land

MDPI

A Method to Determine the Optimal Period for Field-Scale Yield Prediction Using Sentinel-2 Vegetation Indices

Roberto Colonna [1,2,*], Nicola Genzano [3,*], Emanuele Ciancia [2,4], Carolina Filizzola [2,4], Costanza Fiorentino [5], Paola D'Antonio [5] and Valerio Tramutoli [1,2]

1   School of Engineering, University of Basilicata, 85100 Potenza, Italy; valerio.tramutoli@unibas.it
2   Satellite Application Centre (SAC), Space Technologies and Applications Centre (STAC), 85100 Potenza, Italy; emanuele.ciancia@cnr.it (E.C.); carolina.filizzola@cnr.it (C.F.)
3   Department ABC (Architecture, Built Environment and Construction Engineering), Politecnico di Milano, Via Ponzio 31, 20133 Milano, Italy
4   Institute of Methodologies for Environmental Analysis-National Research Council (CNR-IMAA), C.da Santa Loja, Tito Scalo, 85050 Potenza, Italy
5   School of Agricultural, Forest, Food, and Environmental Sciences (SAFE), University of Basilicata, Via dell'Ateneo Lucano 10, 85100 Potenza, Italy; costanza.fiorentino@unibas.it (C.F.); paola.dantonio@unibas.it (P.D.)
*   Correspondence: roberto.colonna@unibas.it (R.C.); nicola.genzano@polimi.it (N.G.)

Abstract: This study proposes a method for determining the optimal period for crop yield prediction using Sentinel-2 Vegetation Index (VI) measurements. The method operates at the single-field scale to minimize the influence of external factors, such as soil type, topography, microclimate variations, and agricultural practices, which can significantly affect yield predictions. By analyzing historical VI data, the method identifies the best time window for yield prediction for specific crops and fields. It allows adjustments for different space–time intervals, crop types, cloud probability thresholds, and variable time composites. As a practical example, this method is applied to a wheat field in the Po River Valley, Italy, using NDVI data to illustrate how the approach can be implemented. Although applied in this specific context, the method is exportable and can be adapted to various agricultural settings. A key feature of the approach is its ability to classify variable-length periods, leveraging historical Sentinel-2 VI compositions to identify the optimal window for yield prediction. If applied in regions with frequent cloud cover, the method can also identify the most effective cloud probability threshold for improving prediction accuracy. This approach provides a tool for enhancing yield forecasting over fragmented agricultural landscapes.

Keywords: remote sensing agriculture; crop monitoring techniques; field-level forecasting; phenological analysis; high-resolution vegetation data; S2 imagery applications; ideal timing acquisition; NDVI; clear pixel procedure; agricultural productivity

1. Introduction

The history of agricultural yield prediction is very long [1], and the research area is very wide [2]. The various methods can be divided into two main categories: physical models (also known as process-based or mechanistic) and data-driven (statistical and machine learning) models [3,4]. Process-based models use physical laws that simulate the growth of crops and can be used to predict yield under a variety of conditions. Data-driven models use statistical methods to learn from historical data [3]. Additionally, some models, such as deterministic and stochastic models, can fall into either category [1]. Deterministic models provide a single, specific output for a given set of inputs, assuming no randomness in the process, and are often based on precise relationships derived from physical laws [5]. Stochastic models, on the other hand, incorporate randomness or uncertainty, providing a range of possible outcomes based on probabilistic inputs, which can better reflect the

inherent variability in agricultural systems, such as weather or biological processes [6]. Data-driven models (especially the statistical ones) are the most used [7] since they can be developed more easily and, if well-articulated, can guarantee reliable results.

In recent years, remote sensing has become essential for providing timely, high-quality information on crop growth [8], while machine learning (ML) techniques have gained prominence in this field [9,10]. Convolutional Neural Networks (CNNs) have been particularly effective in extracting features from satellite imagery time-series [11], often combined with Long Short-Term Memory (LSTM) networks to capture spatial and temporal growth patterns [12]. Among the most used methods for yield prediction are those based on decision trees [2]. Ensemble methods, which combine multiple models, either process-based or data-driven, have also demonstrated good performance [13–17].

1.1. Challenges and Limitations in Yield Prediction Models

The primary difficulties encountered by statistical and ML models arise from the complex interplay of factors influencing crop yield. These factors can be broadly categorized into three main groups: (1) genetic factors, which include crop variety, seed quality, and plant resistance traits; (2) environmental factors, such as temperature, precipitation, and biotic stressors (e.g., pests and diseases); and (3) management practices, including planting date, irrigation, fertilization, pest control strategies, and soil health management (e.g., nutrient management, organic matter content). Each of these categories plays a crucial role in determining yield outcomes, and their interactions add further complexity to the prediction models [18]. The use of remote sensing in data-driven models introduces additional variables, such as the choice of satellite sensors and the spatial, temporal, and spectral resolution. Other factors include the vegetation indices used and the quality and timing of measurements. Due to this wide range of variables involved, yield forecasting models, while effective in specific regions, often underperform when applied to different areas [19]. Moreover, many of the crop yield models are designed for application at the national or regional level [2,7,20], as coarse-scale yield data are more readily available, while field-scale ground-truth estimates are much harder to obtain.

However, the recent scientific literature highlights that these models, while effective on a broader scale, often underperform when applied at the field scale [21–23]. This is because, when transitioning from coarse to fine scales, new challenges emerge. First, parameters validated on a coarse scale are often not adaptable to finer scales [24]. Most importantly, among the numerous factors influencing crop yield, some are specific to individual fields. Elements such as soil type, topography, microclimate variations, and agricultural practices significantly affect yield predictions [25,26], especially in regions with a highly fragmented agricultural landscape, such as Italy [27,28]. Depending on these factors, the forecasting performance of the field-scale models can vary significantly.

1.2. Performance of Field-Scale Models and Role of Sentinel-2

In 2023, Leukel et al. [29] conducted a systematic review of field-scale yield predictions, analyzing 23 models from various studies. Model performance, indicated by R^2 values, ranged from 0.23 to 0.92 and was influenced by factors such as crop type, location, lead time to harvest, the type of tool (e.g., satellite or UAV), the VI used, and the method applied. The highest-performing models were those utilizing UAVs or high spatio-temporal resolution satellites on demand [30–34]. Among models using free satellite data, the best was Sharifi's Gaussian process regression model (integrating Sentinel-2 and Landsat data) [35], which was based on 5 years of field-scale yield data from Iran, which achieved R^2 values between 0.69 and 0.84 depending on temporal training settings and agricultural field differences. The review's findings indicated that high resolution and the development of site-specific models can significantly improve field-scale yield forecasts.

To this aim, the Sentinel-2 (S2) satellite constellation offers significant advantages [36,37], providing free-of-charge imagery with improved features [38]. This accessibility benefits users working with low-income crops (like wheat) [39,40]. Previously, Landsat was the

only free option for field scale modeling in Europe, offering two scenes per month at a 30 m resolution. Since June 2017, S2 has provided six images per month with a 5-day revisit time and 10 m resolution [41,42], allowing for more detailed VI assessments and reducing the impact of persistent cloud cover [41]. Although S2 has enhanced yield estimation compared to Landsat [43], its long-term data availability remains limited, and crop rotation further restricts consistent data collection [44]. Combining S2 with Landsat can improve data availability but may introduce inhomogeneity and uncertainty, especially in the near-infrared band [45]. Nevertheless, S2 has been successfully used in yield estimation across various crops in several studies [36,46–49].

1.3. Importance of Timing in Satellite-Based Yield Forecasting

In EO data-driven models, the model's greatest flexibility lies in selecting optimal images based on the time-series of its training data. The timing of satellite measurements is one of the few variables where the model can make adjustments, making it a critical factor for yield forecasting. Leukel et al.'s review [29] also focused on the timing of satellite measurements across various field-scale models, specifically referring to the "prediction horizon", which is the time between the forecast and harvest. The prediction horizons proposed by the studies varied significantly, even for the same crop (wheat or corn) and geographical region (e.g., USA, China, Australia), ranging from weeks to months before harvest. Some studies included in the review, such as Fieuzal et al.'s (2020) [50], found no major differences in predictive accuracy across different crop stages, suggesting limited benefits from using images acquired after mid-season. Despite this, research often focuses on forecasting near harvest time, which is a trend also pointed out in the review by Schauberger et al. (2020) [7]. Li et al. (2021) [51] found that for winter wheat in China, the most predictive phase was between emergence and tillering, while for rice, it was between tillering and booting. Overall, these reviews and case studies suggest that identifying the optimal period for yield prediction depends on factors like crop type, methods, and site-specific conditions.

In ML models, the added value of identifying the most representative period for yield forecasting lies in the ability to deliver predictions earlier. This can provide a significant advantage in terms of timely decision-making and resource allocation. In statistical models and those integrated with process-based inputs, an additional benefit arises: the ability to select the most representative satellite images based on the identified period. This improves the accuracy of the models by ensuring that the data used are closely aligned with the critical phases of crop development.

1.4. Aim of the Study

The aim of this contribution is to propose an exportable method to determine the best time period/phenological phase for field-scale yield forecasting by S2. The method can be applied to any VI and crop type. In this case, we employed the NDVI [52–54], which is the most frequent index used for this purpose [55]. A wheat field in the Po River Valley (Northern Italy) is used as an example to show how the approach works. Fifteen variable-sized Max Value Composite (MVC) [56] periods, ranging from January to May, were composed and ranked according to their ability to explain the final yield. As a preliminary step, since the study area was located in a particularly humid region, an NDVI-based clear pixel procedure (NDVI-CPP) was also developed. This procedure leverages the difference in NDVI between cloud cover and vegetation to optimize cloud detection in Sentinel-2 images by fine-tuning the cloud probability (CP) mask threshold.

