 10.17605/OSF.IO/E4VAU.
- Asseng, S., Zhu, Y., Basso, B., Wilson, T., Cammarano, D., 2014. Simulation modeling: applications in cropping systems. In: Van Alfen, N.K. (Ed.), *Encyclopedia of Agriculture and Food Systems*. Academic Press, Oxford, pp. 102–112. <https://doi.org/10.1016/B978-0-444-52512-3.00233-3>.
- Baldrian, P., 2017. Microbial activity and the dynamics of ecosystem processes in forest soils. *Curr. Opin. Microbiol.* 37, 128–134. <https://doi.org/10.1016/j.mib.2017.06.008>.
- Barton, L., Wolf, B., Rowlings, D., Scheer, C., Kiese, R., Grace, P., Stefanova, K., Butterbach-Bahl, K., 2015. Sampling frequency affects estimates of annual nitrous oxide fluxes. *Sci. Rep.* 5, 1–9. <https://doi.org/10.1038/srep15912>.
- Bedoussac, L., Journet, E.P., Hauggaard-Nielsen, H., Naudin, C., Corre-Hellou, G., Jensen, E.S., Prieur, L., Justes, E., 2015. Ecological principles underlying the increase of productivity achieved by cereal-grain legume intercrops in organic farming. *A Rev. Agron. Sustain. Dev.* 35, 911–935. <https://doi.org/10.1007/s13593-014-0277-7>.
- Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.* 249, 11–29. [https://doi.org/10.1016/S0022-1694\(01\)00421-8](https://doi.org/10.1016/S0022-1694(01)00421-8).
- Black, R., Cullen, K., Fay, B., Hale, T., Lang, J., Mahmood, S., Smith, S., 2021. Taking stock: A global assessment of net zero targets. (<https://eciu.net/analysis/report-s/2021/taking-stock-assessment-net-zero-targets>).
- Boote, K.J., Jones, J.W., White, J.W., Asseng, S., Lizaso, J.I., 2013. Putting mechanisms into crop production models. *Plant, Cell Environ.* 36, 1658–1672. <https://doi.org/10.1111/pce.12119>.
- Brisson, N., Bussi re, F., Ozier-Lafontaine, H., Tournebize, R., Sinoquet, H., 2004. Adaptation of the crop model STICS to intercropping. Theoretical basis and parameterisation. *Agronomie* 24, 409–421. <https://doi.org/10.1051/agro:2004031>.
- Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussi re, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudill re, J.P., H nault, C., Maraux, F., Seguin, B., Sinoquet, H., 2003. An overview of the crop model stics. *Eur. J. Agron.* 18, 309–332. [https://doi.org/10.1016/S1161-0301\(02\)00110-7](https://doi.org/10.1016/S1161-0301(02)00110-7).
- Brown, H., Huth, N., Holzworth, D., 2018. Crop model improvement in APSIM: using wheat as a case study. *Eur. J. Agron.* 100, 141–150. <https://doi.org/10.1016/j.eja.2018.02.002>.
- Buis, S., Lecharpentier, P., Vezy, R., Giner, M., 2023. CroptimizR: A Package to Estimate Parameters of Crop Models. (<https://github.com/SticsR Packs/CroptimizR>).

- Callaway, R.M., 2007. Positive Interactions and Interdependence in Plant Communities. Springer, Netherlands, Dordrecht. <https://doi.org/10.1007/978-1-4020-6224-7>.
- Cappelli, S.L., Domeignoz Horta, L.A., Gerin, S., Heinonsalo, J., Lohila, A., Raveala, K., Schmid, B., Shrestha, R., Tiisanen, M.J., Thitz, P., Laine, A.L., 2024. Potential of undersown species identity versus diversity to manage disease in crops. *Funct. Ecol.* 38, 1497–1509. <https://doi.org/10.1111/1365-2435.14592>.
