agronomy

Article

ARMOSA Model Parametrization for Winter Durum Wheat Cultivation under Diverse Cropping Management Practices in a Mediterranean Environment

Pasquale Garofalo [1,*], Marco Parlavecchia [1], Luisa Giglio [1], Ivana Campobasso [1], Alessandro Vittorio Vonella [1], Marco Botta [2], Tommaso Tadiello [2], Vincenzo Tucci [3], Francesco Fornaro [1], Rita Leogrande [1], Carolina Vitti [1], Alessia Perego [2], Marco Acutis [2] and Domenico Ventrella [1]

1 Council for Agricoltural Research and Economics—Agriculture and Environment, 70125 Bari, Italy; marcoparlavecchia87@libero.it (M.P.); luisa.giglio@crea.gov.it (L.G.); ivana.campobasso@libero.it (I.C.); vittorio.vonella@crea.gov.it (A.V.V.); francesco.fornaro@crea.gov.it (F.F.); rita.leogrande@crea.gov.it (R.L.); carolina.vitti@crea.gov.it (C.V.); domenico.ventrella@crea.gov.it (D.V.)

2 Department of Agricultural and Environmental Sciences—Production, Landscape, Agroenergy, University of Milan, 20133 Milan, Italy; marco.botta@unimi.it (M.B.); tommaso.tadiello@unimi.it (T.T.); alessia.perego@unimi.it (A.P.); marco.acutis@unimi.it (M.A.)

3 Department of Soil, Plant and Food Sciences, University of Bari Aldo Moro, 70121 Bari, Italy; tuccivi@hotmail.it

* Correspondence: pasquale.garofalo@crea.gov.it; Tel.: +39-080-547-5011

Abstract: In anticipation of climate changes, strategic soil management, encompassing reduced tillage and optimized crop residue utilization, emerges as a pivotal strategy for climate impact mitigation. Evaluating the transition from conventional to conservative cropping systems, especially the equilibrium shift in the medium to long term, is essential. ARMOSA, a robust crop simulation model, adeptly responds to varied soil management practices such as no tillage, minimum tillage, and specific straw management options such as chopping and incorporating crop residue into the soil (with or without prior nitrogen and water addition before ploughing). It effectively captures dynamic fluctuations in total organic carbon over an extended period. While challenges persist in precisely predicting grain yield due to climatic intricacies, ARMOSA stands out as a valuable and versatile tool. The model excels in comprehending and simulating wheat cultivar responses in dynamic agricultural ecosystems, shedding light on phenological patterns, biomass accumulation, and soil organic carbon dynamics. This research significantly advances our understanding of the intricate complexities associated with past wheat cultivation in diverse environmental conditions. ARMOSA's ability to inform decisions on conservation practices positions it as a valuable asset for researchers, agronomists, and policymakers navigating the challenges of sustainable agriculture amidst climate change. Its real-world significance lies in its potential to guide informed decisions, contributing to global efforts in sustainable agriculture and climate resilience.

Keywords: long-term experiment; modelling; agronomy; calibration; soil organic carbon; sustainability

check for updates

Citation: Garofalo, P.; Parlavecchia, M.; Giglio, L.; Campobasso, I.; Vonella, A.V.; Botta, M.; Tadiello, T.; Tucci, V.; Fornaro, F.; Leogrande, R.; et al. ARMOSA Model Parametrization for Winter Durum Wheat Cultivation under Diverse Cropping Management Practices in a Mediterranean Environment. *Agronomy* **2024**, *14*, 164. https://doi.org/10.3390/agronomy14010164

Academic Editors: Mavromatis Theodoros, Thomas Alexandridis and Vassilis Aschonitis

Received: 4 December 2023
Revised: 8 January 2024
Accepted: 9 January 2024
Published: 11 January 2024

1. Introduction

Cereals serve as the primary global food source, with the European Union ranking as the largest wheat producer [1]. Italy, following Canada, currently has the second-highest global production of durum wheat, annually producing 4.2 million tons, primarily concentrated in the southern regions and islands, which account for 65.6% [2].

Global warming (GW) significantly impacts climate conditions, leading to temperature rises, reduced rainfall, increased severity and frequency of extreme events, and elevated atmospheric carbon dioxide levels [3–5]. Consequently, strategies are imperative to adapt to and mitigate the repercussions on crop yield and product quality. Adaptation strategies seek to minimize adverse effects on agricultural production, while mitigation strategies aim to reduce greenhouse gas emissions and sustain or augment the organic carbon content in soil. Integrated analyses are crucial for adjusting cropping systems to evolving climate conditions, especially in regions such as the Mediterranean basin with consistent agronomic and pedoclimatic characteristics.

Recent decades have witnessed detrimental impacts of GW on wheat yields in lower-latitude regions and increased yields in higher-latitude zones [6]. Global projections affirm this trend [7]. A limited body of literature has explored the specific impacts of GW on Italian production [8,9], emphasizing the necessity for accurate climate data, especially in regions with high pedoclimatic and topographical variability, such as Italy [10].

In Northern Apulia's Foggia province, durum wheat is commonly rotated with tomato or irrigated high-income crops over two or three years. Traditional agronomic practices involve mouldboard ploughing, additional tillage, and straw and stubble management through burial or burning [11]. Evaluating soil organic matter dynamics and crop yield in response to tillage through field experiments is expensive and time-consuming. Calibration and validation of process-based models offer a viable solution, aiding in the assessment of diverse soil and crop management practices [12–15].

The literature has explored the effects of soil and crop management practices of conservation agriculture (CA) on crop growth, yield, and nutrient dynamics [16,17]. However, most simulation models struggle to depict the long-term effects of CA practices such as no tillage (NT), minimum tillage (MT), and traditional methods [18–20].

The integration of process-based crop modelling with climate data is vital for understanding the effects of GW on agricultural production. ARMOSA (Analysis of cRopping systems for Management Optimization and Sustainable Agriculture) has emerged as a suitable model for simulating various soil-management practices in diverse environmental conditions [15,21,22]. Its adaptability and reliability in simulating durum wheat growth in the Mediterranean environment make it well suited for this study.

Operating on a daily time-step basis, ARMOSA provides fine-grained analyses, accurately capturing daily fluctuations in pedoclimatic conditions. Its holistic approach, encompassing water balance, evapotranspiration, and nitrogen and carbon cycling, offers a comprehensive understanding of the crop–soil system. The model's modularity enhances adaptability, facilitating adjustments for different experimental setups and changing scenarios.

This research aimed to evaluate ARMOSA's current performance, assessing its ability to replicate essential aspects of durum wheat growth, including phenology, biomass accumulation, grain yield, and soil organic carbon dynamics. It introduced a performance scoring system, conducted comparisons with observed data, and assessed robustness through a validation process. Utilizing experimental data from a long-term durum wheat experiment (LTE) in Foggia, Southern Italy, spanning from 1977 to the present, the study explored the impacts of different straw and soil management practices, including nitrogen and water addition, no tillage, and minimum tillage. It specifically investigated cultivar-specific variations and soil/straw option impacts on wheat growth and soil organic carbon dynamics.

2. Materials and Methods

2.1. Experimental Field

All the field experiments were carried out at Podere 124 (P124) Experimental Station, located in Foggia, Apulia region, Southern Italy (latitude, 41°88′7″ N; longitude, 15°83′05″ E; altitude, 90 m a.s.l.), in two experimental parcels: P124_P30, used to calibrate the model, and P124_P32, used to validate it.

The soil, a vertisol of alluvial origin [23], is classified as silty clay with the following physicochemical properties: 48.5% clay, 38.7% silt, 12.8% sand, bulk density: 1.11 t m^{-3}, organic matter: 2.1%; total N: 0.122%; $NaHCO_3$-extractable P: 41 ppm; NH_4OAc extractable K_2O: 1598 ppm; pH: 8.3; water content at field capacity: 0.396 m^3 m^{-3}; water content at permanent wilting point: 0.195 m^3 m^{-3}; available water: 202 mm m^{-1}.

The climate is classified as "accentuated thermo-Mediterranean" [24], characterized by temperatures below 0 °C in winter and above 40 °C in summer, with an annual average of 550 mm of rainfall, mostly concentrated in winter months [11]. The daily meteorological data of temperature, humidity, rainfall, wind parameters, and solar radiation were recorded at the meteorological station located at P124.

2.2. LTE Datasets

The LTE dataset used to parametrize ARMOSA and to check the robustness and reliability of the model consisted of data from winter durum wheat in a continuous cropping system since 1977. The winter durum wheat underwent three different straw management practices, namely: (i) chopping and incorporation of the crop residue into the soil by ploughing (T2); (ii) chopping, addition of 150 kg of mineral nitrogen per hectare on the straw, and incorporation of the crop residue into the soil by ploughing (T5); and (iii) chopping, addition of 150 kg of mineral nitrogen per hectare, and application of 500 m^3 ha^{-1} of irrigation water on the straw, followed by incorporation of the crop residue into the soil (T8).

The experimental design utilized a randomized block design with five replications, each consisting of an 8 m × 10 m cropped area with a spacing of 15 cm (between two rows) × 5 cm (along the rows). These replications were situated within a single experimental plot (total area of 3500 m^2), referred to as P_30.

For all experimental treatments, sowing, which took place in the first half of November, was preceded by fertilization with superphosphate (100 kg P_2O_5 ha^{-1}), ploughing (with soil incorporation of the chopped straw), harrowing with a disc harrow, and tilling with a rotary tiller. A dose of 100 kg N ha^{-1} was supplied to the crop as top dressing in the first half of March, and the harvest was performed in the middle two weeks of June.

Before harvesting, plant samples were collected over an area of 2 m^2 to estimate the total aboveground dry biomass (TDM) by placing the sample in a ventilated oven at 78 °C until a constant weight was reached.

The wheat harvest took place using a plot combine, which determined, thanks to a portable module, the grain yield for each replication and the corresponding moisture content (from which the dry weight of grain was calculated).

In addition, from 1983 to 2009, the soil organic carbon content (TOC; kg ha^{-1}; 0–40 cm depth) was determined discontinuously on three soil samples of about 500 g each for each replication.

In P_30, the following cultivars (cvs) succeeded each other over the harvesting years: Valgerardo 1978–1982, Appulo 1983–1987, Latino 1988–1992, Appio 1993–1996, Simeto 1997–2000 and 2007–2013, Ofanto 2001–2006, Claudio 2014–2018, and Saragolla 2019–2021.