2. Materials and Methods

2.1. Study Area, Data Collection and Preprocessing

The following analysis was carried out in the Google Earth Engine environment using Sentinel-2 satellite data, in particular the collection "Harmonized Sentinel-2 MultiSpectral Instrument, Level-1C" (S2) and the "Sentinel-2 Cloud Probability" (CP) collection. The

term "Harmonized" indicates that the satellite scenes prior to the "04.00 update", which occurred on 25 January 2022, have been aligned with subsequent ones to ensure radiometric measurements' homogeneity. The field object of this study is located in the municipality of Conselice (RA), in Emilia-Romagna (Italy). The field has an extension of 11.144 hectares (about 1114 Sentinel-2 pixels). The centroid coordinates of the field are Lon.: 11.788; Lat.: 44.536, and it is located in a flat area in the Po River valley at an altitude of 5 m above sea level. The field is represented in Figure 1.
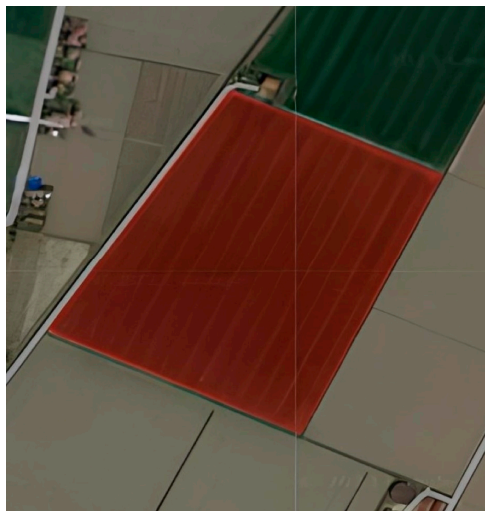

Figure 1. Investigated field (red polygon).

The crop type examined was soft winter wheat planted in the field during the 2017/2018, 2019/2020, and 2021/2022 seasons. The sowing dates were 31 October 2017; 4 November 2019; and 2 November 2021. The field was rain-fed. In all growing seasons, the same treatments aimed at controlling weeds were applied: one before sowing and another in April. The absence of weeds and the condition of the crops were monitored through routine field inspections in November, January, and April. The annual field-level yields, as recorded by the farm during the post-harvest weighing process, are reported in Table 1.

Table 1. Annual wheat yields of the investigated field (q = metric quintal; ha = hectare).

Year	Yield (q/ha)
2018	57
2020	79
2022	65

Through a Google Earth Engine (GEE) script, the collection of satellite images was first pre-processed to filter the spatial, temporal, and spectral intervals of interest and eliminate images of inadequate quality and duplicates (sometimes present within the collections). Spatially, the field under investigation falls within the Military Grid Reference System (MGRS) tile 32TQQ. Temporally, a filter was applied to include the scenes falling in the November/June period of the 2017/2018, 2019/2020, and 2021/2022 seasons. Furthermore, a filter was applied to include the bands necessary for calculating the NDVI index (B4 and B8). Finally, by exploiting the collection image properties "GENERAL_QUALITY" and "GEOMETRIC_QUALITY", the tiles that did not pass the Sentinel-2 On-Line Quality Control (OLQC) [42] were eliminated.

2.2. Cloud Detection and Atmospheric Conditions

The issue of cloud identification in satellite observations has been tackled using various cloud masking techniques for optical sensors over the years (e.g., [57–62]), with

recent improvements driven by ML advancements [62–65]. However, all cloud masks have technical limitations, leading to omission and commission errors [66]. Additionally, optical band imagery (used for vegetation monitoring) can be contaminated not only by the presence of tropospheric clouds but also by the presence of cirrus clouds and mist. Mist, in particular, can alter the vegetation indices due to its variable impact on red and NIR bands based on vapor characteristics (droplet size and density) [67]. Then, the ideal cloud mask depends on the use case (the bands you want to mask and the content of the ground cells you want to observe). Thinking from this perspective, in 2017, the s2cloudless algorithm was developed [68]. Since 2020, this has been integrated into the Google Earth Engine (GEE) as Sentinel-2: Cloud Probability [69,70], allowing users to set a threshold to convert the cloud probability into a mask. The authors recommend applying a CP threshold of 0.4 (40%) [66], but customization is suggested for optimal results based on the intended application.

To manage cloud cover, we propose an alternative solution practiced by several authors (e.g., [71–73]), which is the Maximum Value Composite (MVC) procedure [74]. In a predetermined time period (e.g., 1 month), MVC selects for each pixel the greatest VI value. Considering that vegetation indexes (like NDVI) assume values around zero in the presence of clouds, such a process is expected to automatically filter out cloud recurrences in the time interval for which the computation is applied. However, in humid regions, such as the Po River Valley, this approach is also not risk-free, as the agricultural field, or part of it, may be cloudy and/or foggy during the entire predetermined time interval. To address this issue, we implemented a specific cloud detection method, the NDVI-based clear pixel procedure (NDVI-CPP), in this study.

2.3. NDVI-Based Clear Pixel Procedure (NDVI-CPP)

Once the S2 collection was properly preprocessed, a cloud probability (CP) function was created with the CP threshold as the variable of the cloud mask. The CP threshold was made to vary from 1 to 100. Within the CP function, firstly, the cloud mask using the CP collection was applied to the entire S2 collection. Then, the function for calculating the NDVI was programmed and mapped on the entire S2 collection. At this point, a new collection of NDVI scenes with a 5-day revisit frequency and 10 meters' spatial resolution was obtained. Figure 2 shows an example of an NDVI index map created in false colors (red, yellow, and green).
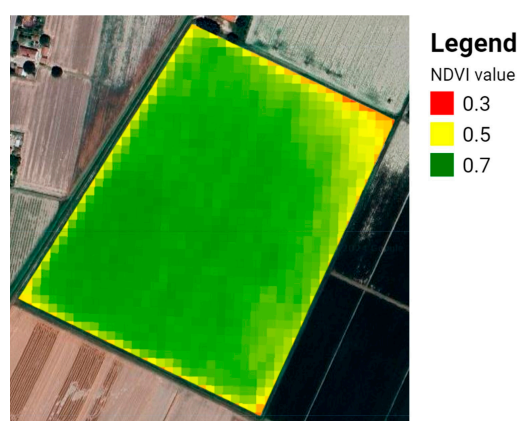

Figure 2. Example of NDVI map made in the Google Earth Engine environment (scene of 29 March 2020).

Thus, a series of MVC scenes on a monthly basis was composed. Following this, for each monthly MVC (mMVC), the spatial average of the entire field (FmMVC) was calculated. The aim of the FmMVC computation was to use it as an optimization criterion for the cloud probability threshold setting. When using the Sentinel-2 Cloud Mask, setting a CP threshold

that is too high leads to a cloud omission error (cloudy pixels not masked), and setting a CP threshold that is too low leads to a cloud commission error (clear pixels masked).

The application of the monthly MVC procedure protects from cloud commission errors (no pixels are masked), but cloud omission errors are still possible (in the case of persistent adverse weather conditions for the entire monthly sequence). However, knowing that the NDVI index takes values close to 0 in the presence of clouds and higher values in the presence of vegetation, we assume that, for each monthly MVC collection (or NDVI collection), as CP threshold increases, starting from 0 (all pixels masked), the FmMVC (or F-NDVI, i.e., the field NDVI spatial mean) increases until a maximum is reached, which represents the CP value for which all clear pixels are shown. A possible subsequent decrease in FmMVC (F-NDVI) indicates the presence of (unmasked) pixels compromised by atmospheric conditions. Then, the best balance between omission and commission errors (optimal CP threshold) can be found by setting the value corresponding to the maximum FmMVC as the CP threshold (see Figure 3).
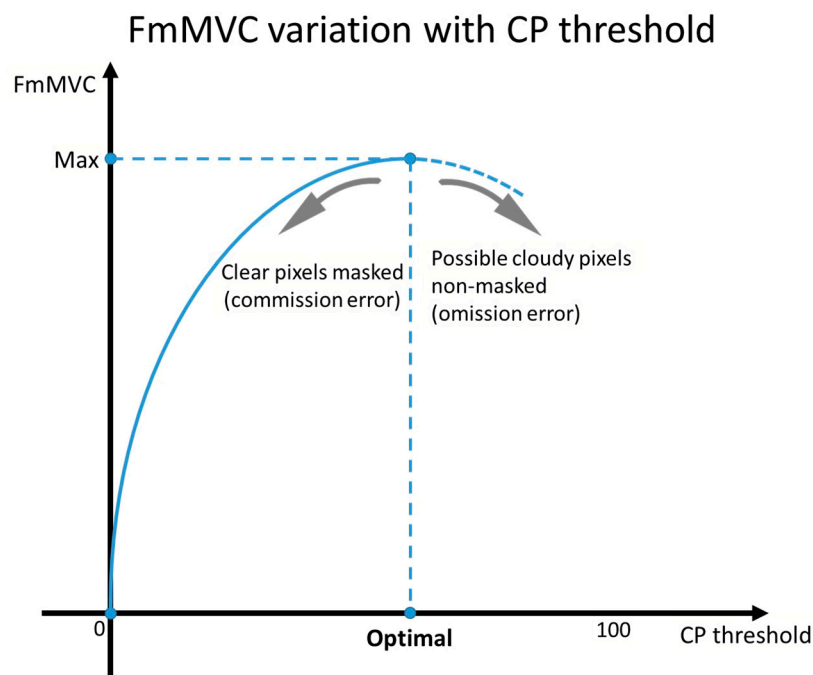

Figure 3. The theoretical FmMVC variation as the CP threshold varies and the optimal setting of the CP threshold as a function of FmMVC (by taking the maximum value).