- Chib, S., 2001. Chapter 57 - Markov Chain Monte Carlo methods: computation and inference. In: Heckman, J.J., Leamer, E. (Eds.), *Handbook of Econometrics*. Elsevier, pp. 3569–3649. [https://doi.org/10.1016/S1573-4412\(01\)05010-3](https://doi.org/10.1016/S1573-4412(01)05010-3) volume 5.
- Corre-Hellou, G., Faure, M., Launay, M., Brisson, N., Crozat, Y., 2009. Adaptation of the STICS intercrop model to simulate crop growth and N accumulation in pea-barley intercrops. *Field Crops Res.* 113, 72–81. <https://doi.org/10.1016/j.fcr.2009.04.007>.
- Crystal-Ornelas, R., Thapa, R., Tully, K.L., 2021. Soil organic carbon is affected by organic amendments, conservation tillage, and cover cropping in organic farming systems: a meta-analysis. *Agric. Ecosyst. Environ.* 312, 107356. <https://doi.org/10.1016/j.agee.2021.107356>.
- Dai, Y., Dickinson, R.E., Wang, Y.P., 2004. A two-big-leaf model for canopy temperature, photosynthesis, and stomatal conductance. *J. Clim.* 17, 2281–2299. [https://doi.org/10.1175/1520-0442\(2004\)017<2281:ATMFC2>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2281:ATMFC2>2.0.CO;2).
- Demarty, J., Chevallier, F., Friend, A.D., Viovy, N., Piao, S., Ciais, P., 2007. Assimilation of global MODIS leaf area index retrievals within a terrestrial biosphere model. *Geophys. Res. Lett.* 34. <https://doi.org/10.1029/2007GL030014>.
- Dumont, B., Leemans, V., Mansouri, M., Bodson, B., Destain, J.P., Destain, M.F., 2014. Parameter identification of the STICS crop model, using an accelerated formal MCMC approach. *Environ. Model. Softw.* 52, 121–135. <https://doi.org/10.1016/j.envsoft.2013.10.022>.
- Dzotsi, K.A., Basso, B., Jones, J.W., 2013. Development, uncertainty and sensitivity analysis of the simple SALUS crop model in DSSAT. *Ecol. Model.* 260, 62–76. <https://doi.org/10.1016/j.ecolmodel.2013.03.017>.
- Evensen, G., 2003. The ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn.* 53, 343–367. <https://doi.org/10.1007/s10236-003-0036-9>.
- Fankhauser, S., Smith, S.M., Allen, M., Axelsson, K., Hale, T., Hepburn, C., Kendall, J.M., Khosla, R., Lezaun, J., Mitchell-Larson, E., Obersteiner, M., Rajamani, L., Rickaby, R., Seddon, N., Wetzer, T., 2022. The meaning of net zero and how to get it right. *Nat. Clim. Change* 12, 15–21. <https://doi.org/10.1038/s41558-021-01245-w>.
- Fatchi, S., Leuzinger, S., Körner, C., 2014. Moving beyond photosynthesis: from carbon source to sink-driven vegetation modeling. *N. Phytol.* 201, 1086–1095. <https://doi.org/10.1111/nph.12614>.
- Fer, I., Gardella, A.K., Shiklomanov, A.N., Campbell, E.E., Cowdery, E.M., De Kauwe, M. G., Desai, A., Duveneck, M.J., Fisher, D.B., Haynes, K.D., Hoffman, F.M., Johnston, M.R., Kooper, R., LeBauer, D.S., Mantooh, J., Parton, W.J., Poulter, B., Quaife, T., Raiho, A., Schaefer, K., Serbin, S.P., Simkins, J., Wilcox, K.R., Viskari, T., Dietze, M.C., 2021. Beyond ecosystem modeling: a roadmap to community cyberinfrastructure for ecological data-model integration. *Glob. Change Biol.* 27, 13–26. <https://doi.org/10.1111/gcb.15409>.