The consistency of ARMOSA was probed on a separate dataset with the same parametrization process. Here, figures were gathered from another LTE consisting of wheat in a continuous cropping system since 2003, cultivated under two CA schemes: NT and MT.

The experimental design was planned in a randomized block design containing three replications for each treatment with an area of 500 m^2 (20 m × 25 m) arranged in one experimental plot (P_32) with a total surface of 4450 m^2.

NT consisted of sowing (in the first half of November) with a no-till seeder and without further disturbance of soil.

Under MT, a single field operation before sowing was performed using a combined farm device with a subsoiler and a rotary cultivator disturbing the first layer of soil at a depth of 0–0.10 m.

For all soil management, straw and stubble were chopped after the harvest and spread back.

Mineral nitrogen fertilization was split into two doses: basal dressing before sowing in the form of di-ammonium phosphate (36 and 92 kg ha^{-1} of N and P_2O_5, respectively) and ammonium nitrate as top dressing (68 kg ha^{-1}) in the first half of March. Weeds were kept under control using chemicals applied before sowing and after emergence.

The cvs that followed one another over the years in P_32 were: (i) Simeto, from 2003 to 2010; (ii) Claudio, from 2011 to 2018; and (iii) Saragolla from 2018 to 2020.

As for P_30, the plot harvester collected data about grain yield and moisture (the dry weight of grain was calculated accordingly) for each replication, over a period from 10 June to 25 June of the examined growing seasons.

Even for P_32, TOC (0–30 cm depth) from 2002 to 2020 (not continuously) was determined using three soil samples of about 500 g each for each replication.

When the emergence, flowering and physiological maturity stages were verified in the experimental plots of P_30 and P_32, the corresponding calendar days were recorded (specific for each growing season, but common to all cvs). Growing degree days (GDD; °C) were computed as the daily mean temperature minus the base temperature (0 °C) for the specific phenological stage accordingly.

2.3. The ARMOSA Model

ARMOSA is a cropping-system model that simulates crop- and soil-related variables at a daily time-step as affected by pedoclimatic conditions and agronomic management. The software is written in Java and structured with a high level of modularity. The model simulates water balance, evapotranspiration, and N and C cycling in the soil layers as well as crop development and growth. The soil properties (i.e., texture, bulk density, and initial soil organic carbon) are set for each layer of the profile. The water dynamics are simulated with the approach consisting of cascading with travel time [25].

The reference evapotranspiration is estimated using the Penman–Monteith, Priestley–Taylor, or Hargreaves equation. Crop evapotranspiration is estimated using the FAO 56 approach [26]. The actual evapotranspiration is determined by the water stress factor [27], influencing both dry-matter production and partitioning.

The simulation of crop growth and development with ARMOSA adheres to the WOFOST approach [28], with two notable differences: (i) the canopy is divided into 5 layers with varying light interception, and (ii) development is described using the BBCH scale.

Carbon- and nitrogen-related processes are simulated much as they are in the SOILN model [29], with some enhancements. Each input of organic matter is independently simulated based on specific decomposition rates, C and N concentrations, and depth of incorporation in the soil.

The required input data are as follows: daily weather data, soil properties (texture, bulk density, soil organic carbon (TOC), with the option to enter the measured water retention parameters), cropping system information (i.e., crop type and rotation, sowing and harvesting dates), data on fertilizers (i.e., mineral or organic, amount, timing, application depth, carbon-to-nitrogen ratio, ammonia and nitrate content), irrigation (i.e., water amount, timing, automatic irrigation as a function of water depletion threshold), tillage, and crop residue management.

The effect of tillage is simulated as a function of tillage type (depth, degree of soil-layer mixing, and perturbation) as reported in the WEPP model [30]. As reported in [22], the mixing of two or more adjacent soil layers results in mixing of pools (either inorganic or organic) and mixing of state variables (e.g., soil water content).

The tillage operation leads to an increase in the mineralization rate of the organic carbon pools as it enhances the microporosity, in accordance with [31]. The soil hydrological parameters of the water retention curve are computed daily as a function of the daily values of bulk density and soil organic carbon [32].

The decomposition of the crop residues is simulated according to the specific decomposition rate and amount of biomass that remains in the soil at the harvesting date (percentage of the simulated biomass of the crop organs, leaves, stem, and roots).

2.4. The ARMOSA Model

To adapt the predictive algorithms of durum wheat growth implemented in ARMOSA to the data harvested in the LTE, the adjustment of the crop coefficients was assessed according to a trial-and-error procedure to reflect reasonable simulations or to bring the model output closer to the observed data. The calibration of ARMOSA was conducted initially for nitrogen and carbon cycling and subsequently for biomass accumulation and grain yield, using the genetic simplex method according to [33].

The selection of parameters to calibrate was performed through the screening method of Morris as modified by [34].

Based on this sensitivity analysis, the mineralisation rates of soil organic matter fractions (litter and stable pools), the parameter PCO_2 (the potential CO_2 assimilation parameter), and GDD from emergence to flowering were calibrated separately for the cvs. The maximization of the Nash–Sutcliffe modelling efficiency NSE [35] was chosen as the objective function. Initially, the NSE for simulated and observed soil organic carbon (TOC) data was maximized, followed by maximizing the NSE for observed and simulated yield and biomass data, separately for each of the cvs.

The test benchmark for calibrating ARMOSA was T2, on which the model was primarily modelled. Subsequently, fine tuning of crop parameters was further implemented to closely align the model outputs with the collected data on biomass at harvest, yield, and TOC dynamics in the LTE, including T5 and T8 as well as T2.

After calibrating ARMOSA, its reliability in replicating the biomass and yield at harvest of cvs and TOC dynamic was assessed by means of appropriate evaluative indices, as follows.

$$RMSD = \sqrt{\frac{\sum_{=i}^{n}(X_{obs,i} - X_{model,i})^2}{n}} \tag{1}$$

where:

RMSD is the root mean square error, or the measure of the difference between values predicted by a model and the values actually observed from the environment that is being modelled [36];

$X_{obs,i}$ is the observed value;

$X_{model,i}$ is the forecast value.

$$NRMSE = \frac{RMSE}{\overline{X}_{obs}} * 100 \tag{2}$$

where:

NRMSE is the relative root mean square error, and it can be interpreted as a fraction of the overall range that is typically resolved by the model [37];

$\overline{X}_{obs}$ is the average of the observation values.

$$EF = 1 - \frac{\sum_{=i}^{n}(X_{obs,i} - X_{model,i})^2}{\sum_{=i}^{n}\left(X_{obs,i} - \overline{X}_{obs}\right)} \tag{3}$$

where:

EF is the Nash-Sutcliffe efficiency [38], a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance.

$$d = 1 - \frac{\sum_{=i}^{n}(X_{obs,i} - X_{model,i})^2}{\sum_{=i}^{n}\left(\left|X_{model,i} - \overline{X}_{obs}\right| + \left|X_{obs,i}\overline{X}_{obs}\right|\right)^2} \tag{4}$$

where:

d is the index of agreement [39].

The index of agreement can detect additive and proportional differences in the observed and simulated means and variances.

$$CRM = 1 - \frac{\sum_{=i}^{n} X_{model,i}}{\sum_{=i}^{n} X_{obs,i}} \tag{5}$$

where:

CRM is the coefficient of residual mass [40], which can assume positive values indicating an underestimation of the model outcome or negative values if there is an overestimation of the model output, while values close to zero indicate the absence of trends.

For each evaluation index, a score ranging between −1 (worst) and 1 (best) was assigned, with 0.5 for the middle value.

$$GSD = \begin{cases} 1\ if\ 25 > GSD > 0; \\ 0.5\ if\ 25 < GSD < 40; \\ -1\ if\ GSD > 40. \end{cases} \tag{6}$$

$$EF = \begin{cases} 1\ if\ 1.0 > EF > 0.4; \\ 0.5\ if\ 0 < EF < 0.4; \\ -1\ if\ EF < 0. \end{cases} \tag{7}$$

$$d = \begin{cases} 1\ if\ 1.0 > EF > 0.7; \\ 0.5\ if\ 0.4 < EF < 0.7; \\ -1\ if\ d < 0.4. \end{cases} \tag{8}$$

$$CRM = \begin{cases} 1\ if\ 0.01 > CRM > -0.01; \\ 0.5\ if - 0.1 < CRM < 0.1; \\ -1\ if\ 0.1 < CRM < -0.1. \end{cases} \tag{9}$$

The comparison using these indices focused on specific phenological stages, including the dates of emergence, flowering, and physiological maturity; dry biomass at harvest; grain yield; and TOC.

To rank the aforementioned valuation indices, less stringent criteria were applied than those reported in other modelling exercises [41,42]. Indeed, the authors of those works performed a comparison between the observed and simulated datasets for a specific growing season and individual cvs, which is less challenging than calibration across multiple growing years and/or cvs.

The detailed analysis of the four evaluative indices presented a challenge in expressing a quick and easily interpretable assessment of ARMOSA's performance. As a result, a conclusive evaluation based on the aggregation of scores related to individual indicators (−1, 0.5, and 1) was implemented.

This final score reflecting the reliability of ARMOSA in replicating wheat growth and productivity assumed the following criteria: Very good = total score from 3.5 to 4; Good = total score from 2.5 to 3; Fair = total score from 1.5 to 2; Bad = total score from 0 to 1.

The model's robustness, tested during the validation step, was assessed by investigating parameters of the 1:1 regression model (i.e., R^2, angular coefficient (β) and significance of the regression model) applied to the yield (averaged for each cv and soil tillage option) and TOC of P_32.

3. Results and Discussion

3.1. Calibration

The growth and productivity of wheat showed high variability both among cvs and across the growing seasons sown within the same cvs (Figure 1a,b).

