Advancing Blackmore's methodology to delineate management zones from Sentinel 2 images

Arthur Lenoir[a,b,1], Bertrand Vandoorne[b], Ali Siah[b], Benjamin Dumont[a2]

[a]University of Liège, Junia, UMRT 1158 BioEcoAgro - Conduction, optimization and design of cropping systems meeting multi-criteria objectives, B-5030 Gembloux, Belgium

[b]Laboratoire Charles Viollette (EA 7394) Joint Research Unit 1158 BioEcoAgro Institut Supérieur d'Agriculture Junia, Université de Lille, Université de Liège, UPJV, Université d'Artois, ULCO, INRAE Lille Cedex France 48 Bd Vauban 59046

Acknowledgments:

Authors would like to acknowledge the team of the **"Ferme 3.0"** (experimental platform from the Chambre d'Agriculture de la Somme), **Messrs. Rivenet** (farmers in Hauts-de France (Fr)) and **Messrs. Goffaux** (farmers in Wallonia (Be)), for their support providing the yield maps used in this study.

Abstract:

Improving agricultural nitrogen management is one of the key objectives of the recent Green Deal in Europe. Current technological developments in agriculture offer new opportunities to improve nitrogen fertilization practices. The aim of this study was to adapt to Sentinel-2 data a proven delineation method initially developed for yield maps, in order to facilitate precise nitrogen management by farmers.

The study was conducted in two steps. Firstly, an analysis at annual level was conducted to assess the

[1] Corresponding author. E-mail address : arthur.lenoir@uliege.be (Arthur Lenoir) – ORCID : 0000-0002-1248-2268
[2] Benjamin Dumont: 0000-0001-8411-3990

relationship between vegetation indices and yield at the subfield scale, for different sensing period. The second step consisted in performing a pluri- annual analysis through the delineation of management zones and compare the results achieved from yield maps and from NDVI maps.

Among different vegetation indices, NDVI proved to be an interesting candidate for subfield detection of yield variation, specifically when the index was sensed was sensed around the second half of May. In this area, this period usually corresponds to phenological development between the flag leaf stage and heading stage, just prior the initiation of winter wheat flowering. Using NDVI maps within Blackmore's delineation approach instead of yield maps . allowed to reach an accuracy of 69% on zone classification. However, as yields and NDVI distribution do not respond to similar statistical distributions, we considered that the delineation threshold used to differentiate high from low yielding zones had to be adapted. The adaptation of the "performance threshold" in favor of the median NDVI, made it possible to achieve a higher accuracy (71%) of the delineation. But above all, the improvement lies also in a more robust satellite-based delineation.

Keywords:

Management zones, NDVI, spatio-temporal delineation, winter-wheat, Sentinel-2

Introduction

In Europe, high yielding crop productions are highly dependent on nitrogen (N) supplies (Jensen et al. 2011). The inappropriate management or the excessive use of N contributes to both serious economic and environmental consequences (Fowler et al., 2013; Willett et al., 2019 ; Schulte-Uebbing et al., 2022 ; Alexander et al., 2023). Improving Nitrogen Use Efficiency (NUE) remains therefore a major objective of modern N fertilization methods in crop productions (Zhang et al., 2015). Despite substantial improvements of NUE along the past decades, actual N fertilization approaches still fail to contain losses of N under acceptable thresholds (Lassaletta et al., 2014). Indeed, current methods struggle to precisely grasp temporal and spatial variability of N requirements. Fertilization decisions based on past performances and taken ahead of unknown climate conditions are likely to lead to inefficient N management; yet, they remain a common practice. In Europe, Ray et al. (2015) reported 31 to 51% yield variations due to interannual climate variability. Along growing periods, this interannual variability reveals large changes in weather conditions and crop N requirements in terms of quantities and dynamics (Gastal et al., 2015). Furthermore, substantial soil variability can be observed at the subfield scale, which will inherently react differently to climate conditions, and strengthen the uncertainty of fertilization practices (Delin et al., 2005). In this regard, the concept of management zone has been proposed as a suitable approach to improve both the economic and environmental aspects of fertilization (Basso et al., 2016).