To illustrate this point with an example, consider a cloudy monthly dataset under investigation and examine how the Maximum Value Composite (MVC) changes as the cloud probability (CP) threshold is adjusted. For this analysis, the month of March 2018 is selected, and for simplicity, the CP threshold is varied in increments of 10 (see Table 2).

Table 2. FmMVC variation according to the CP threshold variation in March 2018.

FmMVC as CP Threshold Varies										
CP threshold (%)	10%	20%	30%	40%	50%	**60%**	70%	80%	90%	100%
FmMVC 2018-03	No Data (entire field masked)	0.2965	0.3262	0.3441	0.4064	0.4725	**0.4725**	0.4700	0.4515	0.4275	0.3368

In the proposed example, the optimal CP threshold value is 60%. Lower values reveal the masking of clear pixels within the MVC, while higher values correspond to the inclusion of pixels compromised by atmospheric conditions. Thus, to extend the procedure to all the periods under investigation and determine a valid CP threshold for the entire three-season

MVC collection, we calculated the temporal mean FmMVC value by averaging the already calculated FmMVC values temporally over the three seasons under investigation. The months from January to May were considered months for which we were certain that there was sufficient plant cover. We call this synthetic value the t-mean FmMVC (temporal-mean FmMVC).

Then, we set the optimal CP threshold by taking the value for which the t-mean FmMVC was the maximum as the CP threshold varied from 1 to 100. Finally, the optimal CP threshold was applied to the S2 cloud mask collection, and the corresponding S2 image collection was composed. The implemented procedure is schematized in Figure 4.
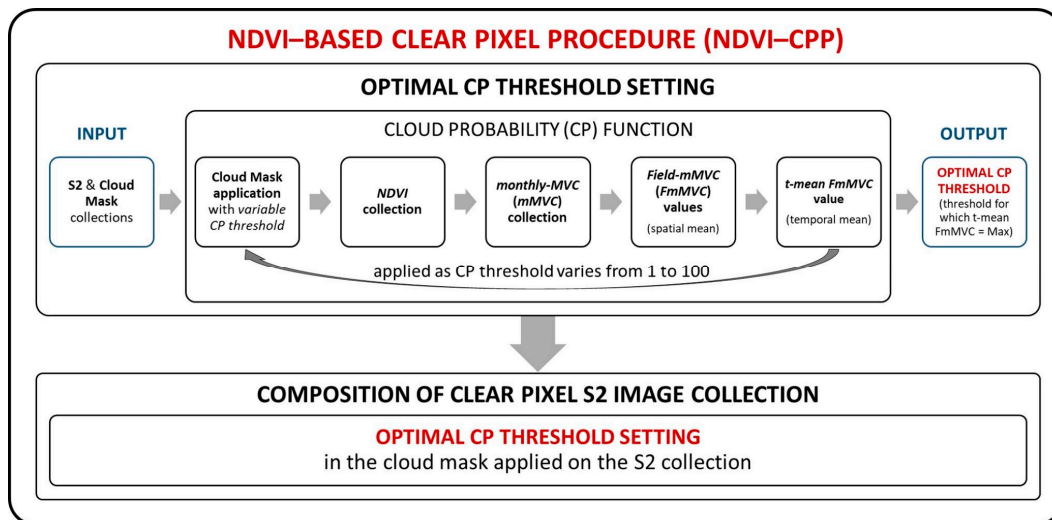

Figure 4. Implementation scheme of the NDVI-CPP.

The NDVI-CPP is designed to operate in particularly humid areas that typically experience fog or persistent cloud cover. Where these conditions are absent, the use of the iterative NDVI-CPP becomes unnecessary. In such cases, the user can select a cloud probability (CP) threshold at their discretion and proceed to the subsequent phase of determining the optimal period for yield prediction.

2.4. Optimal Yield Forecast Period Determination

The "clear pixel S2 image collection" was then used as an input for the determination of the optimal NDVI reference period for the yield forecast. To find the MVC temporal combination that best correlated with the field yield, firstly, the corresponding NDVI collection and clear pixel monthly MVC mosaic collection were generated. Then, for each clear pixel mMVC mosaic from January to May (2018, 2020, and 2022), we calculated the spatial mean, obtaining the relative FmMVC values. After that, we calculated the 3 mean FmMVC values[1] (1 per year) in each of the 15 variable-sized periods (for a total of 45 FmMVC values), as shown in Figure 5.
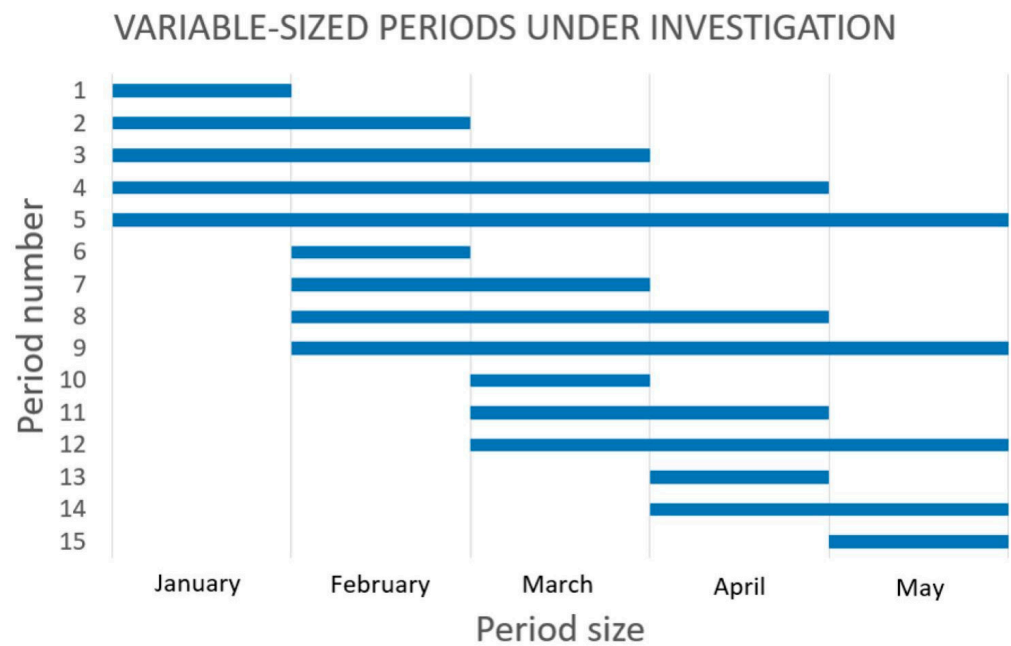
Figure 5. Start–end months of the 15 variable-sized periods to test for the determination of the optimal reference period for yield forecast.

Then, to obtain a synthetic value capable of measuring the performance of each time interval, we cyclically calculated and classified the 15 Pearson correlation coefficients (R) between the three seasonal yields and the three mean FmMVC values for each variable-sized period. The steps of the method are shown in Figure 6.

Finally, in order to obtain indicative time references for the crop phenological phases, we downloaded the daily temperature data from the meteorological station of Conselice (Ravenna, Italy) [75] starting from the season 2008/2009. After the download, the data were averaged by month in order to compare the three seasons under investigation with the monthly average of the 10 previous seasons.
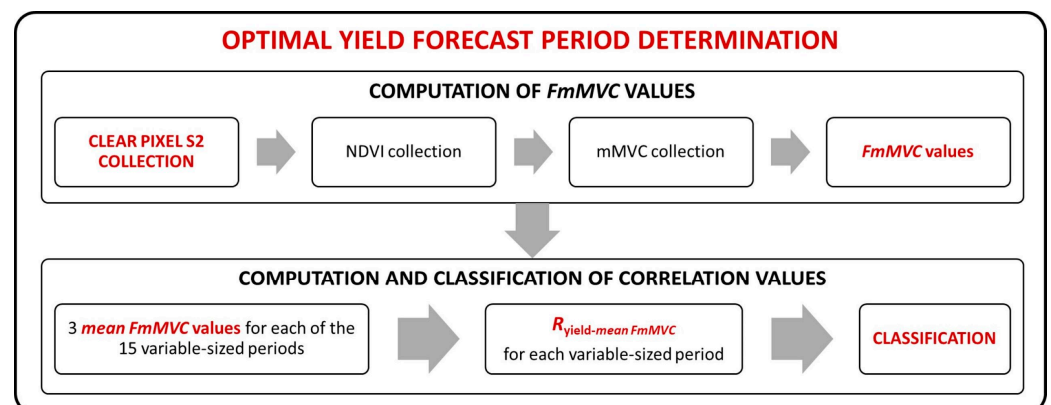

Figure 6. Diagram of the method implemented for determining the optimal yield forecast period.