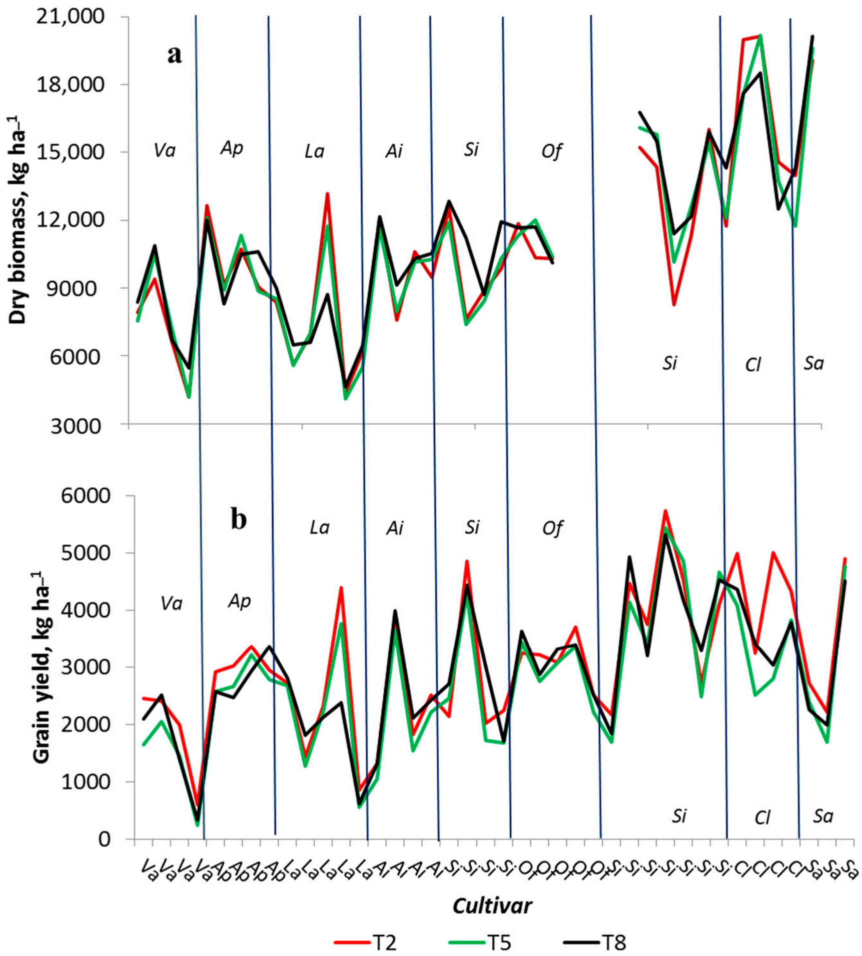

Figure 1. Trend of the total dry biomass (**a**) and grain yield (**b**) at the harvest of durum wheat following one another in the growing years for P_30. *Va* stands for Valgerardo, *Ap* for Appulo, *La* for Latino, *Ai* for Appio, *Si* for Simeto, *Cl* from Claudio, and *Sa* for Saragolla.

Valgerardo and Latino exhibit significant reductions in growth in certain years, adversely affecting productivity. For Valgerardo, dry biomass accumulation halted in 1982, with values ranging between 4157 kg ha^{-1} (T5) and 5447 kg ha^{-1} (T8) in 1982. The grain yield behaved accordingly, with values well below 1000 kg ha^{-1} for all straw treatments.

The following year, a storm occurring just before harvest led to lodging of the plants and resulted in grain loss. Therefore, data from this year were excluded from the reported modelling exercise.

For Latino, dry biomass and productivity at harvest in 1992 remained below 5000 kg ha^{-1} and 1000 kg ha^{-1}, respectively.

Ofanto and Appulo achieved a fair stability of growth and productivity over the growing years, with comparable values in terms of TDM (slightly higher than 10,000 kg ha^{-1} for both) and grain yield (around 3000 kg ha^{-1}).

Simeto and Claudio demonstrated the highest yield potential among the cvs, as evidenced by their high productivity in certain years (peaking over 5000 kg ha^{-1}) compared to the other cvs.

However, even for these two cvs, some growing seasons proved to be unfavourable for the growth and accumulation of biomass, with limited grain yield falling below 2000 kg ha^{-1} for Simeto.

Ultimately, Saragolla was the cv that provided some of the highest (4508 kg ha^{-1} in 2021; T2) and lowest yield values (1692 kg ha^{-1} in 2020; T5) even if, for the worst performances, the corresponding TDM was not especially unfavourable (from 11,723 kg ha^{-1} to 13,974 kg ha^{-1}).

As with Valgerardo, a storm shortly before harvesting heavily compromised grain harvesting in 2001 (cv Ofanto) and 2018 (cv Claudio for T2 and T5 treatments). Consequently, the wheat data from these growing seasons were not considered for model parametrization.

The calibrated values achieved by trial and error for the coefficients of parameters underlying crop growth involved the following: (i) assimilation of CO_2; (ii) conversion into biomass; (iii) separation in the various organs of the plant; (iv) development of the canopy and intercepted radiation; (v) root length; (vi) senescence (Tables 1 and 2).

Table 1. Calibrated values of crop parameters by cultivar. Only modified values are shown in the table.

Parameter	Default Value	Cultivars							
		Appio	Appulo	Claudio	Latino	Ofanto	Saragolla	Simeto	Valgerardo
SPar	12	-	14	-	-	14	19	-	-
EAIfactor	0.5	-	-	-	-	0.4	-	-	-
$LAITH_{min}$	4	-	-	-	-	3	-	-	-
MaintenanceLeaves	0.05	-	-	0.01	-	-	0.02	0.01	-
MaintenanceRoots	0.015	0.05	-	0.01	-	-	0.01	0.01	0.03
MaintenanceStem	0.015	0.05	0.005	0.01	-	-	0.01	0.01	0.09
MaintenanceStorage	0.01	0.05	0.07	0.003	-	-	0.03	0.003	0.01
$PARAgeD_{LAI}$	0.3	0.08	0.7	-	0.2	0.2	0.2	0.43	-
$MaxCO_2Net$	1200	-	1500	-	-	-	-	-	1000
PCO_2	0.0052	-	0.015	0.003	0.001	0.004	0.0099	0.0025	0.0009
MaxRootDepth	800	-	-	900	1000	1000	-	-	-
SLA	0.017	-	0.005	-	-	-	-	-	-
$T_{max}CO_2$	40	36	37	36	36	36	36	33	25
$T_{Off}CO_2$	40	-	-	-	-	-	-	37	36
DeathAgeingLeavesStart	60	-	-	40	40	-	-	-	-
A_{crit}	0.053	0.043	0.043	-	-	-	-	-	-
A_{min}	0.022	-	-	-	-		0.012	-	-
A_{max}	0.083	-	-	-	-	0.05	-	-	-
K_{ET} BBCH 50	1.05	-	1.1	-	-	-	0.95	-	1.1
K_{ET} BBCH 78	1.05	-	1.1	-	-	-	0.95	-	1.1
K_{ET} BBCH 97	0.9	-	1.85	-	-	-	0.7	-	-

In addition to these parameters, the coefficients of algorithms governing the simulation of evapotranspiration (Table 2), specific partitions for each phenological phase (Table 2), and degree days (GDD; Table 3) for achieving the phenological stages were also modified.

For the emergence, flowering, and maturity stages, an excellent match between the observed and simulated data was achieved, both in terms of similarity of values averaged for all growing seasons and in terms of inter-annual variability (Table 4).

Accurate calibration of crop phenology is considered the primary, basic step in the application of crop simulation models [43]. In our modelling exercise, the emergence and flowering stages of wheat, as formalized by ARMOSA, attained the highest scores, with the latter capable of capturing both the averaged GDD to reach these phenological stages and variability across years.

ARMOSA effectively formalized GDD to reach the maturity stage, albeit with a slight penalty from the low score of EF and a moderate score of d. However, this process was well depicted by NRMSE and CRM figures.

Table 2. Calibrated values of plant partition parameters by cultivar. Only modified values are shown in the table.

Parameter	Default Value	Cultivars							
		Appio	**Appulo**	**Claudio**	**Latino**	**Ofanto**	**Saragolla**	**Simeto**	**Valgerardo**
FDM_{Leaves} BBCH 40	0.4	0.5	-	-	-	-	-	-	-
FDM_{Leaves} BBCH 47	0.4	-	-	-	-	-	-	0.3	-
FDM_{Leaves} BBCH 58	0.4	0.3	-	-	-	-	0.3	-	-
FDM_{Leaves} BBCH 61	0.1	0	0.3	0.3	-	0	0.2	-	0.2
FDM_{Leaves} BBCH 75	0	-	0.2	-	-	-	-	-	-
FDM_{Stem} BBCH 0	1	-	0	-	-	-	-	-	-
FDM_{Stem} BBCH 40	0.6	0.5	-	-	-	-	-	-	0.5
FDM_{Stem} BBCH 47	0.6	-	-	-	-	-	-	0.7	0.5
FDM_{Stem} BBCH 58	0.6	0.4	-	-	-	-	-	-	0.4
FDM_{Stem} BBCH 61	0	-	0.2	0.3	-	-	0.6	-	-
FDM_{Stem} BBCH 75	0	-	0.1	-	-	-	0.2	-	-
FDM_{Stem} BBCH 80	0	-	-	-	-	-	0.1	-	-
$FDM_{Storage}$ BBCH 0	0	-	1	-	-	-	-	-	-
$FDM_{Storage}$ BBCH 40	0	-	-	-	-	-	-	-	0.1
$FDM_{Storage}$ BBCH 50	0	-	-	-	-	-	-	-	0.1
$FDM_{Storage}$ BBCH 58	0	0.3			-	-	-	-	0.2
$FDM_{Storage}$ BBCH 61	0	1	0.5	0.4	0.9	1	0.1	0.9	0.8
$FDM_{Storage}$ BBCH 75	0.9	1	0.7	1	1	1	0.6	1	1
$FDM_{Storage}$ BBCH 80	0	-	-	-	-	-	0.9	-	-
FDM_{Shoot} BBCH 0	0.5	-	0.3	-	-	-	-	-	-
FDM_{Shoot} BBCH 9	0.5	-	0.3	-	-	-	-	-	-
FDM_{Shoot} BBCH 29	0.55	-	0.5	-	-	-	-	-	-
FDM_{Shoot} BBCH 56	0.9	-	-	-	-	-	-	0.85	1

Table 3. Calibrated values of phenological-stage-specific parameters by cultivar. Only modified values are showed in the table.