In 2000, Blackmore proposed an interesting method to integrate both spatial and temporal dimensions in site-specific management of N (Blackmore, 2000). The method aims to delineate management zones (MZ) from yield maps time-series. Three types of zones are thus defined according to the yield level and its degree of stability. Unstable zones (U) are parts of the field where yield variations are strong over time. They oppose to stable zones that are subdivided according to their yield levels. Low and stable zones (LS) are parts of the field that always underperform compare to annual mean yields. Oppositely, high and stable zones (HS) are the stable part of the field where yield always outperform the annual mean yields. Methodology derived from Blackmore's approach were successfully used to

64 support better diagnostic of NUE at the subfield scale over large areas (Basso et al., 2019; Maestrini
65 and Basso, 2018; Martinez-Feria and Basso, 2020) and, coupled to the use of crop model, used to
66 improve management of N (Basso et al., 2013, 2011).
67 However, in its original form, the implementation of such method relies on yield monitoring. The
68 complex use of yield maps has been reported through different researches (Lyle et al. 2014, Leroux et
69 al. 2018). Moreover, in west-Europe, the rate of adoption of agricultural technologies remains limited.
70 A recent study carried out in France, revealed although combine harvesters are increasingly equipped
71 with yield monitors (nearly 30% by 2020), only a few farmers (around 1% of those who have equipped
72 their combine) use their yield maps as part of their farm work (Lachia et al. 2020).
73 Alternatively, satellites images provide precise information able to estimate plant traits and estimate
74 plant outcomes such as yield (Azzari et al., 2017; Hatfield et al., 2008; Revill et al., 2019; Vallentin et
75 al., 2022). Sentinel-2 missions launched by European Space Agency, provide freely available images
76 embedding useful wavelengths for agricultural applications (Delloye et al., 2018). To be informative
77 and robust, sensing data must be processed and integrated into agronomic methods. In this regard,
78 different Vegetation Indices (VI) have been developed to estimate plant biophysical variables such as
79 Leaf Area Index (LAI) or chlorophyll content (Baret, 2016; Ollinger, 2011).
80 Most VIs used to estimate LAI, biomass and yield rely on Red and Near Infra-Red (NIR) wavelengths.
81 The Red wavelengths are primarily absorbed by chlorophyll while NIR is scattered by leaf cell walls
82 (Ollinger 2011). Among the vegetation indices commonly used on annual crops, NDVI, the Normalized
83 Difference Vegetation Index (Tucker, 1974) is the most popular one. This index is related to plant vigor
84 and often used to estimate photosynthetic activity or plant biomass through LAI. NDVI is known to
85 saturate in high LAI (from LAI = 2) (Gitelson et al., 2003). To overcome the saturation issue, different
86 vegetation indices were since proposed in the literature (Daughtry et al., 2000; Gitelson et al., 2003;
87 Haboudane, 2004; Huete et al., 2002; Tucker, 1979).
88 Using supplementary wavelengths, found in the visible and the Red-edge part of the light spectrum,
89 these indices were proposed to improve LAI estimation and by extension biomass and yield estimation

(Haboudane, 2004).

Since Sentinel-2 satellites were launched, giving access to high resolution images, only few studies have assessed the ability of VIs to detect the subfield yield variations (Hunt et al., 2019; Revill et al., 2019; Skakun et al., 2021). Yet, each of these studies reported interesting results on the ability of specific wavelengths, embedded in complex statistical models, to retrieve subfield yield variations. They highlighted the higher sensitivity of wavelengths such as NIR, Red-edge and Green over the Red to capture spatial variations of yield . They also revealed the added value of high-resolution sensing to detect spatial variations. So far, no consensus was reported on the superiority of wavelengths accuracy over their spatial resolution to optimally detect the yield spatial variations. In addition, the precision of VIs depends upon the crop, its environment (Diacono et al., 2013; Fang et al., 2019; Maestrini and Basso, 2018; Vannoppen and Gobin, 2021) and the growth stage at which the sensing occurs (Marti et al., 2007; Panek and Gozdowski, 2020).