- Finnish Meteorological Institute, 2024. Finnish Meteorological Institute open data, download observations service. (<https://en.fmi.fi/taiteenlaitos/download-observations>), cited: 2024-08-29, Downloaded: 2024-01-10.
- Fisher, R.A., Koven, C.D., 2020. Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *J. Adv. Model. Earth Syst.* 12, e2018MS001453. <https://doi.org/10.1029/2018MS001453>.
- Franklin, S.E., Lavigne, M.B., Deuling, M.J., Wulder, M.A., Hunt Jr, E.R., 1997. Estimation of forest Leaf Area Index using remote sensing and GIS data for modelling net primary production. *Int. J. Remote Sens.* 18, 3459–3471. <https://doi.org/10.1080/014311697216973>.
- Friedlingstein, P., O'Sullivan, M., Jones, M.W., Andrew, R.M., Bakker, D.C.E., Hauck, J., Landschützer, P., LeQuéré, C., Luijckx, I.T., Peters, G.P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J.G., Ciais, P., Jackson, R.B., Alin, S.R., Anthoni, P., Barbero, L., Bates, N.R., Becker, M., Bellouin, N., Decharme, B., Bopp, L., Brasika, I.B.M., Cadule, P., Chamberlain, M.A., Chandra, N., Chau, T.T.T., Chevallier, F., Chini, L.P., Cronin, M., Dou, X., Enyo, K., Evans, W., Falk, S., Feely, R. A., Feng, L., Ford, D.J., Gasser, T., Ghattas, J., Gkritzalis, T., Grass, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Heinke, J., Houghton, R.A., Hurtt, G.C., Iida, Y., Ilyina, T., Jacobson, A.R., Jain, A., Jarníková, T., Jersild, A., Jiang, F., Jin, Z., Joos, F., Kato, E., Keeling, R.F., Kennedy, D., KleinGoldewijk, K., Knauer, J., Korsbakken, J.I., Körtzinger, A., Lan, X., Lefèvre, N., Li, H., Liu, J., Liu, Z., Ma, L., Marland, G., Mayot, N., McGuire, P.C., McKinley, G.A., Meyer, G., Morgan, E.J., Munro, D.R., Nakaoka, S.I., Niwa, Y., O'Brien, K.M., Olsen, A., Omar, A.M., Ono, T., Paulsen, M., Pierrot, D., Pöckel, K., Poulter, B., Powis, C.M., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T.M., Schwinger, J., Séférian, R., Smallman, T.L., Smith, S.M., Sospedra-Alfonso, R., Sun, Q., Sutton, A.J., Sweeney, C., Takao, S., Tans, P.P., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G. R., van Ooijen, E., Wanninkhof, R., Watanabe, M., Wimart-Rousseau, C., Yang, D., Yang, X., Yuan, W., Yue, X., Zaehle, S., Zeng, J., Zheng, B., 2023. Global carbon budget 2023. *Earth Syst. Sci. Data* 15, 5301–5369. <https://doi.org/10.5194/essd-15-5301-2023>.
- Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W., Sitch, S., 2004. Terrestrial vegetation and water balance—hydrological evaluation of a dynamic global vegetation model. *J. Hydrol.* 286, 249–270. <https://doi.org/10.1016/j.jhydrol.2003.09.029>.
- Gielen, B., Acosta, M., Altimir, N., Buchmann, N., Cescatti, A., Ceschia, E., Fleck, S., Hörtnagl, L., Klumpp, K., Kolari, P., Lohila, A., Loustau, D., Marañón-Jimenez, S., Manise, T., Matteucci, G., Merbold, L., Metzger, C., Moureaux, C., Montagnani, L., Nilsson, M.B., Osborne, B., Papale, D., Pavelka, M., Saunders, M., Simioni, G., Soudani, K., Sonnentag, O., Tallec, T., Tuittila, E.S., Peichl, M., Pokorný, R.,
- Vincke, C., Wohlfahrt, G., 2018. Ancillary vegetation measurements at ICOS ecosystem stations. *Int. Agrophys.* 32, 645–664. <https://doi.org/10.1515/intag-2017-0048>.