Parameter	Default Value	Cultivars							
		Appio	**Appulo**	**Claudio**	**Latino**	**Ofanto**	**Saragolla**	**Simeto**	**Valgerardo**
GDD_{sum} Emergence	50	90	-	70	70	70	-	90	60
GDD_{sum} Tillering	400	250	-	250	350	450	300	250	200
GDD_{sum} Flowering	350	250	300	300	-	-	300	-	-
GDD_{sum} Phys. maturity	600	300	220	300	250	200	300	300	350
T_{base} Emergence	5	7	7	-	-	-	-	-	-
T_{base} Tillering	5	-	7	-	-	-	-	-	-
T_{base} Flowering	8	-	5	-	-	-	-	-	-
T_{base} Phys. maturity	8	-	6	-	-	6	7	7	-

A simulation model's accuracy in replicating crop phenology correlates with its ability to capture the genetic variability underlying canopy development and biomass accumulation within the same framework [44].

Biomass accumulation is linked to the amount of radiation intercepted by the leaf surface, which, in turn, is responsible for converting assimilated CO_2 into carbohydrates, a cultivar-specific trait.

In light of this, the coefficients of certain algorithms governing canopy development and senescence, CO_2 conversion into dry matter, maintenance respiration, and water and temperature stress for each cultivar were adjusted to best align with the simulation of biomass accumulation based on data gathered in the LTE (see Table 1).

Regarding phenology, the calibration phase demonstrated the proficiency of ARMOSA in faithfully replicating the total dry biomass at harvest, averaged for all soil treatments (Table 5).

Table 4. Comparison between observed and simulated data for the phenological stages recorded for all cultivars and treatments of P_30. Observed and simulated data on phenology were equal for all cvs. White, light grey, and grey cells indicate the best (1), middle (0.5), and worst (0) scores, respectively.

Parameter	Unit	Obs	Mean		St.Dev.		RMSD	NRMSE	EF	d	CRM	Score
		n°	Obs	Sim	Obs	Sim	(GDD)	(%)				
Emergence	GDD (°C)	43	352	347	18	20	11	3.06	0.70	0.92	−0.01	Very good
Flowering	GDD (°C)	43	129	131	8	10	9	6.82	0.21	0.74	0.01	Very good
Maturity	GDD (°C)	43	158	157	9	8	9	6.01	−0.36	0.62	0.00	Good

Very good = total score from 3.5 to 4; Good = total score from 2.5 to 3.

Table 5. Comparison between simulated and observed data on total dry biomass and performance evaluation indices of the model applied to straw treatments. White and light grey, indicate the best (1) and middle (0.5) scores, respectively.

Parameter	Unit	Obs	Mean		St.Dev.		RMSD	NRMSE	EF	d	CRM	Score
Treatment		n°	Obs	Sim	Obs	Sim	(kg ha^{-1})	(%)				
T2	kg ha^{-1}	36	10,835	10,475	±4005	±3076	2916	26.91	0.45	0.81	0.03	Very good
T5	kg ha^{-1}	36	10,824	11,509	±3884	±4303	2877	26.58	0.44	0.86	−0.06	Good
T8	kg ha^{-1}	37	11,124	11,873	±3696	±4163	2653	23.85	0.47	0.87	−0.07	Very good
P_30	kg ha^{-1}	109	10,930	11,291	±3829	±3898	2816	25.77	0.45	0.85	−0.03	Very good

Very good = total score from 3.5 to 4; Good = total score from 2.5 to 3.

Indeed, the highest score was observed for three out of four evaluation indices, with only a negligible deviation of NRMSE from the optimal value (25.77% vs. 25%).

When evaluating ARMOSA's response for each cropping system separately (T2, T5, and T8), a remarkable match between observed TDM and the model output was evident for T2 and T8. There was a narrow deviation from the optimal value of NRMSE for the former and a slight overestimation of the model for the latter. Nonetheless, even the output of ARMOSA in replicating T5 could be deemed satisfactory, with the best performance for EF and d, but with a slight overestimation and deviation of the simulated data compared to the observed data.

The environment of the area under investigation (Mediterranean climate) is characterized by erratic rainfall patterns, leading to prolonged drought conditions, especially during the spring–summer period.

Additionally, common agronomic practices for durum wheat in the Mediterranean area do not include irrigation. The sum of these conditions subjects the crop to extremely variable water supply and water stress among years and within the same growing season [45–47].

Examining the ratio between the standard deviation and the mean value of TDM revealed that certain cvs were more susceptible to climatic erraticism (e.g., Valgerardo, Latino, Appio) than others (Ofanto and Appulo; Table 6).

Table 6. Comparison between simulated and observed data on total dry biomass and performance evaluation indices of the model applied to individual cultivars. White, light grey, and grey cells indicate the best (1), middle (0.5), and worst (0) scores, respectively.

Parameter	Unit	Obs	Mean		St.Dev.		RMSD	NRMSE	EF	d	CRM	Score
cv		N°	Obs	Sim	Obs	Sim	(kg ha^{-1})	(%)				
Appio	kg ha^{-1}	12	9148	8713	±2377	±939	2473	27.03	−0.18	0.39	0.05	Fair
Appulo	kg ha^{-1}	12	10,346	10,706	±1473	±667	1625	15.18	−0.33	0.32	−0.03	Fair
Claudio	kg ha^{-1}	13	15,911	14,709	±3792	±3948	3368	21.17	0.15	0.79	0.08	Good
Latino	kg ha^{-1}	15	7393	8953	±2607	±2546	2250	30.43	0.20	0.81	−0.21	Fair
Ofanto	kg ha^{-1}	12	10,981	11,340	±827	±2320	2318	21.11	−7.58	0.41	−0.03	Good
Saragolla	kg ha^{-1}	9	15,517	17,035	±3207	±5599	6427	41.42	−3.52	0.40	−0.12	Bad
Simeto	kg ha^{-1}	24	11,346	11,808	±2834	±2200	1971	17.37	0.50	0.83	−0.06	Very good
Valgerardo	kg ha^{-1}	12	7411	7045	±2187	±1953	737	9.94	0.88	0.97	0.05	Very good

Very good = total score from 3.5 to 4; Good = total score from 2.5 to 3; Fair = total score from 1.5 to 2; Bad = total score from 0 to 1.

The variability observed in the experimental yield data was influenced by climatic variables, including rainfall, which exhibited high variability with differences of up to 400 mm across growing seasons, and temperature (especially during the flowering and grain-filling period).

Lower yields were recorded in years with lower precipitation during the crop growing period (around 300 mm in 1982 and 1992). The best performances were noted in growing seasons where precipitation ensured water inputs exceeding 430 mm, particularly in 1991, 1998, and several years ranging from 2018 to 2021.

Detrimental effects of temperatures on productivity were observed in years when grain yield was not satisfactory (i.e., 1989, 1995, 2007, and 2020). In these instances, average mean temperatures reached peaks of 24–28 °C between the beginning of flowering and the waxy maturity stage of the seed (mid-April to the second ten days of May), mainly due to anomalies in maximum temperatures (heatwaves) leading to pollination failure and/or reductions in grain mass.

Accordingly, a meticulous calibration of the crop coefficients related to the mechanisms of adaptation to temperature and rainfall pattern and any water/temperature stress (i.e., WSPar, or susceptibility of the crop to water stress; TmaxCO2, or the maximum temperature threshold for the optimal development of the crop; TOffCO2, that is, the temperature above which crop growth ceases; and KET, which represents the crop coefficient at specific phenological stages of the crop) was performed for each cv.

Among 8 cvs, ARMOSA was able to accurately replicate TDM at the end of growing season for 4 of them and produced fairly good replications for 3 cvs; there was only one cv for which the simulation was not satisfactory.

It should be noted that for Saragolla, we investigated only three growing seasons (from 2019 to 2021), leading to a limited number of observations not sufficient to optimize ARMOSA's response for this cv.

Simeto and Valgerardo were the cvs for which ARMOSA accurately simulated both the inter-annual variability and the average TDM observed in the field, with a slight overestimation for Simeto.

For the remaining cvs, there was a mixed response; for some of them, ARMOSA was efficient in replicating the biomass accumulation at harvest, returning negligible differences between the observed and simulated mean data, but less effective in capturing the variability between the various years (see NRMSE, EF, and d for Appulo, Claudio, and Ofanto).

For other cvs, the simulations comprehensively captured the inter-annual variability (i.e., Claudio and Latino) but overestimated or underestimated the average trend of TDM.

Definitively, by analysing the response of ARMOSA in simulating TDM at harvest, it emerged that the calibration process correctly trained the cropping-system model to effectively replicate the data observed in the field across the LTE under P_30 treatments.

Thus, the correct estimate of TDM by ARMOSA and therefore of biomass incorporated in the soil was the first key point for an adequate simulation of TOC dynamics.

In previous studies, ARMOSA was calibrated and validated on a wide range of climate and soil conditions throughout Europe under conventional systems and CA, simulating TOC dynamics with very good or even excellent results [22].

Thus, the calibration step for the TOC dynamic focused only on two parameters controlling the evolution of soil organic matter, namely Khumus (1.4×10^{-4}) and CMicrob-Efficiency (0.4), leaving all the other parameters unchanged.

ARMOSA replicated the dynamics of TOC quite favourably, achieving the "Good" score for all the treatments under investigation (Table 7; Figure 2). This result was attributed to the accurate estimation of the mean value of TOC (averaged for all treatments; 64,965 vs. 64,758 kg ha^{-1}, Table 7).

Table 7. Comparison between simulated and observed data on TOC (0–40 cm) for P_30 and performance evaluation indices of the model applied to each treatment. White, light grey, and grey cells indicate the best (1), middle (0.5), and worst (0) scores, respectively.

Parameter	Unit	Obs	Mean		Dev.st		RMSD	NRMSE	EF	d	CRM	Score
Treatment		N°	Obs	Sim	Obs	Sim	(kg ha^{-1})	(%)				
T2	kg ha^{-1}	8	66,345	64,549	±5738	±4279	6371	9.60	−0.41	0.50	0.03	Good
T5	kg ha^{-1}	13	64,313	65,060	±5700	±4390	6071	9.44	−0.23	0.57	−0.01	Good
T8	kg ha^{-1}	13	64,226	65,127	±5517	±3373	5780	9.00	−0.19	0.36	−0.01	Good
P_30	kg ha^{-1}	34	64,758	64,965	±5537	±3883	6035	9.32	−0.22	0.48	0.00	Good

Good = total score from 2.5 to 3.