In the context of West-Europe, there is a clear room to improve the management of nitrogen fertilization, notably through the concept of management zones. Provided it could be deployed at the large scale and adapted to the high yielding agronomic context of western Europe, Blackmore's approach might provide an important opportunity. However, large uncertainty remains on the ability of VIs sensed from Sentinel-2 images to replace yield data in Blackmore's delineation approach. In this regard, the present study aims to address the following research questions: (i) Which vegetation index and which timeframe of records are best suited to capture the subfield yield variations ? (ii) How to optimally implement Blackmore's delineation method using directly the information from vegetation index ?

Materials and methods:

Study area

The study was conducted on three farms situated in Belgium (Wallonia region) and France (Hauts-de-France region) (Table 1). Both regions have similar pedoclimatic and agronomic contexts. Soils are

characterized by their high productivity (deep soils characterized by high silk content). Climate is temperate with annual rainfall ranging from 451 mm to 868 mm (Table 1). In both regions, winter wheat is the main crop grown on arable land but regional crop rotations are also composed of barley, rapeseed, potatoes, and sugar beet (Agreste 2020; SPW, 2022). In those countries, the most common fertilization approaches consist in balancing annually the difference between soil N supply and crop N demand with N fertilization (Dumont et al., 2015b; Machet et al., 2017; Meynard et al., 1997; Ravier et al., 2016). Average amount ranges between 170 and 190 kgN/ha in the two above mentioned regions (Agreste, 2019 ; EAW 2022).

Table 1 : Features of the studied fields

	Site A	Site B	Site C
Localization	Wallonia (Be)	Hauts-de-France (Fr)	Hauts-de-France (Fr)
Soil type[3]	Stagnic albe-luvisol	Calcosol	Fluvisol
Texture[4]	Loam	Clay-loam	Silt-loam
Soil depth[5]	>125 cm	40 [25-85] cm	>100 cm
Mean annual rainfall (mm)[6]	533 [451 – 721]	680 [530 - 717]	740 [666 - 868]
Mean annual temperature (°C)[3]	10.3	11.1	11.3
Average solar radiation ($MJ.m^{-2}.day^{-1}$)	10.6	11.8	11.5
Altitude (m)	300	130	25
Number of fields	15	3	17
Fields average surfaces (ha)	5.9 [0.97 - 24]	9.8 [6.4 – 11.9]	9.7 [0.7 - 20.1]

Collection and filtering of yield map data

On the three farms, yield maps were collected between 2015 and 2021 on fields where winter wheat was grown and uniform N fertilizations were supplied. Yield maps were acquired from three different monitoring systems mounted on John Deere T560 (Site A), New Holland CX 7.80 (Site B) and Claas Lexion 750 Terra Trac (Site C). In total, yield maps from 35 plots were included in the analysis. Yield map records varied from one to three years due to different crop rotations; thus, of the 35 fields, 46

[3] GIS SOL 2019; Dumont et al. 2015
[4] Legrain & Block 2009
[5] Le Bas 2021
[6] Météo-France, IRM 2021

yield maps were available, for a total mapping of 312 ha. Of the 35 fields, only six had sufficient yield map history to allow the delineation method to be implemented. Therefore, the initial stage of the analysis was conducted on the 35 fields while analysis based on delineation was carried out on the six fields only.

A two-step filtering procedure was performed on yield maps when raw data were accessible. The filtering methodology was largely inspired by Leroux et al. (2018). The first filtering stage, considered as a global outlier treatment is proposed by Lyle et al. (2014). As yield is estimated from the product of grain flow entering the harvester and from machine speed, yield values corresponding to lower cutting width than the maximum were removed. Yield values having an inter-record distance out of the 90 % interval were also removed. Finally, from the pre-filtered data and to ensure data consistency, only records in the 90 % yield confidence interval were conserved.