3. Results

3.1. Optimal Cloud Probability Threshold

The output of the NDVI-CPP was the optimal CP threshold value to use in the "Sentinel-2: Cloud Probability" mask (representing the best balance between omission and commission errors), i.e., the CP threshold value for which the t-mean FmMVC value was the maximum. As shown in the graphs in Figure 7, after the execution of the procedure, the optimal threshold was found to be equal to 0.66 (66%).

Then, by applying the 66% CP threshold value to the investigated S2 time-series, were computed all the clear pixel mMVC mosaics. Note that the t-mean FmMVC values reported on the y-axis in Figure 7 are not a reliable reference for evaluating the effectiveness of the CP threshold level, as they represent data averaged over three years, which include mostly clear-sky images that tend to flatten the values. The effectiveness of the CP threshold can only be assessed in cloudy scenes. As an example, in Figure 8, we show the difference in applying three different CP thresholds in March 2018 (which was a particularly cloudy month). The scenes displayed are the monthly RGB composites generated using the B4 (red), B3 (green), and B2 (blue) band values of the same pixels employed for the mMVC mosaics (i.e., the greenest ones).
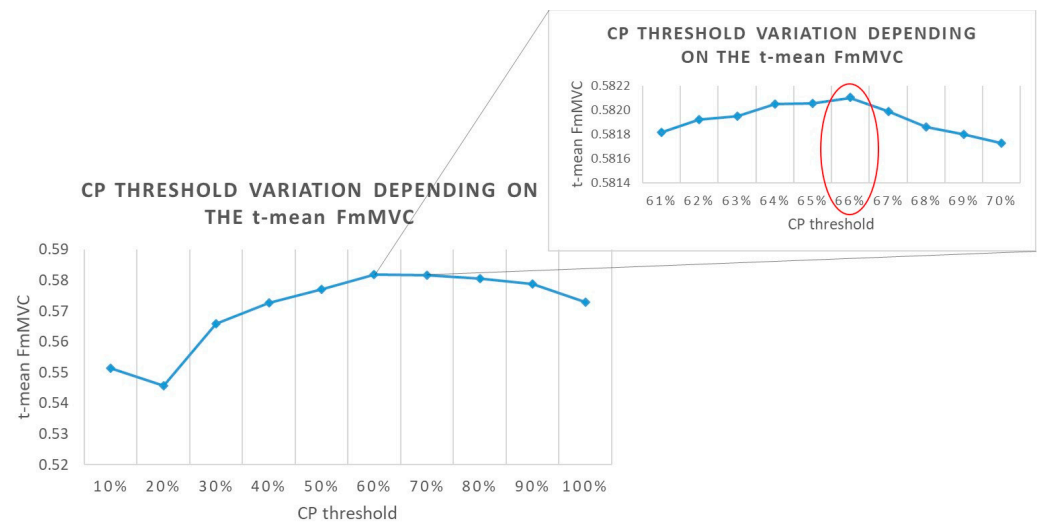

Figure 7. "Sentinel-2 Cloud Probability" (CP) threshold variation as a function of the t-mean FmMVC measured over the investigated field. On the left, the variation at 10% intervals and, on the right, the zoom for the variation of a single percentile between 61% and 70% are given, interval in which the maximum value (in red circle) is recorded.
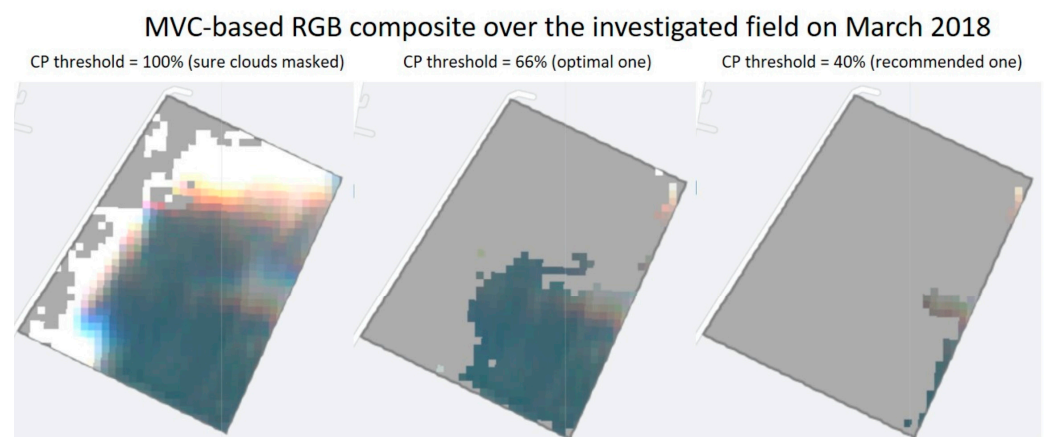

Figure 8. March 2018 RGB composites, over the field under investigation, at 3 different CP threshold percentages: 100% (on the **left**), 66% (in the **center**), and 40% (on the **right**). The scenes are generated using the R = B4, G = B3, and B = B2 band values of the same pixels employed for the MVC (i.e., the pixel having the maximum NDVI).

The 100% CP represents the case in which only clouds identified for sure are masked; the 66% CP is the cloud probability optimized for the purpose as previously discussed; and CP 40% is the threshold value recommended by the s2cloudless authors for general applications [68]. The unmasked clouds on the left of the figure are present in all six scenes

of the month, so the MVC was unable to mask them. Instead, from the figure on the right, it can be seen that lowering the threshold to 40% meant that non-cloudy pixels were masked as a result. When the optimal threshold of 66% is applied, the FmMVC value is 0.4757, while the thresholds of 100% and 40% correspond to FmMVC values of 0.3369 and 0.3441, respectively. These values would have also significantly distorted the mean FmMVC–yield correlations, which are shown in Figure 9 in the next section, leading to a decrease in the correlations ranging from 1% to 5% across all variable-size periods that include the month of March. In particular, the FmMVC–yield correlation for the month of March alone would have dropped from 95.2% to 90.1% (with a CP threshold of 100%) to 90.3% (with a CP threshold of 40%).
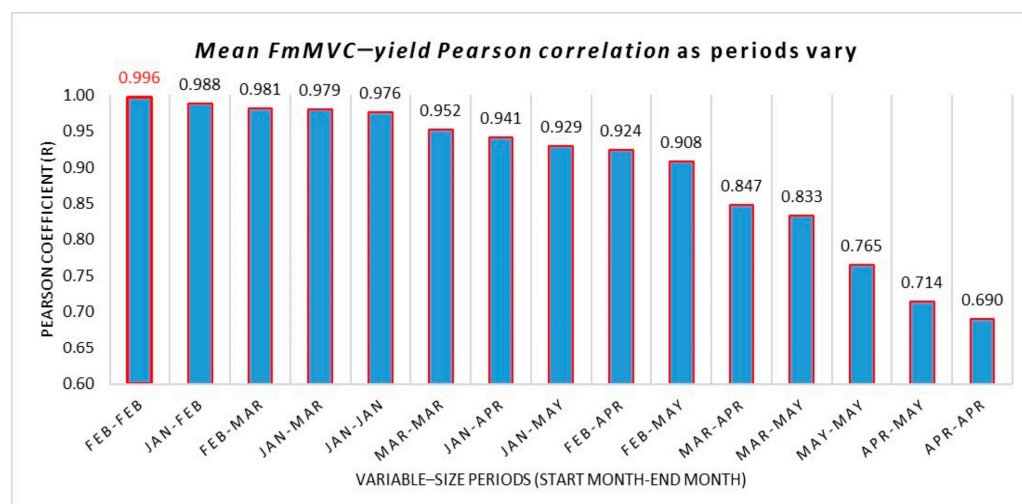

Figure 9. Pearson correlation coefficient mean for FmMVC–yield correlation for each of the 15 variable-sized periods sorted in descending order.

3.2. Optimal NDVI Reference Period

Figure 9 shows the Pearson correlation coefficients between the mean FmMVC and yield for each of the 15 variable-sized periods investigated. The MVC of the period is calculated using the 66% CP threshold. The correlation coefficients are sorted in descending order by value.

The best correlated month is February, with a correlation index close to unity. However, we also generally notice a better correlation during the first 3 months of the year compared to the April–May period. The correlation coefficients are, in fact, always greater than 95% when the period includes only the months from January to March and decreases sharply below 80% when the period includes only the months of April and May. The high correlation coefficients for FmMVC–yield also highlight the validity of the NDVI-CPP in optimizing cloud detection.

Focusing on the winter wheat phenological phases, after sowing, the first phase encompasses germination, emergence, and tillering. This is the slowest and longest phase, depending on the weather conditions; it may either complete or be interrupted before the winter dormancy. Most of the processes that contribute to grain yield are completed during this phase [76–78]. Stem elongation (or extension), a period of rapid plant growth, typically occurs in mid-March, but its timing and intensity are highly dependent on the air temperature [79]. In Italy, winter wheat generally reaches maturity in early June, with harvesting usually completed by late June or early July.

Growth stages are closely linked to weather conditions, mainly air temperature. Some indicative temperature thresholds for soft wheat phases are as follows: the minimum germination temperature is 1 °C, with a maximum of 37 °C and an optimal range of 20–25 °C. For stem elongation, the thermal threshold is 5–10 °C, with an optimal range of 15–22 °C, while the ideal flowering temperature is 18–24 °C [80]. Figure 10 shows the

monthly average temperatures recorded at the Conselice weather station, comparing the three seasons under investigation with the average of the previous 10 seasons.