- Gouttevin, I., Krinner, G., Ciais, P., Polcher, J., Legout, C., 2012. Multi-scale validation of a new soil freezing scheme for a land-surface model with physically-based hydrology. *Cryosphere* 6, 407–430. <https://doi.org/10.5194/tc-6-407-2012>.
- Guillaume, S., Bergez, J.E., Wallach, D., Justes, E., 2011. Methodological comparison of calibration procedures for durum wheat parameters in the STICS model. *Eur. J. Agron.* 35, 115–126. <https://doi.org/10.1016/j.eja.2011.05.003>.
- Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive metropolis algorithm. *Bernoulli* 7, 223–242.
- Hartig, F., Dyke, J., Hickler, T., Higgins, S.I., O'Hara, R.B., Scheiter, S., Huth, A., 2012. Connecting dynamic vegetation models to data – an inverse perspective. *J. Biogeogr.* 39, 2240–2252. <https://doi.org/10.1111/j.1365-2699.2012.02745.x>.
- Jégo, G., Pattey, E., Bourgeois, G., Morrison, M.J., Drury, C.F., Tremblay, N., Tremblay, G., 2010. Calibration and performance evaluation of soybean and spring wheat cultivars using the STICS crop model in Eastern Canada. *Field Crops Res.* 117, 183–196. <https://doi.org/10.1016/j.fcr.2010.03.008>.
- Jokinen, P., Pirinen, P., Kaukoranta, J.P., Kangas, A., Alenius, P., Eriksson, P., Johansson, M., Wilkman, S., 2021. Til. Suom. Ilma ja Meres ä 1991–2020. Rap. - Rapp. - Rep. 2021, 8. <https://doi.org/10.35614/isbn.9789523361485>.
- Kherif, O., Seghouani, M., Justes, E., Plaza-Bonilla, D., Bouhache, A., Zemmouri, B., Dokukin, P., Latati, M., 2022. The first calibration and evaluation of the STICS soil-crop model on chickpea-based intercropping system under Mediterranean conditions. *Eur. J. Agron.* 133, 126449. <https://doi.org/10.1016/j.eja.2021.126449>.
- Korhonen, P., Palosuo, T., Persson, T., Höglind, M., Jégo, G., Van Oijen, M., Gustavsson, A.M., Bélanger, G., Virkajärvi, P., 2018. Modelling grass yields in northern climates - a comparison of three growth models for timothy. *Field Crops Res.* 224, 37–47. <https://doi.org/10.1016/j.fcr.2018.04.014>.
- Lal, R., 2015. Sequestering carbon and increasing productivity by conservation agriculture. *J. Soil Water Conserv.* 70, 55A–62A. <https://doi.org/10.2489/jswc.70.3.55A>.
- Lecharpentier, P., Vezy, R., Buis, S., Giner, M., 2023. SticsOnR: Manage STICS simulations running the executable or javaStics. (<https://github.com/SticsRPacks/SticsOnR>), 10.5281/zenodo.4443130.
- Ma, L., Hurtt, G., Ott, L., Sahajpal, R., Fisk, J., Lamb, R., Tang, H., Flanagan, S., Chini, L., Chatterjee, A., Sullivan, J., 2022. Global evaluation of the ecosystem demography model (ED v3.0). *Geosci. Model Dev.* 15, 1971–1994. <https://doi.org/10.5194/gmd-15-1971-2022>.
- Martin Belda, D., Anthoni, P., Wårlind, D., Olin, S., Schurgers, G., Tang, J., Smith, B., Arneth, A., 2022. LPJ-GUESS/LSMv1.0: a next-generation land surface model with high ecological realism. *Geosci. Model Dev.* 15, 6709–6745. <https://doi.org/10.5194/gmd-15-6709-2022>.