The second filtering stage is a local outlier treatment relying on a clustering approach (Leroux et al. 2018). The filtering algorithm was downloaded from Leroux (2020). Few adaptations were implemented in the algorithm to include the anisotropic treatment of data suggested in Leroux et al. 2018. The performance of the two stages filtering is summarized in Table 2.

Table 2 :Distribution and spatial features of yield maps before and after outliers' treatment

		Yield performances				Spatial trends of yield	
	Removed records	Yield range ($Mg.ha^{-1}$)	Mean yield ($Mg.ha^{-1}$)	Spatial CV	Skewness	Sill	Range
Raw data	0	[0 - 50]	7.32	35%	- 0.70	1889	8455
Global filtering	-13%	[1.5 – 17.2]	7.78	21%	-0.57	2.49	361
Global + local filtering	-19%	[1.5 – 17.2]	7.86	19%	-0.49	2.28	77

Satellite data collection

Sentinel-2 images sensed from the Multi-Spectral Instrument (MSI) were downloaded from the platform of the French scientific structure Theia (*Pôle Theia, 2022*). Only Sentinel-2A images were

153 available in 2016 while Sentinel-2B satellite was launched in 2017, which increased the revisit time up
154 to five days. Downloaded images from the Theia platform were already corrected with the MAJA
155 software (Hagolle et al., 2015), allowing access to images that were orthorectified, georeferenced (by
156 100 km x 100 km tiles) and benefited from atmospheric correction. This treatment corresponds to "2A
157 treatment" or "Bottom of Atmosphere (BoA)" correction. Only wavelengths with resolutions of 10 m
158 and 20 m were downloaded. Acquisition period was set, from 2016 to 2021, between the 1st of April
159 to 30th of June and only images with less than 70% of cloud coverage were selected to ensure maximal
160 availability of pixels during the period. Along each growing periods, satellite images were co-registered
161 using Arosics software (Scheffer et al. 2017). The 20 m resolution bands were automatically
162 downscaled by the software to a 10 m resolution using a cubic method. Each image available along the
163 years 2016 – 2021 was cropped using the different fields masks and combined with their corresponding
164 yield maps into data cubes.
165 Thereby, a yield map data cube contains the yield map and satellite images recorded at each available
166 sensing date. Missing pixels within the yield maps were used to filter corresponding pixels of satellite
167 images. On the contrary, missing pixels of the satellite images, within the boundaries of yield maps,
168 were conserved to guarantee the integrity of yield maps.
169
170 *Table 3: Number of downloaded images from Theia platform per years and months*

Year	April	May	June	Total
2016	1	2	2	5
2017	2	3	1	6
2018	2	3	2	7
2019	4	2	2	8
2020	7	11	4	22
2021	4	2	8	14
Average	3	4	3	

171
172 **Vegetation indices**
173 Vegetation indices were computed from Sentinel-2 images to assess their ability to retrieve subfield
174 yield variations. Obviously, being the most popular one, the NDVI - Normalized Difference Vegetation

Index – proposed by Tucker (1974) was assessed. However, although widely used, NDVI is known to saturate at high LAI, which would limit its accuracy in finely detecting yield subfield spatial variations (Haboudane, 2004). To overcome this limit, various indexes have been proposed (Darra et al. 2023). The Enhanced Vegetation Index (EVI) has been introduced as it was less sensitive to saturation (Ibid). EVI embed NIR and Red wavelengths, similarly as NDVI, but introduces blue wavelength and a correction coefficient to limit the saturation effects (Huete et al., 2002). Different vegetation indices using similar wavelengths as NDVI, namely OSAVI (Daughtry et al. 2000) and MSAVI (QI et al.1994) were tested in early stage of the study. As Spearman correlation were used (cf. Statistical analysis section), these VI yielded very similar results as NDVI and were not kept further. The Modified Chlorophyll Absorption in Reflectance Index (MCARI) was developed to detect subtle changes in chlorophyll variations (Daughtry et al. 2000). It was expected that chlorophyll content would indicate the vigor state of the crop which would result in yield differences at the subfield scale. Improvements of this index were proposed by Haboudane (2004) to create new LAI sensitive indices (MCARI 1 and MCARI 2).