The average monthly temperatures for the three seasons under investigation were generally higher than the standard values, particularly in February 2020 and 2022, when temperatures exceeded 10 °C. Germination likely began shortly after sowing, supported by the November temperatures (averaging around 10 °C in all seasons). Tillering progressed slowly in the colder months, with possible dormancy setting in December and lasting until February. The increase in temperatures anticipated the occurrence of stem elongation in at least two of the three seasons investigated in February rather than in March. Flowering would have occurred in April across all seasons, with temperatures around 17–19 °C, which are slightly below the optimal range but still sufficient. By June, maturity was reached in all three seasons, supported by temperatures well above 24 °C.
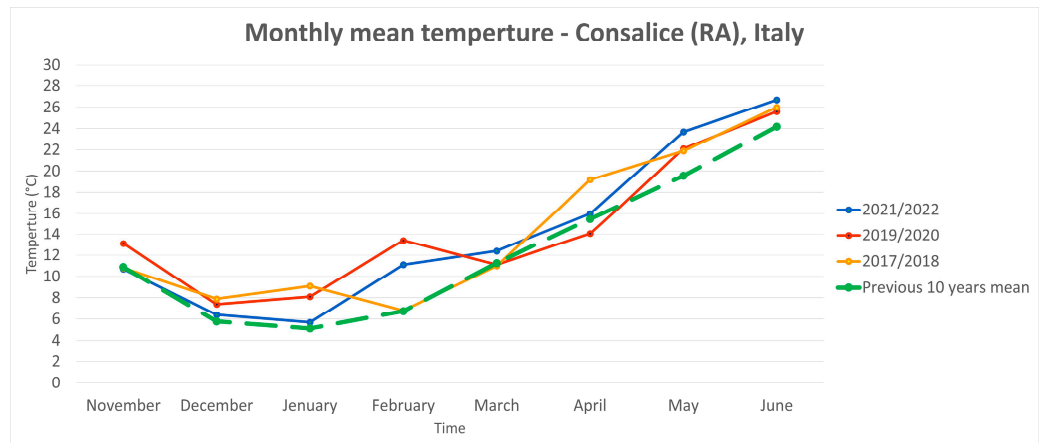

Figure 10. Plot of the three seasons under investigation and the average of the 10 previous seasons (starting from 2008/2009) with the monthly mean temperature of the meteorological station of Conselice (Ravenna, Italy).

Therefore, in terms of phenological phases, it seems that the following is true:

(a) The best yield–NDVI correlation can be associated with the stem elongation phase;
(b) The timing of this phase cannot be predicted based solely on long-term analysis (e.g., 10 years or more). Instead, field-scale analyses over shorter time periods are required to account for any more rapid interannual variations in air temperature.

To further delve into the whole dynamic, we plot (in Figure 11) the FmMVC by season and how it performed after the entire input collection was cloud-masked with a 66% CP threshold.
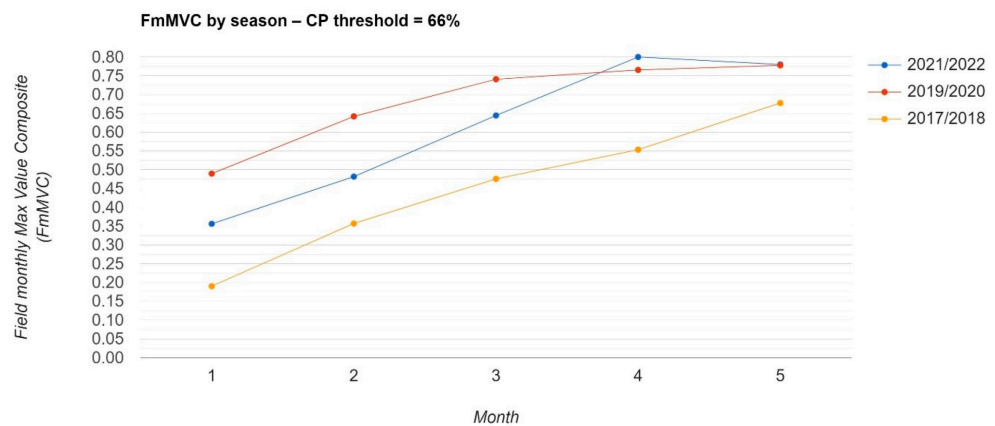

Figure 11. FmMVC by season calculated over the wheat field under investigation using the 66% CP threshold value.

While the NDVI-based index for the 2017/2018 season exhibits a consistently linear growth pattern, the 2019/2020 season shows this linear trend only until March, and the 2021/2022 season maintains it until April. This behavior can be partially linked to the well-known NDVI saturation effect [81], which seems to occur around values of 0.75/0.8 depending on the field's vegetation cover. Conversely, the slight decrease observed in the index from April to May 2022 suggests that part of the non-linear behavior is also due to the natural decline in vegetation vigor during that period.

In Figure 12, we show the MVC maps corresponding to the best-correlated month (i.e., February) for each year under investigation. Under the hypothesis that the MVC value for this month serves as a proxy for the expected yield, these maps provide spatial information on areas likely to be more productive, as well as on the ones that may require additional management practices.
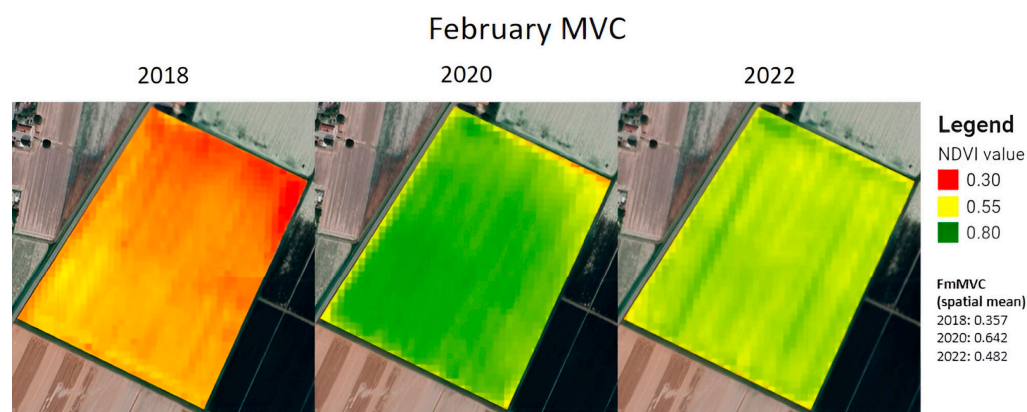

Figure 12. February (NDVI-based) Maximum Value Composite for the years 2018, 2020, and 2022.

The maps are also consistent with the field yields recorded in the 2018 (57 q/ha), 2020 (79 q/ha), and 2022 (65 q/ha) seasons.

4. Discussion

One limitation of this study is related to the limited dataset. On one hand, field-scale yield data (especially in Europe, where such information is often controlled in the framework of the EC aids in agriculture) is difficult to obtain from farmers. On the other hand, the analysis was limited to only 3 years of agricultural harvests. Considering that seven agricultural seasons have passed since the two twin Sentinel-2 satellites became fully operational in the second half of 2017 [42], and given that crop rotation is a mandatory (and necessary) practice in Italy [82], this limitation remains intrinsic to the use of Sentinel-2 data until the continuation of the mission increases the available dataset. Other limitations of this study include the need to run the method multiple times when using different indices or fields, treating them as separate variables. Additionally, the method does not directly account for temperature data, which, in cases of significant annual variation, can be important in determining the optimal period and should be addressed separately.

Although many scientific reviews have explored the identification of the most representative forecasting period (e.g., [7,29], our literature research does not reveal any model specifically designed to determine the optimal measurement period tailored to individual fields. In machine learning models, this is understandable, as it can be addressed for each model by adjusting input periods [50], which can help determine the most advantageous time to deliver forecasts to the end user [7]. However, the agricultural sector is constantly evolving, with frequent changes in field-level conditions (e.g., changes in practices, planting dates, etc.). Consequently, homogeneous data are often insufficient to train ML models [83]. In such cases, forecasts may also rely on pre-existing statistical models [7,84,85], potentially integrated with process-based models [36], and anchored by fixed parameters for key

inputs, such as the optimal time interval for image acquisition based on the crop growth phase [29].

5. Conclusions

In this study, we proposed a method to determine the optimal field-scale yield prediction period by Sentinel-2 VI measurements. To illustrate its application, this method was applied to a wheat field in the Po River Valley (Emilia Romagna, Italy) using the NDVI index. The method was specifically developed to operate at the single-field level in order to minimize the influence of external factors such as soil type, topography, microclimate variations, and agricultural practices, which can significantly affect yield predictions. These factors are particularly relevant in areas with highly fragmented agricultural landscapes, such as Italy.

As a preliminary step, Sentinel-2's 5-day temporal resolution was used to generate mosaics free from clouds or humidity. A self-developed "NDVI-based Clear Pixel Procedure" (NDVI-CPP) was applied to optimize cloud detection by adjusting the cloud probability (CP) threshold, which improved yield prediction accuracy compared to scenes without this adjustment. For the case study, the optimal CP threshold was set at 66%. However, this step can be omitted in areas with less persistent fog or cloud cover.