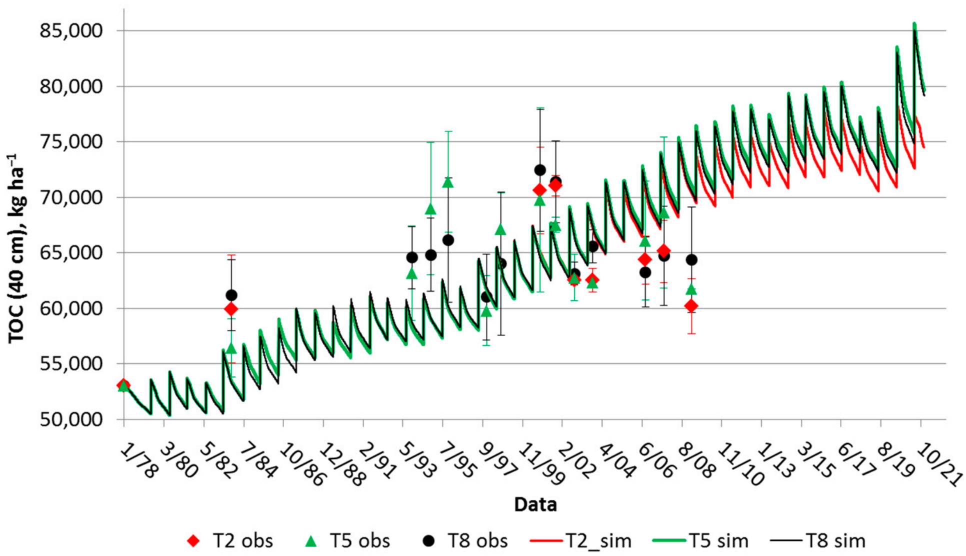

Figure 2. Comparison between simulated and observed data on TOC (0–40 cm) for P_30. Bars indicate the standard deviation.

While the CRM index indicated a perfect alignment of the simulated values with the measured ones, it is noteworthy that ARMOSA tended to slightly underestimate the data collected in the initial course of the LTE and then overestimate the data in the middle period of the LTE (Figure 2).

Measuring the robustness of ARMOSA in formalizing TOC dynamics in the last part of the LTE was not possible due to the absence of soil sampling, which occurred during the validation phase (as discussed in the next section).

The high variability of measured TOC, both between consecutive years and within the same sampling (indicated by a high standard deviation), was highlighted (Figure 2).

The source of this erraticism may be a series of conditions associated with the sampling time and sampling point. The sampling dates over the years ranged from the beginning of September to the end of November. During this period, straw could be intact (i.e., early September) or already partially degraded (i.e., late November), a state also related to the time of their burial with respect to the soil sampling. This could affect the amounts of organic matter and organic carbon in the shallow layers of soil as well as the sampling point, which could be affected by the substantial content (and dynamics) of crop residues [22–48].

This might explain the diminished correspondence between the measured and simulated variability of TOC (indicated by low EF and d scores). Nevertheless, ARMOSA successfully captured the high variability of this variable between the beginning and end of the growing period, attributed to the dynamic degradation of straw.

Divergent results emerged from the simulation of grain yield (refer to Table 8).

Table 8. Comparison between simulated and observed data on grain yield at harvest for P_30 and performance evaluation indices of the model applied to each treatment. White, light grey, and grey cells indicate the best (1), middle (0.5), and worst (0) scores, respectively.

Parameter	Unit	Obs	Mean		Dev.st		RMSD	NRMSE	EF	d	CRM	Score
		N°	Obs	Sim	Obs	Sim	(kg ha^{-1})	(%)				
T2	kg ha^{-1}	40	3074	2832	±1214	±1382	1175	38.22	0.04	0.78	0.08	Good
T5	kg ha^{-1}	40	2735	3114	±1206	±1549	1413	51.66	−0.41	0.69	−0.13	Bad
T8	kg ha^{-1}	41	2902	3265	±1116	±1509	1411	48.62	−0.63	0.66	−0.13	Bad
Total	kg ha^{-1}	121	2904	3072	±1178	±1481	1338	46.07	−0.30	0.71	−0.06	Fair

Good = total score from 2.5 to 3; Fair = total score from 1.5 to 2; Bad = total score from 0 to 1.

Although the simulated total score of yield averaged for all treatments was "Fair", only for T2 was a good result achieved, while, for the other two treatments, the outcome was not adequate.

This pattern was consequently confirmed for the simulated yield of several cultivars. Among the eight cultivars, half did not attain a satisfactory score, three achieved a fairly good score, and only one reached the maximum score (see Table 9). NRMSE ranged from a minimum of 24.45% for Latino to a maximum of 66.51% for Claudio. The latter had a poor fit in the calibration test with EF (−9.93) and CRM (−0.23), which were the worst among the simulated varieties. In addition to Latino, the calibration of Simeto allowed satisfactory performance in terms of EF (0.1) and d (0.77), followed by Valgerardo (0.18 and 0.83 for EF and d, respectively).

Table 9. Comparison between simulated and observed data on grain yield and performance evaluation indices of the model applied to individual cultivars. White, light grey, and grey cells indicate the best (1), middle (0.5), and worst (0) scores, respectively.

Parameter	Unit	Obs N°	Mean Obs	Mean Sim	Dev.st Obs	Dev.st Sim	RMSD (kg ha^{-1})	NRMSE (%)	EF	d	CRM	Score
Appio	kg ha^{-1}	12	2325	2361	±1031	±393	1239	53.3	−0.58	0.1	−0.02	Bad
Appulo	kg ha^{-1}	12	2903	3114	±306	±353	1564	19.39	−2.72	0.23	−0.06	Fair
Claudio	kg ha^{-1}	13	3754	4618	±786	±2624	2497	66.51	−9.93	0.37	−0.23	Bad
Latino	kg ha^{-1}	15	2135	2029	±1093	±912	524	24.54	0.71	0.92	0.05	Very good
Ofanto	kg ha^{-1}	15	3092	2641	±437	±989	1066	34.47	−5.38	0.42	0.15	Bad
Saragolla	kg ha^{-1}	9	3049	2966	±1293	±1091	2095	68.71	−1.96	0.06	0.03	Bad
Simeto	kg ha^{-1}	33	3477	3818	±1274	±1292	1190	34.22	0.1	0.77	−0.10	Fair
Valgerardo	kg ha^{-1}	12	1600	1973	±817	±1044	708	44.25	0.18	0.83	−0.23	Fair

Very good = total score from 3.5 to 4; Fair = total score from 1.5 to 2; Bad = total score from 0 to 1.

The unsatisfactory outcome for Saragolla should also be highlighted, as EF and d deviated significantly from the optimum values, despite the simulation of the mean yield aligning with observed data (CRM of 0.03).

Calibration of ARMOSA was focused on the parameters controlling the partition of the biomass between the different organs, therefore reflecting the grain and the maintenance respiration of the same (Table 2).

The observed data showed that the grain yield was not linearly related to the biomass produced at harvest.

Several authors reported poor performance when calibrating crop simulation models on wheat yield across different sites, years, and cultivars, especially in hot–arid environments.

Specifically, some authors claimed that grain production depended on genetic coefficients that were not only site-specific [49] but also year-specific [50,51].

Our results after the calibration of ARMOSA confirm what was reported by [52], who stated that it was difficult to accurately predict the production of wheat with low levels and/or in environments characterized by high temperatures.

The simulation of grain production becomes challenging when situations of water and/or thermal stress occur during seed formation [53].

In the climatic conditions of the experimental site, recurrent periods of low rainfall and heat waves significantly compromised the potential productivity of the crop. Additionally, the occurrence of short but intense storms and strong gusts of wind resulted in lodging of the crop. These extreme events during seed filling, which significantly impact the final yield, are rarely formalized by crop growth simulation models [54].

Nevertheless, the 1:1 regression line depicting observed and simulated data (Figure 3) demonstrated the commendable fitness of ARMOSA in capturing the variability of the average grain yield among cultivars (Table 8), evidenced by an R^2 value of 0.82 and an angular coefficient of 1.06.

The calibration procedure involved an intricate adjustment of parameters underlying crop growth, encompassing CO_2 assimilation, biomass conversion, organ separation, canopy development, intercepted radiation, root length, and senescence. Phenological stages, including emergence, flowering, and maturity, achieved an excellent match between observed and simulated data, underscoring the importance of accurate calibration in capturing genetic variability.

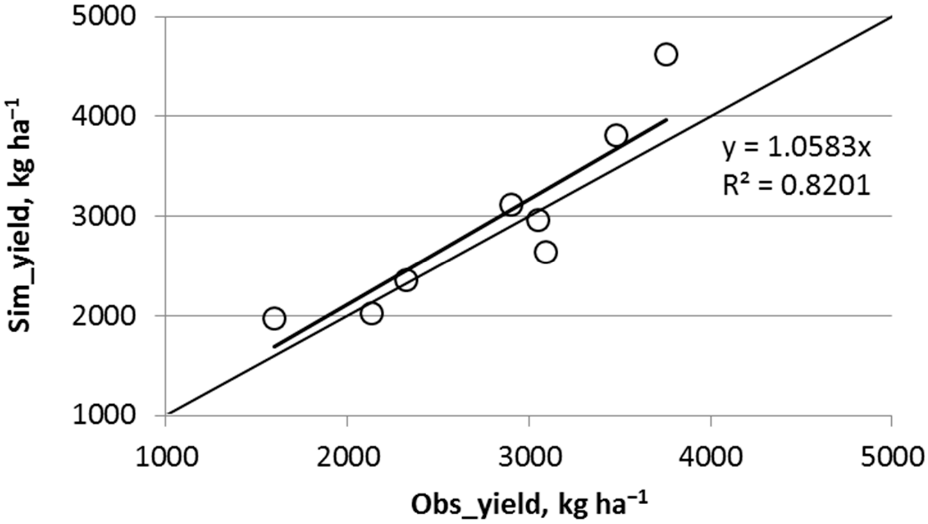

Figure 3. Linear regression (thin line) between observed grain yield (Obs_yield) and simulated grain yield (Sim_yield) of P_30. Empty circles indicate the yield averaged for each cultivar. Thin black line indicates 1:1 fit.