Table 4: Equations of the vegetation indices used in the study. By default, NIR corresponds to the wavelength sensed at 833 nm and Red-edge corresponds to the wavelength sensed at 704 nm.

Index	Equations	Spatial resolution (m)	Source	
NDVI	$\frac{(\mathrm{NIR}-\mathrm{Red})}{(\mathrm{NIR}+\mathrm{Red})}$	10	Tucker 1979	(1)
EVI	$2.5 \times \frac{(\mathrm{NIR}-\mathrm{Red})}{((\mathrm{NIR}+6\times\mathrm{Red}-7.5\times\mathrm{Blue})+1)}$	10	(Huete et al., 2002)	(2)
MCARI	$(\text{Red-edge - Red}) - 0.2 \times(\text{Red-edge - Green}) \times \left(\frac{\text{Red-edge}}{\text{Red}}\right)$	20	Daughtry et al. 2000	(3)
MCARI1	$1.2 \times (2.5 \times (\mathrm{NIR} - \mathrm{Red}) - 1.3 \times (\mathrm{NIR} - \mathrm{Green}))$	10	Haboudane 2004	(4)
MCARI2	$\frac{1.5 \times 2.5(\mathrm{NIR} - \mathrm{Red}) - 1.3(\mathrm{NIR} - \mathrm{Green})}{\sqrt{(2\mathrm{NIR}+1)^2 - (6\mathrm{NIR} - 5\sqrt{\mathrm{Red}}) - 0.5}}$	10	Haboudane 2004	(5)
CIgreen	$\frac{NIR}{Green} - 1$	10	(Gitelson et al., 2003)	(6)
CIrededge	$\frac{NIR_{865}}{Rededge} - 1$	20	(Gitelson et al., 2003; Revill et al., 2020)	(7)

The Chlorophyll Indices based on Red-edge wavelength ($CI_{rededge}$) and Green wavelength (CI_{green}) were also developed for chlorophyll detection (Gitelson et al. 2003). They were used by (Revill et al., 2020)

194 to detect subfield spatial variations in of wheat yields and obtained interesting results compare to
195 NDVI. All vegetation indices were systematically computed on each satellite images used in this study.

196

197 **Blackmore's approach of management zones delineation**

198 The methodology proposed by Blackmore (2000) was used to delineate management zones. This
199 approach considers both spatial (eq. 8 and eq. 9) and temporal (eq. 10) yield variations.

$$s_i = \left(\frac{y_i}{\bar{y}}\right) \times 100 \tag{8}$$

$$\bar{s}_i = \frac{\sum_{t-1}^{n} S_{i_t}}{n} \tag{9}$$

$$\mathrm{CV}_i = \frac{\left(\frac{\sum_{t=1}^{t=n} s_{i_t}^2 - \left(\sum_{t=1}^{t=n} s_{i_t}\right)^2}{n(n-1)}\right)^{0.5}}{\bar{s}_i} \times 100 \tag{10}$$

200 Where s_i is the standardized yield (y) on location *i*, $\bar{y}$ is the annual mean yield of the field, $\bar{s}_i$ the average
201 of standardized yield over *n* years on each location *i*, *t* is one of *n* years considered in the delineation.
202 *CV* is the temporal coefficient of variation.
203 Once computed, $\bar{s}_i$ and CV_i are opposed to thresholds to classify each field pixel in one of the three
204 categories, leading to the management zones (Table 5). The thresholds are conserved as in the original
205 method proposed by Blackmore (2000). When CV_i is below the threshold of 30%, the pixel is classified
206 in the "stable" part of the field; above the 30% threshold, the pixel is considered "Unstable" (U). Among
207 stable locations, $\bar{s}_i$ above the 100% threshold is considered High, leading to "High and Stable" (HS)
208 zones, while pixels under the 100% threshold are considered Low, leading to the "Low and Stable" (LS)
209 zones (Blackmore 2000). Thus, each field can be delineated in three types of management zones,
210 namely the HS, LS and U. At least three years of record, that is three yield maps or satellite maps, are
211 required to run the algorithm.
212 The original methodology, *i.e.*, using the annual mean yield $\bar{y}$, was implemented from yield map
213 records. Adaptation of the methodology to substitute yield information by the information originating

from vegetation index will be proposed in this paper.