The core of the proposed method is its ability to classify variable-length periods using historical VI compositions from Sentinel-2, identifying the optimal time window for yield prediction for specific crops and fields. In the example provided, which focuses on a wheat field in Northern Italy, 15 NDVI periods of varying lengths were composed on a monthly basis, with February showing the strongest correlation with yield prediction. Upon analyzing air temperature data, it was hypothesized that this stronger correlation was due to stem elongation occurring earlier (in February rather than March) because of unusually high temperatures. More generally, the satellite-detected vegetation cover during the January–March period was more representative than that observed in April–May, even though the latter is closer to the harvest.

Beyond the results obtained on this specific application, which are not directly transferable due to the variability between fields, the broader objective of this work is to present an exportable method, particularly applicable in areas with highly fragmented landscapes that often lack sufficient homogeneous data to test and validate ML models. The method allows users to determine, for any specific application, the following:

1.　The optimal cloud probability threshold;
2.　The optimal period for yield prediction.

And the following can also be applied:

1.　Different space–time intervals;
2.　Other types of crops;
3.　Other variable-threshold cloud mask collections;
4.　A different MVC time (15 days, 3 weeks, 6 weeks and so on);
5.　No (or any other) time-based composite technique (like MVC).

Once the optimal CP threshold (which requires NDVI and ground vegetation cover to be applied) has been set, the optimal period for yield prediction can be determined (in the same space–time interval):

6.　For any Vegetation Index (VI).

The generalized process for "optimal yield forecast period determination using the NDVI-CPP" is outlined in Figure 13.
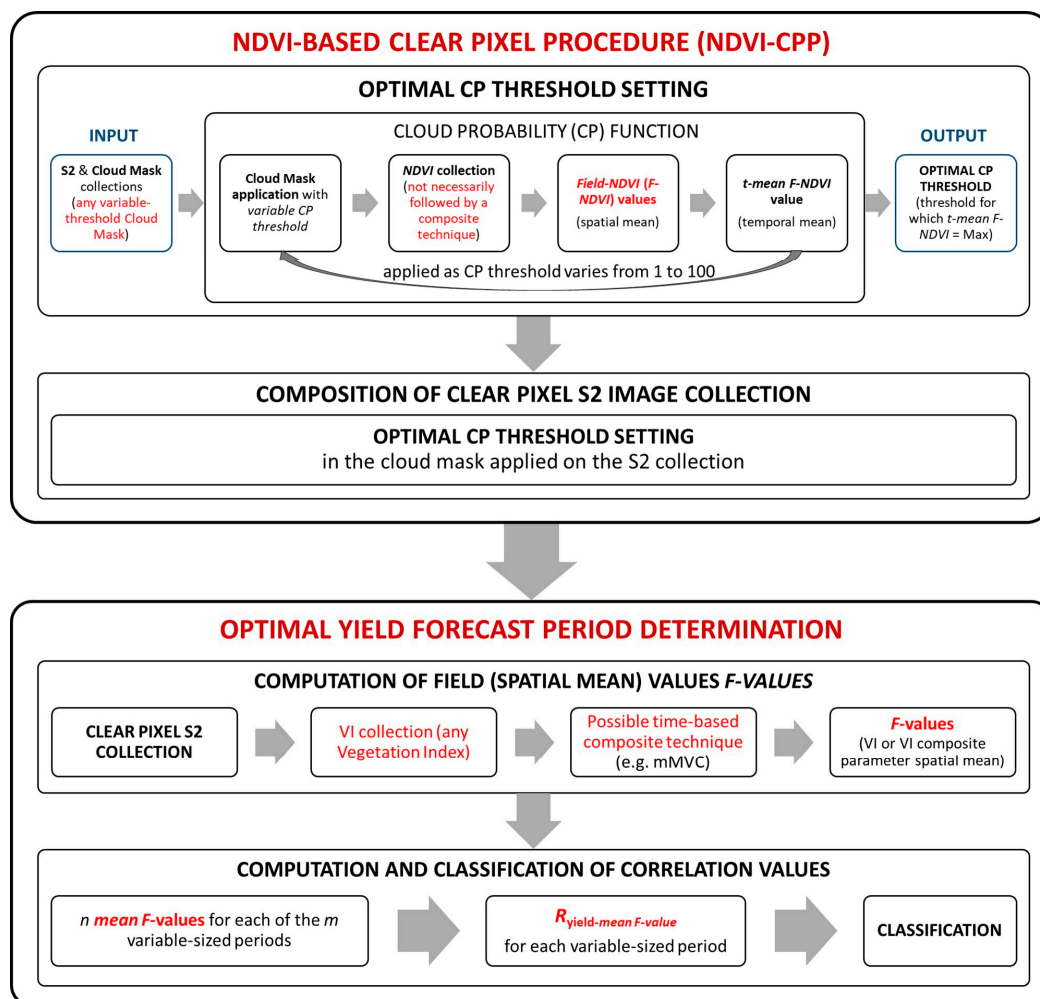
Figure 13. Flowchart of the (general) procedure for the "optimal yield forecast period determination using the NDVI-CPP". In red, the changes made (generalizations) compared to Figures 4 and 6 are shown. For the meaning of the acronyms, see Section 2. Materials and Methods.

Author Contributions: Conceptualization, R.C.; methodology, R.C. and V.T.; software, R.C.; validation, R.C., E.C., and N.G.; formal analysis, R.C. and C.F. (Carolina Filizzola); investigation, R.C.; resources, R.C., C.F. (Costanza Fiorentino) and V.T.; data curation, R.C.; writing—original draft preparation, R.C.; writing—review and editing, V.T., P.D., and C.F. (Costanza Fiorentino); visualization, R.C. and C.F. (Carolina Filizzola); supervision, V.T., P.D.; project administration, V.T. and N.G. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: This study is a re-analysis of existing data, which are openly available at locations cited in the reference section. Publicly available datasets were analyzed in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Note

1   Calculated as the mean of the individual FmMVC values; e.g.: $MEAN_{JAN-FEB} = (FmMVC_{JAN} + FmMVC_{FEB})/2$.

References

1.   Sargun, K.; Mohan, S. Modeling the crop growth-a review. *Mausam* **2020**, *71*, 103–114.
2.   Darra, N.; Anastasiou, E.; Kriezi, O.; Lazarou, E.; Kalivas, D.; Fountas, S. Can Yield Prediction Be Fully Digitilized? A Systematic Review. *Agronomy* **2023**, *13*, 2441. [CrossRef]

3. Chang, Y.; Latham, J.; Licht, M.; Wang, L. A data-driven crop model for maize yield prediction. *Commun. Biol.* **2023**, *6*, 439. [CrossRef] [PubMed]

4. Joshi, A.; Pradhan, B.; Gite, S.; Chakraborty, S. Remote-Sensing Data and Deep-Learning Techniques in Crop Mapping and Yield Prediction: A Systematic Review. *Remote Sens.* **2023**, *15*, 2014. [CrossRef]

5. Wallach, D.; Makowski, D.; Jones, J.W.; Brun, F. *Working with Dynamic Crop Models*, 3rd ed.; Academic Press: London, UK, 2018; pp. 3–43. [CrossRef]

6. Wallach, D.; Thorburn, P.; Asseng, S.; Challinor, A.J.; Ewert, F.; Jones, J.W.; Rotter, R.; Ruane, A. Multimodel ensembles improve predictions of crop–environment–management interactions. *Glob. Chang. Biol.* **2018**, *24*, 5072–5083. [CrossRef] [PubMed]

7. Schauberger, B.; Jägermeyr, J.; Gornott, C. A systematic review of local to regional yield forecasting approaches and frequently used data resources. *Eur. J. Agron.* **2020**, *120*, 126153. [CrossRef]

8. Azzari, G.; Jain, M.; Lobell, D.B. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sens. Environ.* **2017**, *202*, 129–141. [CrossRef]

9. Chlingaryan, A.; Sukkarieh, S.; Whelan, B. Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review. *Comput. Electron. Agric.* **2018**, *151*, 61–69. [CrossRef]

10. Cheng, E.; Zhang, B.; Peng, D.; Zhong, L.; Yu, L.; Liu, Y.; Xiao, C.; Li, C.; Li, X.; Chen, Y.; et al. Wheat yield estimation using remote sensing data based on machine learning approaches. *Front. Plant Sci.* **2022**, *13*, 1090970. [CrossRef]

11. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [CrossRef]

12. Jiang, H.; Hu, H.; Zhong, R.; Xu, J.; Xu, J.; Huang, J.; Wang, S.; Ying, Y.; Lin, T. A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level. *Agric. For. Meteorol.* **2020**, *284*, 107872. [CrossRef]

13. Shahhosseini, M.; Hu, G.; Archontoulis, S.V. Forecasting Corn Yield with Machine Learning Ensembles. *Front. Plant Sci.* **2020**, *11*, 1120. [CrossRef] [PubMed]

14. Wolanin, A.; Mateo-García, G.; Camps-Valls, G.; Gómez-Chova, L.; Meroni, M.; Duveiller, G.; Liangzhi, Y.; Guanter, L. Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. *Environ. Res. Lett.* **2020**, *15*, 024019. [CrossRef]

15. Meraj, G.; Kanga, S.; Ambadkar, A.; Kumar, P.; Singh, S.K.; Farooq, M.; Johnson, B.A.; Rai, A.; Sahu, N. Assessing the yield of wheat using satellite remote sensing-based machine learning algorithms and simulation modeling. *Remote Sens.* **2022**, *14*, 3005. [CrossRef]