Beyond phenology, the calibration phase scrutinized biomass accumulation and cultivar-specific adaptation to environmental stressors. Different cultivars exhibited varying susceptibility to climatic erraticism, necessitating meticulous calibration of crop coefficients related to temperature, rainfall patterns, and water/temperature stress. ARMOSA demonstrated varying success in replicating total dry biomass at the end of the growing season for different cultivars, reflecting the intricacies of cultivar-specific responses to environmental variations.

The simulation of grain yield emerged as a challenging aspect, with ARMOSA demonstrating a tendency to slightly overestimate yields and exhibiting broader sensitivity to climate patterns than the actual plant dynamics. The nonlinear relationship between grain yield and biomass produced at harvest added an extra layer of complexity to the calibration process. Despite these challenges, ARMOSA presented a commendable ability to capture the variability of average grain yield among cultivars, demonstrating the model's aptitude in predicting wheat productivity under diverse conditions.

3.2. Validation

ARMOSA's performance in simulating phenology remained consistent during validation, achieving maximum scores for emergence and flowering.

While the formalization of maturity stage did not attain the same degree of excellence (EF of −1.05 and CRM of 0.46), ARMOSA closely aligned with the observed mean values (156 days vs. 155 days; Table 10).

Table 10. Comparison between observed and simulated data for the phenological stages recorded for all cultivars and treatments of P_32 in the validation step. Observed and simulated data on phenology were equal for all cvs. White, light grey, and grey cells indicate the best (1), middle (0.5), and worst (0) scores, respectively.

Parameter	Unit	Obs	Mean		Dev.st		RMSD	NRMSE	EF	d	CRM	Score
		n°	Obs	Sim	Obs	Sim	(GDD)	(%)				
Emergence	GDD	14	365	368	27	35	13	3.56	0.74	0.95	−0.01	Very good
Flowering	GDD	14	123	128	7	11	10	8.13	−1.11	0.63	−0.04	Very good
Maturity	GDD	14	155	156	6	7	9	5.81	−1.05	0.46	−0.01	Good

Very good = total score from 3.5 to 4; Good = total score from 2.5 to 3.

Indication on the reliability of ARMOSA in replicating the productivity of the cvs (Simeto, Claudio, and Saragolla) throughout the validation process were drawn from the results of the 1:1 regression (Table 11).

Table 11. Comparison between observed and simulated data on grain yield in the validation step and main parameters of the related linear regression.

Parameter	Unit	Obs n°	Mean Obs	Mean Sim	Dev.st Obs	Dev.st Sim	*R-Squared*	p-Val (Fit)	β	p-Val (β)
Simeto	kg ha^{-1}	8	3267	4416	±957	±720	0.87	<0.001	1.24	<0.001
Claudio	kg ha^{-1}	7	4300	4392	±617	±2027	0.86	<0.001	1.02	<0.001
Saragolla	kg ha^{-1}	2	3089	2867	±656	±402	0.99	<0.001	0.92	<0.001
NT	kg ha^{-1}	17	3703	4202	±953	±1481	0.86	<0.001	1.08	<0.001
MT	kg ha^{-1}	17	3684	4246	±963	±1478	0.87	<0.001	1.11	<0.001
P_32	kg ha^{-1}	34	3676	4224	±944	±1457	0.87	<0.001	1.1	<0.001

The average value of the grain yield of Claudio was aligned between the model output and the observed data (4300 kg ha^{-1} vs. 4392 kg ha^{-1}). Although the standard deviation was much higher in ARMOSA than in the LTE data, the model reasonably captured the observed variability among years (see dispersion around the 1:1 regression line). What turned out to be off scale were the outcomes related to a single growing season for NT and MT, in which the simulated values (8154 kg ha^{-1}, as mean) were much higher than the observed productivity (4565 kg ha^{-1}).

For Saragolla, ARMOSA was inclined to slightly underestimate the actual yield (β = 0.92), but with an excellent fit between simulated and observed data (R^2 = 0.99), even if the compared growing seasons numbered only two for a total of four yield productivity figures.

For Simeto, the overestimation of grain production by ARMOSA was around 24% (3267 kg ha^{-1} vs. 4416 kg ha^{-1}). As for Claudio, a very high inconsistency between the output and the actual grain yield was observed for one growing season (2349 kg ha^{-1} vs. 5919 kg ha^{-1} as mean), but Simeto definitively proved to be the most difficult cv for ARMOSA to predict (although not dramatically) in the validation phase.

Evaluating ARMOSA overall for NT and MT treatments, the tendency of the model to slightly overestimate (+10%) the observed grain productivity was highlighted, to which was added the larger variability generated by the model, as computed by the coefficient of variation (ratio between the standard deviation and the mean), which was approximately 35% for ARMOSA and 26% for the LTE.

Examining ARMOSA's overall performance for different treatments highlighted a tendency to slightly overestimate yield (+10%), coupled with increased variability compared to observed plant dynamics (CV, defined as the ratio of the standard deviation to the mean equal to 34% for ARMOSA and 26% for the LTE).

Testing the response of ARMOSA in formalizing TOC (Figure 4b), it emerged how the model responded differently to the two soil treatments (NT and MT) and how the outputs aligned with what was observed during the LTE.

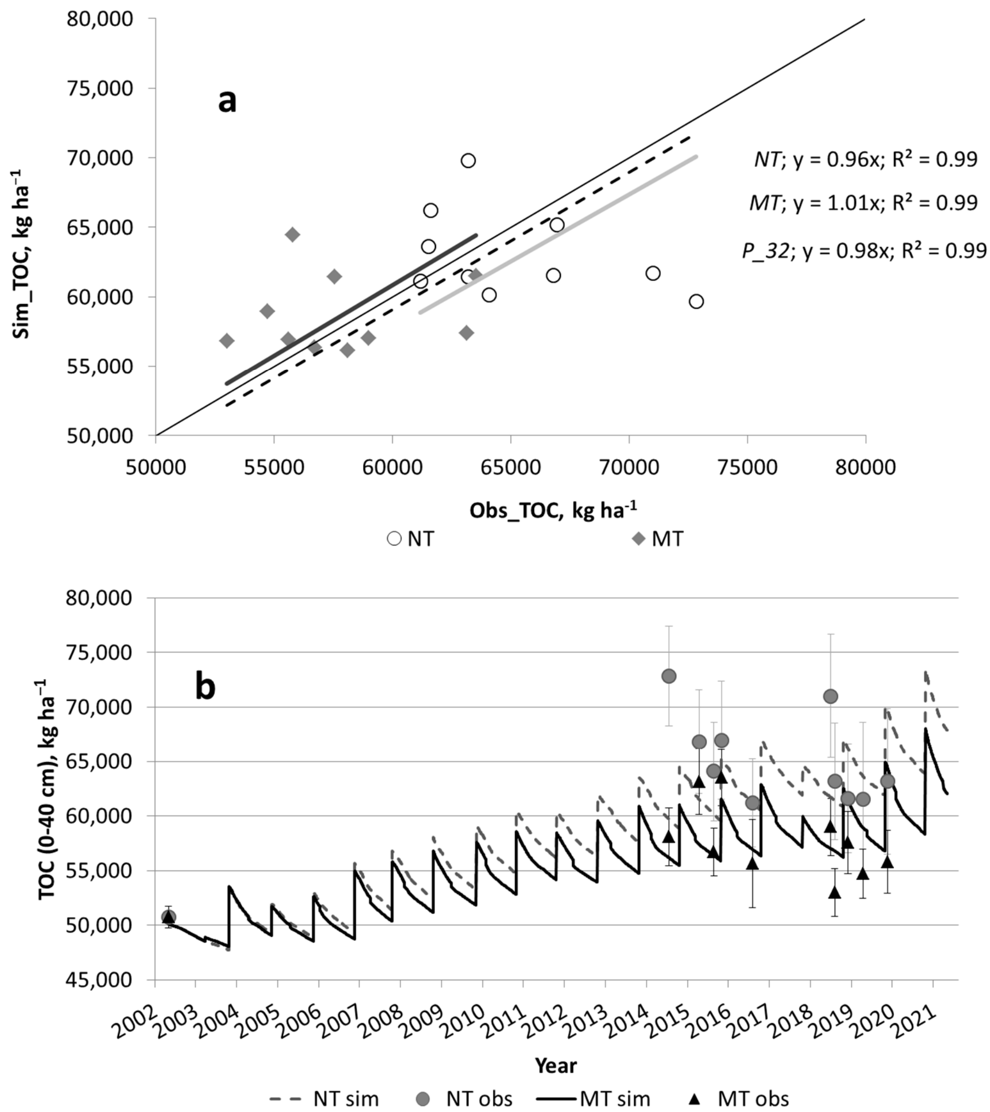

Figure 4. Linear regression between observed TOC (Obs_TOC) and simulated yield (Sim_TOC) achieved by *NT* (grey line), *MT* (thin black line), P_32 (dashed line) in the validation step (**a**); thin black line indicates 1:1 fit. TOC (0–40 cm) dynamics of observed (obs) and simulated (sim) *NT* and *MT* verified across experimental years of LTE (**b**).

Indeed, in the LTE, TOC went from about 51,000 kg ha^{-1} at the beginning of the experimental test (2002) to 63,200 kg ha^{-1} in NT and 55,800 kg ha^{-1} in MT, respectively, in 2020.

ARMOSA did not go far from the observed data, returning TOC values of 63,045 kg ha^{-1} and 65,247 kg ha^{-1} for NT and MT, respectively, for 2020.

This contrasts with the results of comparing the simulated and observed data for some of the several experimental years (i.e., 2015 and 2019), in which substantial differences were recorded between ARMOSA outputs and actual soil TOC content.

This is because TOC determined by laboratory analysis is strongly affected by the organic substances deriving from the total or partial degradation of crop residues, the content of which can be extremely variable depending on the sampling point [48].

This also explains the extreme variability of the figures (see standard deviation in Figure 4b) observed for each sampling, both in NT and in MT.

Concerning TOC, what has been achieved represents a judicious compromise between the performances obtained in the calibration and validation steps, as optimizing one phase over the other would have diminished the overall modelling capacity of the model with respect to TOC.

The TOC pattern in MT, though showing an increase, suggests a more moderate impact on carbon sequestration than NT. Some soil disturbance in MT may accelerate decomposition, but the overall effect remained positive for organic carbon accumulation. This resulted in an annual increase in TOC ranging from 114 kg ha^{-1} to 290 kg ha^{-1} when simulating CA practices such as NT and MT. Similar findings were also indicated by other long-term modelling exercises with ARMOSA [22], where, under current climatic conditions, the TOC increase reached up to 320 kg ha^{-1}.