- Mathers, C., Black, C.K., Segal, B.D., Gurling, R.B., Zhang, Y., Easter, M.J., Williams, S., Motew, M., Campbell, E.E., Brummitt, C.D., Paustian, K., Kumar, A.A., 2023. Validating DayCent-CR for cropland soil carbon offset reporting at a national scale. *Geoderma* 438, 116647. <https://doi.org/10.1016/j.geoderma.2023.116647>.
- Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovov, V., Stockmann, U., Sulaeman, Y., Tsui, C.C., Vågen, T.G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. *Geoderma* 292, 59–86. <https://doi.org/10.1016/j.geoderma.2017.01.002>.
- van Oijen, M., 2017. Bayesian methods for quantifying and reducing uncertainty and error in forest models. *Curr. For. Rep.* 3, 269–280. <https://doi.org/10.1007/s40725-017-0069-9>.
- Pimentel, D., Cerasale, D., Stanley, R.C., Perlman, R., Newman, E.M., Brent, L.C., Mullan, A., Chang, D.T.I., 2012. Annual vs. perennial grain production. *Agric. Ecosyst. Environ.* 161, 1–9. <https://doi.org/10.1016/j.agee.2012.05.025>.
- Pinnington, E., Amezcua, J., Cooper, E., Dadson, S., Ellis, R., Peng, J., Robinson, E., Morrison, R., Osborne, S., Quaife, T., 2021. Improving soil moisture prediction of a high-resolution land surface model by parameterising pedotransfer functions through assimilation of SMAP satellite data. *Hydrol. Earth Syst. Sci.* 25, 1617–1641. <https://doi.org/10.5194/hess-25-1617-2021>.
- Pinnington, E., Quaife, T., Lawless, A., Williams, K., Arkebauer, T., Scooby, D., 2020. Land variational ensemble data assimilation framework: LAVENDAR v1.0.0. *Geosci. Model Dev.* 13, 55–69. <https://doi.org/10.5194/gmd-13-55-2020>.
- Ravelojaona, N., Jégo, G., Ziadi, N., Mollier, A., Lafond, J., Karam, A., Morel, C., 2023. STICS soil-crop model performance for predicting biomass and nitrogen status of spring barley cropped for 31 years in a gleyic soil from Northeastern Quebec (Canada). *Agronomy* 13, 2540. <https://doi.org/10.3390/agronomy13102540>.
- Rawlins, F., Ballard, S.P., Bovis, K.J., Clayton, A.M., Li, D., Inverarity, G.W., Lorenc, A.C., Payne, T.J., 2007. The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* 133, 347–362. <https://doi.org/10.1002/qj.32>.
- Reichle, R.H., Zhang, S.Q., Kolassa, J., Liu, Q., Todling, R., 2023. A weakly coupled land surface analysis with SMAP radiance assimilation improves GEOS medium-range forecasts of near-surface air temperature and humidity. *Q. J. R. Meteorol. Soc.* 149, 1867–1889. <https://doi.org/10.1002/qj.4486>.
- Ruget, F., Brisson, N., Delécolle, R., Faivre, R., 2002. Sensitivity analysis of a crop simulation model, STICS, in order to choose the main parameters to be estimated. *Agronomie* 22, 133–158. <https://doi.org/10.1051/agro:2002009>.
- Ruimy, A., Jarvis, P.G., Baldocchi, D.D., Saugier, B., 1995. CO₂ fluxes over plant canopies and solar radiation: a review. In: Begon, M., Fitter, A.H. (Eds.), *Advances in*

- Ecological Research. Academic Press, pp. 1–68. [https://doi.org/10.1016/S0065-2504\(08\)60063-X](https://doi.org/10.1016/S0065-2504(08)60063-X) volume 26.