Table 5 : Classification of management zones according to a threshold of performance (first condition) and interannual variability (second condition). The table is inspired by Blackmore 2000.

Management zone	First condition	Second condition
HS (High & Stable yielding zone)	$\bar{s}_i > 100$	$CV_i < 30$
LS (High & Stable yielding zone)	$\bar{s}_i < 100$	$CV_i < 30$
U (Unstable yielding zone)		$CV_i > 30$

Statistical analyses

Selecting a vegetation index to optimally retrieve annual yield variations

A benchmark was performed between the different vegetation indices to assess their ability to catch the yield variations. Compared vegetation indices were systematically computed on the same satellite images for each field and sensing date so we assumed a constant spatial structure. The benchmark was focused on the attribute variations of the indices to assess the sensitivity of each index. As there is no special evidence of linear relationship between VI's and yield maps, Spearman rank correlation between the yield maps and the vegetation indices were preferred over Pearson's. They were computed from the related satellite images of the data cubes.

Optimal sensing window

After identification of the most suited vegetation index (VI) to capture the yield variations, an optimal sensing window was searched for. Indeed, as they rely on different crop biophysical traits that change spatially and temporally, VIs are also expected to vary along the growing period. As the central objective of the study was to substitute yield maps in Blackmore's delineation method, we searched for the sensing period that optimally retrieved the spatial yield variation. To do so, Moran's global index was first used to assess the spatial structure of data and exclude problematic data from analysis, while Moran's I bivariate and Spearman correlation were referred to actually identify optimal sensing windows. Procedures are detailed here below.

Among the satellite images sensed every year, a maximal cloud cover of 70% was tolerated. However, the cloud cover might differently affect VI maps across fields and/or sensing dates. As a result, for a given field, the annual spatial structure of the VI-maps would change along the period according to missing pixels (the one removed due to cloud cover). To detect potentially non-comparable images, each yield map was associated in stacks with each of the related VI maps acquired during the detection period. From the created stacks, the missing pixels from each VI-map were used to filter those pixels from the yield maps. By doing so, different filtered versions of each yield map were obtained for each field and year.

Moran's global index (I) was then computed on each version of the yield map to assess the spatial structure differences among the versions of the yield maps. When a large difference in Moran's I value was observed between different versions of the filtered yield map, the stacks having larger difference than a given threshold were excluded from the analysis, considering that they would bring a bias. This threshold was set as a positive or negative variation larger than 0.3 on standardized Moran's I values. The standardization was carried out per field and year removing the mean and dividing by the standard deviation each Moran's I value The conserved maps were considered spatially comparable as they would offer a sufficiently similar yield map spatial structure.

Ultimately, both Spearman rank correlation and local Moran's I bivariate were computed between conserved yield maps and each of their related VI-maps to identify the best sensing window. Spearman correlation analysis only consider the relationship between satellite-based map and yield map through attribute information. However, the relationship can also be affected by the spatial distribution of each variable. This issue is illustrated in Fig. 1. Local Moran's I bivariate allow to measure the relationship between one variable and the average value of a second variable's neighborhood (Anselin et al., 2010). Moran's indices were ranked per year and field. Then, within the different two- weeks subperiods, the frequencies of each of the three first ranks were computed and expressed as a percentage of the total appearance of the rank over the whole season.