Although simulations for all T2, T5, and T8 options resulted in an increase in TOC over the course of the wheat monocropping, the latter two showed slightly better performance than T2 during the steady-state phase.

The limited positive impact of water and nitrogen additions to straw on the dynamics of TOC accumulation in T5 and T8, as indicated by ARMOSA simulations, may be ascribable to the timing of water and nitrogen supply.

In the simulations, mirroring the experimental conditions, mineral nitrogen and water were applied during the summer period, characterized in the study environment by very high daytime temperatures (up to 40 °C). These conditions promoted water evaporation and reduced the activity of microorganisms involved in the mineralization of organic matter (with a correlated reduction in nitrogen supply), thus mitigating a more disruptive effect on TOC accumulation in the soil.

The better TOC dynamics (even if not dramatic) in T5 and T8, according to ARMOSA simulations, favoured a slightly higher yield performance than T2.

This indicates that soil health, using soil organic carbon as an indicator, also promotes an improvement in crop productivity. This is due to a more gradual release and increased availability of nitrogen from the soil to the crop. Following ARMOSA's recommendations, the surface release of straw (NT) or shallow burial (MT), without prior chopping, favoured higher grain yield. This was associated with the mulching effect of residues on the soil, leading to a reduction in water loss through evaporation.

ARMOSA's performance in simulating phenology remained consistent during validation, achieving maximum scores for emergence and flowering. While the formalization of the maturity stage did not attain the same degree of excellence, ARMOSA closely aligned with the observed mean values. This reinforced the model's reliability in capturing critical phenological events, crucial for understanding crop development and predicting growth patterns.

Examining ARMOSA's overall performance for different treatments highlighted a tendency to slightly overestimate yield, coupled with increased variability compared to observed plant dynamics. The model's broader sensitivity to varying climate patterns was evident, indicating areas for potential refinement in predicting crop performance under diverse conditions.

The simulation of TOC dynamics during validation underscored ARMOSA's ability to capture fluctuations, particularly during long-term evolution under different crop systems. The model closely aligned with observed TOC values for the year 2020 but exhibited discrepancies in certain experimental years.

The observed variability, attributed to challenges in consistently collecting soil samples under identical conditions, during the same period, and at the same sampling points across different experimental years, underscores the difficulties in accurately simulating TOC by ARMOSA.

In light of that, ARMOSA can be considered reliable in the simulation of TOC fluctuation, particularly if one considers the evolution over a period long enough to capture the correct dynamics of TOC under different crop systems [15–55].

This ARMOSA study, with a primary focus on its application in the Mediterranean environment, establishes a robust foundation for comprehending crop dynamics and soil processes. While the specific regional context is integral, the findings carry substantial implications for global agriculture. The model's adaptability and reliability, validated through successful calibration in the Mediterranean, suggest its potential utility across diverse climates and agricultural systems.

Beyond the Mediterranean scope, the study provides extensive insights into crop responses under various conditions and agronomic practices. As it accurately simulates crop- and soil-related variables, ARMOSA emerged as a versatile tool that can be fine-tuned to suit the unique conditions of various regions, making it valuable for researchers, agronomists, and policymakers involved in optimizing crop growth across different environments. A noteworthy aspect is the model's detailed impact analysis of soil organic carbon dynamics due to agronomic practices, amplifying its applicability, particularly in regions striving to enhance soil fertility, water retention, and overall ecosystem resilience.

As it accurately simulates crop- and soil-related variables, ARMOSA emerged as a versatile tool that can be fine-tuned to suit the unique conditions of various regions, making it valuable for researchers, agronomists, and policymakers involved in optimizing crop growth across different environments.

4. Conclusions

In this modelling application, the ARMOSA crop growth simulation model underwent rigorous testing to assess its reliability in replicating key variables of durum wheat growth, including phenology, dry biomass accumulation, and grain yield. The crop was cultivated under five distinct soil and straw management scenarios, with a focus on their implications for TOC dynamics.

The calibration of ARMOSA, performed on eight durum wheat phenotypes, yielded favourable outcomes for phenology and biomass at harvest across most investigated cultivars. However, the simulation of grain yield presented varying degrees of success, with some cultivars being replicated effectively, while others exhibited unsatisfactory results. During the subsequent validation phase, ARMOSA demonstrated a reasonable ability to capture the average grain yield across multiple growing seasons, despite occasional deviations in specific years.

Replicating productivity in hot–arid environments with low grain yields emerged as a challenging aspect of applying simulation models, aligning with previous findings in the field.

In terms of TOC dynamics, ARMOSA demonstrated proficiency in replicating the overall trend of soil organic carbon both in the calibration and validation processes. While challenges existed in precisely capturing year-to-year variability (attributable to the inherent spatial variability of this parameter), ARMOSA effectively mirrored the progression of TOC over the considered timeframe within the LTE.

While improvements are desirable to enhance the model's response to heat waves and prolonged drought effects on final grain yield, the study's global relevance is noteworthy. Definitively, while the study's focus is on a Mediterranean environment, its findings carry broader significance globally. ARMOSA's adaptability, demonstrated in replicating wheat growth and TOC dynamics, suggests potential applicability in different regions and agronomic practices. The resulting insights into crop response under varying conditions contribute to a broader understanding of sustainable agriculture. Regarding the research's global relevance, the findings may guide crop modelling efforts and inform agronomic strategies in different climates, fostering sustainable practices worldwide.

Author Contributions: All the authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: This research is part of the P.S.R. 2014/2020 (Sottomisura 16.2) project (Apulia region), as well as "Smart Future Organic Farm, Un metodo innovativo (monitorabile, misurabile e certificabile) di produzione biologica verso un'agricoltura a 0 emissioni di CO_2" (management document 28 July 2020 n. 030_DIR_2020_00170) and "Water4AgriFood, Miglioramento delle produzioni agroalimentari mediterranee in condizioni di carenza di risorse idriche", PNR 2015–2020, + PON ARS01_00825 "Ricerca e Innovazione" 2014–2020. The authors want to acknowledge Marcello Mastrangelo for his skilful work in laboratory analysis.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. United States Department of Agriculture. Foreign Agricultural Service (USDA). *World Agricultural Production*; Circular Series; USDA: Washington DC, USA, 2018.
2. ISTAT. Istituto Nazionale di Statistica. Agricoltura e Zootecnia. 2019. Available online: http://agri.istat.it (accessed on 8 January 2024).
3. IPCC. Part A: Global and Sectoral Aspects. (Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change). In *Climate Change 2014: Impacts, Adaptation and Vulnerability*; IPCC: Geneva, Switzerland, 2014; p. 1132.
4. Bird, D.N.; Benabdallah, S.; Gouda, N.; Hummel, F.; Koeberl, J.; La Jeunesse, I.; Meyer, S.; Prettenthaler, F.; Soddu, A.; Woess-Gallasch, S. Modelling climate change impacts on and adaptation strategies for agriculture in Sardinia and Tunisia using AquaCrop and Value-at-Risk. *Sci. Total Environ.* **2016**, *543*, 1019–1027. [CrossRef] [PubMed]
5. Liu, K.; Harrison, M.T.; Yan, H.; Liu, D.L.; Meinke, H.; Hoogenboom, G.; Wang, B.; Peng, B.; Guan, K.; Jaegermeyr, J.; et al. Silver lining to a climate crisis in multiple prospects for alleviating crop waterlogging under future climates. *Nat. Commun.* **2023**, *14*, 765. [CrossRef] [PubMed]
6. IPCC. Summary for Policymakers. In *Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems*; IPCC: Geneva, Switzerland, 2019.
7. Rosenzweig, C.; Elliott, J.; Deryng, D.; Ruane, A.C.; Müller, C.; Arneth, A.; Boote, K.J.; Folberth, C.; Glotter, M.; Khabarov, N.; et al. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proc. Natl. Acad. Sci. USA* **2013**, *111*, 3268. [CrossRef] [PubMed]
8. Ciscar, J.C.; Feyen, L.; Ibarreta, D.; Soria, A. *Climate Impacts in Europe: Final Report of the JRC Peseta III Project, EUR 29427 EN*; Publications Office of the European Union: Luxemburg, 2018; ISBN 978-92-79-97218-8. [CrossRef]
9. Hristov, J.; Toreti, A.; Perez Domínguez, I.; Dentener, F.; Fellmann, T.; Elleby, C.; Ceglar, A.; Fumagalli, D.; Niemeyer, S.; Cerrani, I.; et al. *Analysis of Climate Change Impacts on EU Agriculture by 2050*; Publications Office of the European Union: Luxembourg, 2020; ISBN 978-92-76-10617-3. [CrossRef]
10. Mereu, V.; Gallo, A.; Trabucco, A.; Carboni, G.; Spano, D. Modeling high-resolution climate change impacts on wheat and maize in Italy. *Clim. Risk Manag.* **2021**, *33*, 100339. [CrossRef]
11. Ventrella, D.; Stellacci, A.M.; Castrignanò, A.; Charfeddine, M.; Castellini, M. Effects of crop residue management on winter durum wheat productivity in a long term experiment in Southern Italy. *Eur. J. Agron.* **2016**, *77*, 188–198. [CrossRef]
12. Kassam, A.; Friedrich, T.; Derpsch, R. Global Spread of Conservation Agriculture Global Spread of Conservation Agriculture. *Int. J. Environ. Stud.* **2019**, *76*, 29–51. [CrossRef]
13. Liu, S.; Yang, J.Y.; Zhang, X.Y.; Drury, C.F.; Reynolds, W.D.; Hoogenboom, G. Modelling crop yield, soil water content and soil temperature for a soybean-maize rotation under conventional and conservation tillage systems in Northeast China. *Agric. Water Manag.* **2013**, *123*, 32–44. [CrossRef]
14. Shafeeq, M.P.; Aggarwal, P.; Krishnan, P. Modeling the temporal distribution of water, ammonium-N and nitrate-N in the root zone of wheat using HYDRUS-2D under conservation agriculture. *Environ. Sci. Pollut. Res.* **2020**, *27*, 2197–2216. [CrossRef] [PubMed]
15. Puig-Sirera, À.; Acutis, M.; Bancheri, M.; Bonfante, A.; Botta, M.; De Mascellis, R.; Orefice, N.; Perego, A.; Russo, M.; Tedeschi, A.; et al. Zero-tillage effects on durum wheat productivity and soil-related variables in future climate scenarios: A modeling analysis. *Agronomy* **2022**, *12*, 331. [CrossRef]
16. Ngwira, A.R.; Aune, J.B.; Thierfelder, C. DSSAT Modelling of Conservation Agriculture Maize Response to Climate Change in Malawi. *Soil Tillage Res.* **2014**, *143*, 85–94. [CrossRef]
17. Corbeels, M.; Chirat, G.; Messad, S.; Thierfelder, C. Performance and Sensitivity of the DSSAT crop growth model in simulating maize yield under conservation agriculture. *Eur. J. Agron.* **2016**, *76*, 41–53. [CrossRef]
18. Chaki, A.K.; Gaydon, D.S.; Dalal, R.C.; Bellotti, W.D.; Gathala, M.K.; Hossain, A.; Menzies, N.W. How we used APSIM to simulate conservation agriculture practices in the rice-wheat system of the eastern gangetic plains. *Field Crop. Res.* **2022**, *275*, 108344. [CrossRef]