- Rumpel, C., Amiraslani, F., Chenu, C., Garcia Cardenas, M., Kaonga, M., Koutika, L.S., Ladha, J., Madari, B., Shirato, Y., Smith, P., Soudi, B., Soussana, J.F., Whitehead, D., Wollenberg, E., 2020. The 4p1000 initiative: opportunities, limitations and challenges for implementing soil organic carbon sequestration as a sustainable development strategy. *Ambio* 49, 350–360. <https://doi.org/10.1007/s13280-019-01165-2>.
- Scharlemann, J.P., Tanner, E.V., Hiederer, R., Kapos, V., 2014. Global soil carbon: understanding and managing the largest terrestrial carbon pool. *Carbon Manag.* 5, 81–91. <https://doi.org/10.4155/cmt.13.77>.
- Schmidt, M.W.I., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D.A.C., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S.E., 2011. Persistence of soil organic matter as an ecosystem property. *Nature* 478, 49–56. <https://doi.org/10.1038/nature10386>.
- Seidel, S.J., Palosuo, T., Thorburn, P., Wallach, D., 2018. Towards improved calibration of crop models – where are we now and where should we go? *Eur. J. Agron.* 94, 25–35. <https://doi.org/10.1016/j.eja.2018.01.006>.
- Shiklomanov, A.N., Dietze, M.C., Fer, I., Viskari, T., Serbin, S.P., 2021. Cutting out the middleman: calibrating and validating a dynamic vegetation model (ED2-PROSPECT5) using remotely sensed surface reflectance. *Geosci. Model Dev.* 14, 2603–2633. <https://doi.org/10.5194/gmd-14-2603-2021>.
- Tarantola, A., 2005. Inverse problem theory and methods for model parameter estimation. *Society for Industrial and Applied Mathematics*. 10.1137/1.9780898717921.
- Trémeau, J., Olascoaga, B., Backman, L., Karvinen, E., Vekuri, H., Kulmala, L., 2024. Lawns and meadows in urban green space - a comparison from perspectives of greenhouse gases, drought resilience and plant functional types. *Biogeosciences* 21, 949–972. <https://doi.org/10.5194/bg-21-949-2024>.
- Valkama, E., Lemola, R., Känkänen, H., Turtola, E., 2015. Meta-analysis of the effects of undersown catch crops on nitrogen leaching loss and grain yields in the Nordic countries. *Agric. Ecosyst. Environ.* 203, 93–101. <https://doi.org/10.1016/j.agee.2015.01.023>.
- Vezy, R., Munz, S., Gaudio, N., Launay, M., Lecharpentier, P., Ripoche, D., Justes, E., 2023. Modeling soil-plant functioning of intercrops using comprehensive and generic formalisms implemented in the STICS model. *Agron. Sustain. Dev.* 43, 61. <https://doi.org/10.1007/s13593-023-00917-5>.
- Vrugt, J.A., Ter Braak, C.J.F., 2011. DREAM_(D): An adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrol. Earth Syst. Sci.* 15, 3701–3713. <https://doi.org/10.5194/hess-15-3701-2011>.
- Wallach, D., Buis, S., Lecharpentier, P., Bourges, J., Clastre, P., Launay, M., Bergez, J.E., Guerif, M., Soudais, J., Justes, E., 2011. A package of parameter estimation methods and implementation for the STICS crop-soil model. *Environ. Model. Softw.* 26, 386–394. <https://doi.org/10.1016/j.envsoft.2010.09.004>.
- Wallach, D., Buis, S., Seserman, D.M., Palosuo, T., Thorburn, P.J., Mielenz, H., Justes, E., Kersebaum, K.C., Dumont, B., Launay, M., Seidel, S.J., 2024. A calibration protocol for soil-crop models. *Environ. Model. Softw.* 180, 106147. <https://doi.org/10.1016/j.envsoft.2024.106147>.
- Yang, T., Siddique, K.H.M., Liu, K., 2020. Cropping systems in agriculture and their impact on soil health—a review. *Glob. Ecol. Conserv.* 23, e01118. <https://doi.org/10.1016/j.gecco.2020.e01118>.