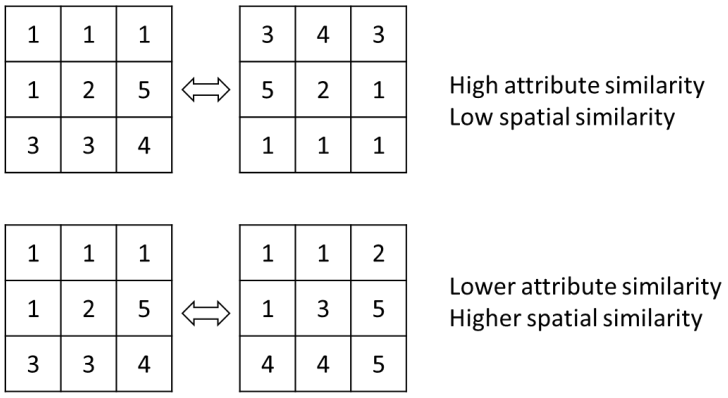

Fig. 1 Graphical representation of heterogeneity analysis on geographic and attribute spaces

Accuracy assessment of the satellite based-delineation

Across the cleaned database of yield maps, it was possible to implement Blackmore's delineation on six fields, for which at least three yield maps were available. The delineation was computed both from yield maps and from the related VI maps. Based upon prior results gained, maps of the most adapted VI and optimal sensing window were used at this step. An exception occurred in 2016; most satellite images being affected by cloud coverage VI maps the most correlated to yield maps, regardless of the sensing window, were selected.

To compute the delineation, both yield maps and VI maps were already projected on a common field grid with a resolution of 10 m, which was the coarser resolution between yield maps and satellite images.

On three out of the six fields, yield maps were not available over the period 2017 – 2018 due to records problem, while Sentinel-2 images were only available starting 2016. For those fields, considering that Blackmore's method of delineation should be robust over years, yield maps recoded in 2015 and Sentinel-2 images sensed in the period 2017 – 2018 were used (**Table S8**).

To assess the accuracy of the satellite-based delineation, 3 x 3 confusion matrices were computed on each field between both yield map-based and VI based delineation (**Fig. 2**) (Tharwat, 2021). The confusion matrices report the number of pixels belonging to each management zones that were correctly or incorrectly classified.

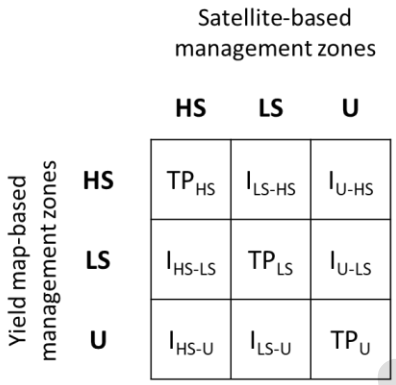

Fig. 2 : Confusion matrix used to assess the accuracy of satellite-based delineation. 'TP' refers to True Positive (the correctly classified pixels), 'I' refers to incorrectly classified pixels.

From these confusion matrices, different metrics were calculated, namely the accuracy, the specificity, and the sensitivity (Makowski et al., 2009). The accuracy (example for HS accuracy is provided eq. 11) measured the part of correctly classified pixels among all pixels (each management zones included). The specificity (eq. 12) measured the proportion of pixels that were correctly not classified in a management zone (equation for HS specificity is provided eq. 5). The sensitivity (eq. 13) measured the proportion of correctly classified pixels among the pixels classified in the same management zone (from satellite images - horizontally on **Fig. 2**). The confusion matrices and the different metrics were computed with the R package "caret" (Kuhn, 2008).

$$\text{Accuracy } (HS) = \frac{TP_{HS} + TP_{LS} + TP_U + I_{LS-U} + I_{U-LS}}{TP_{HS} + TP_{LS} + TP_U + I_{LS-U} + I_{U-LS} + I_{LS-HS} + I_{U-HS} + I_{HS-LS} + I_{HS-U}} \quad (11)$$

$$\text{Specificity (HS)} = \frac{TP_{LS} + TP_U + I_{LS-U} + I_{U-LS}}{TP_{LS} + TP_U + I_{LS-U} + I_{U-LS} + I_{LS-HS} + I_{U-HS}} \quad (12)$$

$$\text{Sensitivity (HS)} = \frac{TP_{HS}}{TP_{HS} + I_{LS-HS} + I_{U-HS}} \quad (13)$$

Adaptation of Blackmore's algorithm to satellite data

Blackmore's delineation method was initially elaborated to classify yield data. The statistical principles involved in the delineation proposed by Blackmore assumes normally distributed data. In Blackmore's algorithm, pixels are classified each year according to the mean yield value of the field (Eq. 8). An example of yield distribution gained on one of this study's field is provided in Fig 3 (left panel). The distribution of NDVI data (optimal sensing time) corresponding to the same field is presented at **Fig. 3 – right panel**.