16. Dhakar, R.; Sehgal, V.; Chakraborty, D.; Sahoo, R.; Mukherjee, J.; Ines, A.; Naresh Kumar, S.; Shirsath, P.; Baidya Roy, S. Field scale spatial wheat yield forecasting system under limited field data availability by integrating crop simulation model with weather forecast and satellite remote sensing. *Agric. Syst.* **2022**, *195*, 103299. [CrossRef]

17. Yang, S.; Li, L.; Fei, S.; Yang, M.; Tao, Z.; Meng, Y.; Xiao, Y. Wheat yield prediction using machine learning method based on UAV remote sensing data. *Drones* **2024**, *8*, 284. [CrossRef]

18. Ngoune, L.; Shelton, C. Factors Affecting Yield of Crops. In *Agronomy—Climate Change and Food Security*; IntechOpen: London, UK, 2020. [CrossRef]

19. Archana, S.; Kumar, P.S. A Survey on Deep Learning Based Crop Yield Prediction. *Nature Environ. Pollut. Technol.* **2023**, *22*, 579–592. [CrossRef]

20. Zhou, W.; Liu, Y.; Ata-Ul-Karim, S.T.; Ge, Q.; Li, X.; Xiao, J. Integrating climate and satellite remote sensing data for predicting county-level wheat yield in China using machine learning methods. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *111*, 102861. [CrossRef]

21. Pang, A.; Chang, M.W.L.; Chen, Y. Evaluation of Random Forests (RF) for Regional and Local-Scale Wheat Yield Prediction in Southeast Australia. *Sensors* **2022**, *22*, 717. [CrossRef]

22. Halder, M.; Datta, A.; Siam, M.K.H.; Mahmud, S.; Sarkar, M.S.; Rana, M.M. A Systematic Review on Crop Yield Prediction Using Machine Learning. In *Intelligent Systems and Networks*; Nguyen, T.D.L., Verdú, E., Le, A.N., Ganzha, M., Eds.; Springer: Singapore, 2023; Volume 752. [CrossRef]

23. Meghraoui, K.; Sebari, I.; Pilz, J.; Ait El Kadi, K.; Bensiali, S. Applied Deep Learning-Based Crop Yield Prediction: A Systematic Analysis of Current Developments and Potential Challenges. *Technologies* **2024**, *12*, 43. [CrossRef]

24. Yu, F.; Wang, M.; Xiao, J.; Zhang, Q.; Zhang, J.; Liu, X.; Ping, Y.; Luan, R. Advancements in Utilizing Image-Analysis Technology for Crop-Yield Estimation. *Remote Sens.* **2024**, *16*, 1003. [CrossRef]

25. Maestrini, B.; Basso, B. Drivers of within-field spatial and temporal variability of crop yield across the US Midwest. *Sci. Rep.* **2018**, *8*, 14833. [CrossRef] [PubMed]

26. Silva, L.; Conceição, L.A.; Lidon, F.C.; Patanita, M.; D'Antonio, P.; Fiorentino, C. Digitization of Crop Nitrogen Modelling: A Review. *Agronomy* **2023**, *13*, 1964. [CrossRef]

27. Cardillo, C.; Cimino, O. Small Farms in Italy: What Is Their Impact on the Sustainability of Rural Areas? *Land* **2022**, *11*, 2142. [CrossRef]

28. Fiorentino, C.; D'Antonio, P.; Toscano, F.; Donvito, A.; Modugno, F. New Technique for Monitoring High Nature Value Farmland (HNVF) in Basilicata. *Sustainability* **2023**, *15*, 8377. [CrossRef]

29. Leukel, J.; Zimpel, T.; Stumpe, C. Machine learning technology for early prediction of grain yield at the field scale: A systematic review. *Comput. Electron. Agric.* **2023**, *207*, 107721. [CrossRef]

30. Nevavuori, P.; Narra, N.; Linna, P.; Lipping, T. Crop Yield Prediction Using Multitemporal UAV Data and Spatio-Temporal Deep Learning Models. *Remote Sens.* **2020**, *12*, 4000. [CrossRef]

31. Shafiee, S.; Lied, L.M.; Burud, I.; Dieseth, J.A.; Alsheikh, M.; Lillemo, M. Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. *Comput. Electron. Agric.* **2021**, *183*, 106036. [CrossRef]

32. Fan, H.; Liu, S.; Li, J.; Li, L.; Dang, L.; Ren, T.; Lu, J. Early prediction of the seed yield in winter oilseed rape based on the near-infrared reflectance of vegetation (NIRv). *Comput. Electron. Agric.* **2021**, *186*, 106166. [CrossRef]

33. Roy Choudhury, M.; Das, S.; Christopher, J.; Apan, A.; Chapman, S.; Menzies, N.W.; Dang, Y.P. Improving Biomass and Grain Yield Prediction of Wheat Genotypes on Sodic Soil Using Integrated High-Resolution Multispectral, Hyperspectral, 3D Point Cloud, and Machine Learning Techniques. *Remote Sens.* **2021**, *13*, 3482. [CrossRef]

34. Sagan, V.; Maimaitijiang, M.; Bhadra, S.; Maimaitiyiming, M.; Brown, D.R.; Sidike, P.; Fritschi, F.B. Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning. *ISPRS J. Photogramm. Remote Sens.* **2021**, *174*, 265–281. [CrossRef]

35. Sharifi, A. Yield prediction with machine learning algorithms and satellite images. *J. Sci. Food Agric.* **2020**, *101*, 891–896. [CrossRef] [PubMed]

36. Zhao, Y.; Potgieter, A.B.; Zhang, M.; Wu, B.; Hammer, G.L. Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling. *Remote Sens.* **2020**, *12*, 1024. [CrossRef]

37. European Space Agency. Sentinel-2 Operations. Available online: https://www.esa.int/Enabling_Support/Operations/Sentinel-2_operations (accessed on 25 May 2024).

38. Segarra, J.; Buchaillot, M.L.; Araus, J.L.; Kefauver, S.C. Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications. *Agronomy* **2020**, *10*, 641. [CrossRef]

39. Hunt, M.L.; Blackburn, G.A.; Carrasco, L.; Redhead, J.W.; Rowland, C.S. High-resolution wheat yield mapping using Sentinel-2. *Remote Sens. Environ.* **2019**, *233*, 111410. [CrossRef]

40. Jewiss, J.L.; Brown, M.E.; Escobar, V.M. Satellite Remote Sensing Data for Decision Support in Emerging Agricultural Economies: How Satellite Data Can Transform Agricultural Decision Making [Perspectives]. *IEEE Geosci. Remote Sens. Mag.* **2020**, *8*, 117–133. [CrossRef]

41. Dhillon, M.S.; Dahms, T.; Kübert-Flock, C.; Steffan-Dewenter, I.; Zhang, J.; Ullmann, T. Spatiotemporal Fusion Modelling Using STARFM: Examples of Landsat 8 and Sentinel-2 NDVI in Bavaria. *Remote Sens.* **2022**, *14*, 677. [CrossRef]

42. European Space Agency. Sentinel-2 Products Specification Document. Available online: https://sentinel.esa.int/documents/247904/685211/sentinel-2-products-specification-document (accessed on 25 May 2024).

43. Faqe Ibrahim, G.; Rasul, A.; Abdullah, H. Sentinel-2 Accurately Estimated Wheat Yield in a Semi-Arid Region Compared with Landsat 8. *Int. J. Remote Sens.* **2023**, *44*, 4115–4136. [CrossRef]

44. Mancini, A.; Solfanelli, F.; Coviello, L.; Martini, F.M.; Mandolesi, S.; Zanoli, R. Time Series from Sentinel-2 for Organic Durum Wheat Yield Prediction Using Functional Data Analysis and Deep Learning. *Agronomy* **2024**, *14*, 109. [CrossRef]

45. Mandanici, E.; Bitelli, G. Preliminary Comparison of Sentinel-2 and Landsat 8 Imagery for a Combined Use. *Remote Sens.* **2016**, *8*, 1014. [CrossRef]

46. Nguyen-Thanh Son, C.-F.; Chen, Y.-S.; Cheng, P.; Toscano, P.; Chen, S.-L.; Tseng, K.-H.; Syu, C.-H.; Guo, H.-Y.; Zhang, Y.-T. Field-Scale Rice Yield Prediction from Sentinel-2 Monthly Image Composites Using Machine Learning Algorithms. *Ecol. Inform.* **2022**, *69*, 101618. [CrossRef]

47. Yli-Heikkilä, M.; Wittke, S.; Luotamo, M.; Puttonen, E.; Sulkava, M.; Pellikka, P.; Heiskanen, J.; Klami, A. Scalable Crop Yield Prediction with Sentinel-2 Time Series and Temporal Convolutional Network. *Remote Sens.* **2022**, *14*, 4193. [CrossRef]

48. Desloires, J.; Ienco, D.; Botrel, A. Out-of-Year Corn Yield Prediction at Field-Scale Using Sentinel-2 Satellite Imagery and Machine Learning Methods. *Comput. Electron. Agric.* **2023**, *209*, 107807. [CrossRef]

49. Amankulova, K.; Farmonov, N.; Mucsi, L. Time-Series Analysis of Sentinel-2 Satellite Images for Sunflower Yield Estimation. *Smart Agric. Technol.* **2023**, *3*, 100098. [CrossRef]