19. Matthews, R.B.; Rivington, M.; Muhammed, S.; Newton, A.C.; Hallett, P.D. Adapting crops and cropping systems to future climates to ensure food security: The role of crop modelling. *Glob. Food Sec.* **2013**, *2*, 24–28. [CrossRef]
20. Corbeels, M.; Berre, D.; Rusinamhodzi, L.; Lopez-Ridaura, S. Can we use crop modelling for identifying climate change adaptation options? *Agric. For. Meteorol.* **2018**, *256–257*, 46–52. [CrossRef]
21. Perego, A.; Giussani, A.; Sanna, M.; Fumagalli, M.; Carozzi, M.; Alfieri, L.; Brenna, S.; Acutis, M. The ARMOSA simulation crop model: Overall features, calibration and validation. *Italian J. Agrometeorol.* **2013**, *3*, 23–38.
22. Valkama, E.; Kunypiyaeva, G.; Zhapayev, R.; Karabayev, M.; Zhusupbekov, E.; Perego, A.; Schillaci, C.; Sacco, D.; Moretti, B.; Grignani, C.; et al. Can conservation agriculture increase soil carbon sequestration? A modelling approach. *Geoderma* **2020**, *369*, 114298. [CrossRef]
23. USDA. *Soil Taxonomy Classification*; United States Department of Agricultura: Washington, DC, USA, 2010.
24. Emberger, L. *Climate Biogeographic Classification*; Collection of Botanical Geological and Zoological Laboratories Works; Montpellier Faculty of Science: Montpellier, France, 1945; Volume 7, pp. 3–43.
25. Savabi, M.R.; Williams, J.R. Water balance and percolation. In *USDA-Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation*; NSERL Report No. 10; Flanagan, D.C., Nearing, M.A., Eds.; National Soil Erosion Research Laboratory, USDA-Agricultural Research Service: West Lafayette, IN, USA, 1995; Chapter 5; pp. 5.1–5.14.
26. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration-Guidelines for Computing Crop Water Requirements-FAO Irrigation and Drainage Paper 56*; FAO-Food and Agriculture Organisation of the United Nations: Rome, Italy, 1998; Available online: https://www.fao.org/3/x0490e/x0490e00.htm (accessed on 10 January 2020).
27. Sinclair, T.R.; Muchow, R.C.; Ludlow, M.M.; Leach, G.J.; Lawn, R.J.; Foale, M.A. Field and model analysis of the effect of water deficits on carbon and nitrogen accumulation by soybean, cowpea and black gram. *Field Crop Res.* **1987**, *17*, 121–140. [CrossRef]
28. van Diepen, C.A.; Wolf, J.; van Keulen, H.; Rappoldt, C. WOFOST: A simulation model of crop production. *Soil Use Manag.* **1989**, *5*, 16–24. [CrossRef]
29. Johnsson, H.; Bergstrom, L.; Jansson, P.E.; Paustian, K. Simulated Nitrogen Dynamics and Losses in a Layered Agricultural Soil. *Agric. Ecosyst. Environ.* **1987**, *18*, 333–356. [CrossRef]
30. Flanagan, D.C.; Gilley, J.E.; Franti, T.G. Water Erosion Prediction Project (WEPP): Development history, model capabilities and future enhancements. *Trans. ASABE* **2007**, *50*, 1603–1612. [CrossRef]
31. Rádics, J.P.; Jóri, I.J.; Fenyvesi, L. Soil CO_2 emission induced by tillage machines. *Int. J. Appl. Sci.* **2014**, *4*, 37–44.
32. Chen, F.; Dudhia, J. Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Weather Rev.* **2001**, *129*, 569–585. [CrossRef]
33. Acutis, M.; Confalonieri, R. Optimization algorithms for calibrating cropping systems simulation models. A case study with simplex-derived methods integrated in the WARM simulation environment. *Ital. J. Agrometeorol.* **2006**, *3*, 26–34.
34. Campolongo, F.; Cariboni, J.; Saltelli, A. An effective screening design for sensitivity analysis of large models. *Environ. Modell. Softw.* **2007**, *22*, 1509–1518. [CrossRef]
35. Nash, J.E.; Sutcliffe, J.V. River Flow Forecasting through Conceptual Model. Part 1—A Discussion of Principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]
36. Fox, D.G. Judging air quality model performance: A summary of the AMS workshop on dispersion models performance. *Bull. Am. Meteorol. Soc.* **1981**, *62*, 599–609. [CrossRef]
37. Jørgensen, S.E.; Kamp-Nielsen, L.; Christensen, T.; Windolf-Nielsen, J.; Westergaard, B. Validation of a prognosis based upon a eutrophication model. *Ecol. Model.* **1986**, *35*, 165–182. [CrossRef]
38. Greenwood, D.J.; Neeteson, J.J.; Draycott, A. Response of potatoes to N fertilizer: Dynamic model. *Plant Soil.* **1985**, *85*, 185–203. [CrossRef]
39. Willmott, C.J.; Wicks, D.E. An empirical method for the spatial interpolation of monthly precipitation within California. *Phys. Geogr.* **1980**, *1*, 59–73. [CrossRef]
40. Loague, K.; Green, R.E. Statistical and graphical methods for evaluating solute transport models: Overview and application. *J. Contam. Hydrol.* **1991**, *7*, 51–73. [CrossRef]
41. Garofalo, P.; Di Paolo, E.; Rinaldi, M. Durum wheat (Triticum durum Desf.) in rotation with faba bean (*Vicia faba* var. minor L.): Long-term simulation case study. *Crop Pasture Sci.* **2009**, *60*, 240–250. [CrossRef]
42. Rinaldi, M.; Garofalo, P.; Rubino, P.; Steduto, P. Processing tomatoes under different irrigation regimes in Southern Italy: Agronomic and economic assessments in a simulation case study. *J. Agrometeorol.* **2011**, *3*, 39–56.
43. Archontoulis, S.V.; Miguez, F.E.; Moore, K.J. A methodology and an optimization tool to calibrate phenology of short-day species included in the APSIM PLANT model: Application to soybean. *Environ. Modell. Softw.* **2014**, *62*, 465–477. [CrossRef]
44. Robertson, M.J.; Carberry, P.S.; Huth, N.I.; Turpin, J.E.; Probert, M.E.; Poulton, P.L.; Bell, M.; Wright, G.C.; Yeates, S.J.; Brinsmead, R.B. Simulation of growth and development of diverse legume species in APSIM. *Aust. J. Agric. Res.* **2002**, *53*, 429–446. [CrossRef]
45. Habash, D.Z.; Kehel, Z.; Nachit, M. Genomic approaches for designing durum wheat ready for climate change with a focus on drought. *J. Exp. Bot.* **2009**, *60*, 2805–2815. [CrossRef]
46. Latiri, K.; Lhomme, J.P.; Annabi, M.; Setter, T.L. Wheat production in Tunisia: Progress, inter-annual variability and relation to rainfall. *Eur. J. Agron.* **2010**, *33*, 33–42. [CrossRef]
47. Kyratzis, A.C.; Pallides, A.; Katsiotis, A. Investigating stability parameters for agronomic and quality traits of durum wheat grown under mediterranean conditions. *Agronomy* **2022**, *12*, 1774. [CrossRef]

48. Zhang, Z.; Sun, Y.; Yu, D.; Mao, P.; Xu, L. Influence of sampling point discretization on the regional variability of soil organic carbon in the Red Soil region, China. *Sustainability* **2018**, *10*, 3603. [CrossRef]
49. Gabrielle, B.; Roche, R.; Angas, P.; Cantero-Martinez, C.; Cosentino, L.; Mantineo, M.; Langensiepen, M.; Hénault, C.; Laville, P.; Nicoullaud, B.; et al. A priori parameterisation of the CERES soil-crop models and tests against several European data sets. *Agronomie* **2002**, *22*, 119–132. [CrossRef]
50. Overman, A.R.; Blue, W.G. Estimating yields and forage N for bahiagrass production in Florida. *Soil Crop Sci. Soc. Fla. Proc.* **1990**, *49*, 113–117.
51. Overman, A.R.; Wilkinson, S.R.; Evers, G.W. Yield response of Bermudagrass and Bahiagrass to applied nitrogen and overseed clover. *Agron. J.* **1992**, *84*, 998–1001. [CrossRef]
52. Timsina, J.; Humphreys, E. Performance of CERES-Rice and CERES-Wheat models in rice-wheat systems: A review. *Agric. Syst.* **2006**, *90*, 5–31. [CrossRef]
53. Langensiepen, M.; Hanus, H.; Schoop, P.; Gräsle, W. Validating CERES-wheat under North-German environmental conditions. *Agric. Syst.* **2008**, *97*, 34–37. [CrossRef]
54. Dettori, M.; Cesaraccio, C.; Motroni, A.; Spano, D.; Duce, P. Using CERES-Wheat to simulate durum wheat production and phenology in Southern Sardinia, Italy. *Field Crop. Res.* **2011**, *120*, 179–188. [CrossRef]
55. Monteleone, M.; Cammerino, A.R.B.; Garofalo, P.; Delivand, M.K. Straw-to-soil or straw-to-energy? An optimal trade off in a long term sustainability perspective. *Appl. Energy* **2015**, *154*, 891–899. [CrossRef]