50. Fieuzal, R.; Bustillo, V.; Collado, D.; Dedieu, G. Combined Use of Multi-Temporal Landsat-8 and Sentinel-2 Images for Wheat Yield Estimates at the Intra-Plot Spatial Scale. *Agronomy* **2020**, *10*, 327. [CrossRef]

51. Li, L.; Wang, B.; Feng, P.; Wang, H.; He, Q.; Wang, Y.; Liu, D.L.; Li, Y.; He, J.; Feng, H.; et al. Crop Yield Forecasting and Associated Optimum Lead Time Analysis Based on Multi-Source Environmental Data Across China. *Agric. For. Meteorol.* **2021**, *308–309*, 108558. [CrossRef]

52. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]

53. Huang, S.; Tang, L.; Hupy, J.P.; Wang, X.; Shao, Y.; Chen, Y. A Commentary Review on the Use of Normalized Difference Vegetation Index (NDVI) in the Era of Popular Remote Sensing. *J. For. Res.* **2021**, *32*, 1–6. [CrossRef]

54. Sharma, M.; Bangotra, P.; Gautam, A.S.; Gautam, S. Sensitivity of Normalized Difference Vegetation Index (NDVI) to Land Surface Temperature, Soil Moisture and Precipitation Over District Gautam Buddh Nagar, UP, India. *Stoch. Environ. Res. Risk Assess.* **2022**, *36*, 1779–1789. [CrossRef]

55. Radočaj, D.; Šiljeg, A.; Marinović, R.; Jurišić, M. State of Major Vegetation Indices in Precision Agriculture Studies Indexed in Web of Science: A Review. *Agriculture* **2023**, *13*, 707. [CrossRef]

56.   Holben, B.N. Characteristics of maximum-value composite images from temporal AVHRR data. *Int. J. Remote Sens.* **1986**, *7*, 1417–1434. [CrossRef]

57.   Hollingsworth, B.V.; Chen, L.; Reichenbach, S.E.; Irish, R.R. Automated Cloud Cover Assessment for Landsat TM Images. In *Imaging Spectrometry II*; International Society for Optics and Photonics: Bellingham, WA, USA, 1996; pp. 170–179. [CrossRef]

58.   Irish, R.R.; Barker, J.L.; Goward, S.N.; Arvidson, T. Characterization of the Landsat-7 ETM+ Automated Cloud-Cover Assessment (ACCA) Algorithm. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 1179–1188. [CrossRef]

59.   Scaramuzza, P.L.; Bouchard, M.A.; Dwyer, J.L. Development of the Landsat Data Continuity Mission Cloud-Cover Assessment Algorithms. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1140–1154. [CrossRef]

60.   Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D., Jr.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Hughes, M.J.; Laue, B. Cloud Detection Algorithm Comparison and Validation for Operational Landsat Data Products. *Remote Sens. Environ.* **2017**, *194*, 379–390. [CrossRef]

61.   Qiu, S.; Zhu, Z.; He, B. Fmask 4.0: Improved Cloud and Cloud Shadow Detection in Landsats 4–8 and Sentinel-2 Imagery. *Remote Sens. Environ.* **2019**, *231*, 111205. [CrossRef]

62.   Xie, F.; Shi, M.; Shi, Z.; Yin, J.; Zhao, D. Multilevel Cloud Detection in Remote Sensing Images Based on Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3631–3640. [CrossRef]

63.   Jeppesen, J.H.; Jacobsen, R.H.; Inceoglu, F.; Toftegaard, T.S. A Cloud Detection Algorithm for Satellite Imagery Based on Deep Learning. *Remote Sens. Environ.* **2019**, *229*, 247–259. [CrossRef]

64.   Segal-Rozenhaimer, M.; Li, A.; Das, K.; Chirayath, V. Cloud Detection Algorithm for Multi-Modal Satellite Imagery Using Convolutional Neural Networks (CNN). *Remote Sens. Environ.* **2020**, *237*, 111446. [CrossRef]

65.   López-Puigdollers, D.; Mateo-García, G.; Gómez-Chova, L. Benchmarking Deep Learning Models for Cloud Detection in Landsat-8 and Sentinel-2 Images. *Remote Sens.* **2021**, *13*, 992. [CrossRef]

66.   Skakun, S.; Wevers, J.; Brockmann, C.; Doxani, G.; Aleksandrov, M.; Batič, M.; Frantz, D.; Gascon, F.; Gómez-Chova, L.; Hagolle, O.; et al. Cloud Mask Intercomparison Exercise (CMIX): An Evaluation of Cloud Masking Algorithms for Landsat 8 and Sentinel-2. *Remote Sens. Environ.* **2022**, *274*, 112990. [CrossRef]

67.   Dombrovsky, L.A.; Solovjov, V.P.; Webb, B.W. Attenuation of solar radiation by a water mist from the ultraviolet to the infrared range. *J. Quant. Spectrosc. Radiat. Transf.* **2011**, *112*, 1182–1190. [CrossRef]

68.   Sentinel Hub. Improving Cloud Detection with Machine Learning. Available online: https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13 (accessed on 25 May 2024).

69.   Google Earth Engine. COPERNICUS Sentinel-2 Cloud Probability Dataset. Available online: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_CLOUD_PROBABILITY#description (accessed on 25 May 2024).

70.   Google Earth Blog. More Accurate and Flexible Cloud Masking for Sentinel-2 Images. Available online: https://medium.com/google-earth/more-accurate-and-flexible-cloud-masking-for-sentinel-2-images-766897a9ba5f (accessed on 25 May 2024).

71.   Zeng, L.; Wardlow, B.D.; Hu, S.; Zhang, X.; Zhou, G.; Peng, G.; Xiang, D.; Wang, R.; Meng, R.; Wu, W. A Novel Strategy to Reconstruct NDVI Time-Series with High Temporal Resolution from MODIS Multi-Temporal Composite Products. *Remote Sens.* **2021**, *13*, 1397. [CrossRef]

72.   Yan, J.; Zhang, G.; Ling, H.; Han, F. Comparison of Time-Integrated NDVI and Annual Maximum NDVI for Assessing Grassland Dynamics. *Ecol. Indic.* **2022**, *136*, 108611. [CrossRef]

73.   Xie, Y.; Chen, Y.; Zhang, Y.; Li, M.; Xie, M.; Mo, W. Response of Vegetation Normalized Difference Vegetation Index to Different Meteorological Disaster Indexes in Karst Region of Guangxi, China. *Heliyon* **2023**, *9*, e20518. [CrossRef]

74.   Google Earth Engine. Make a Greenest Pixel Composite. Available online: https://developers.google.com/earth-engine/tutorials/tutorial_api_06#make-a-greenest-pixel-composite (accessed on 25 May 2024).

75.   Visual Crossing Weather Data Services. Available online: https://www.visualcrossing.com/weather/weather-data-services (accessed on 27 May 2024).

76.   Shang, Q.; Wang, Y.; Tang, H.; Sui, N.; Zhang, X.; Wang, F. Genetic, Hormonal, and Environmental Control of Tillering in Wheat. *Crop J.* **2021**, *9*, 986–991. [CrossRef]

77.   Xie, Q.; Mayes, S.; Sparkes, D.L. Optimizing Tiller Production and Survival for Grain Yield Improvement in a Bread Wheat × Spelt Mapping Population. *Ann. Bot.* **2016**, *117*, 51–66. [CrossRef] [PubMed]

78.   Pais, I.P.; Moreira, R.; Semedo, J.N.; Ramalho, J.C.; Lidon, F.C.; Coutinho, J.; Maçãs, B.; Scotti-Campos, P. Wheat Crop under Waterlogging: Potential Soil and Plant Effects. *Plants* **2022**, *12*, 149. [CrossRef]

79.   Kronenberg, L.; Yates, S.; Boer, M.P.; Kirchgessner, N.; Walter, A.; Hund, A. Temperature Response of Wheat Affects Final Height and the Timing of Stem Elongation under Field Conditions. *bioRxiv* **2020**, 756700. [CrossRef]

80.   Lombardia Region; General Directorate of Agriculture. Disciplinari Colture Cerealicole. Available online: https://www.risoitaliano.eu/customcontents/TRACCIA.pdf (accessed on 7 March 2024).

81.   Prabhakara, K.; Hively, W.D.; McCarty, G.W. Evaluating the Relationship Between Biomass, Percent Ground-Cover and Remote Sensing Indices Across Six Winter Cover Crop Fields in Maryland, United States. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *39*, 88–102. [CrossRef]

82.   EEB/Birdlife Europe. Soil and Carbon Farming in the New CAP: Alarming Lack of Action and Ambition. Available online: https://eeb.org/wp-content/uploads/2022/06/Briefing-Soil-Health-No-Branding-V2.pdf (accessed on 28 August 2024).

83. Meroni, M.; Waldner, F.; Seguini, L.; Kerdiles, H.; Rembold, F. Yield Forecasting with Machine Learning and Small Data: What Gains for Grains. *Agric. For. Meteorol.* **2021**, *308–309*, 108555. [CrossRef]

84. Lai, Y.R.; Pringle, M.J.; Kopittke, P.M.; Menzies, N.W.; Orton, T.G.; Dang, Y.P. An Empirical Model for Prediction of Wheat Yield, Using Time-Integrated Landsat NDVI. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 99–108. [CrossRef]

85. Liu, L.; Basso, B. Linking Field Survey with Crop Modeling to Forecast Maize Yield in Smallholder Farmers' Fields in Tanzania. *Food Sec.* **2020**, *12*, 537–548. [CrossRef]