No evidence of a symmetry between both the yield and NDVI shape distributions can be observed on **Fig. 3**. In that sense, the mean NDVI would not correspond to the mean yield. Consequently, mean NDVI would not classify sub-field zones similarly if used instead of mean yield. In non-normal distribution, the median is often considered a better estimator than mean.

To identify the optimal NDVI threshold splitting distributions in ranges of values leading to low and stable, high and stable or unstable yield zones, various NDVI thresholds were assessed around (1) the mean and (2) the median of NDVI's distributions, up to plus or minus 0.10, with a 0.01 step.

New confusion matrices were then computed for each potential thresholds drawn from the mean and from the median with the aim to identify the optimal value from the derived accuracies and adapt NDVI-based delineation.

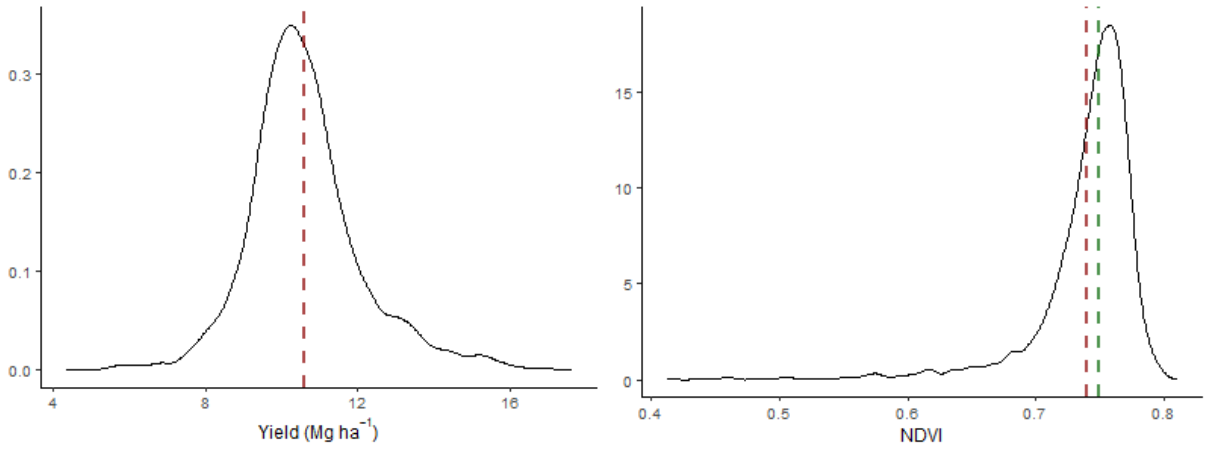

Fig. 3: Example of yield (left panel) and NDVI (right panel) aggregated distributions from different fields available in 2019. Yield followed a normal-like distribution while NDVI followed a beta distribution. Red dotted lines represent the distribution mean and green dotted line (right panel only) represents the distribution median.

Software used

Data treatment and statistical analyses were performed with R software (version 4.2.3), QGIS (version 3.22) and Python (version 3.9). Specifically, satellite images were treated from Python and the package Arosics. Data treatment and visualization were done with R packages dplyr (Wickham et al., 2023a), tidyr (Wickham et al., 2023b), lubridate (Grolemund and Wickham, 2011), sf (Pebesma, 2018) and ggplot2 (Wickham, 2016). Data analysis was performed with sfdep (Parry, 2022) and caret (Kuhn, 2008).

Results :

Benchmarking the vegetation indices and sensing windows

Maps of vegetation indices (VI) were computed on the 35 fields of the study at each sensing date available during the period (1st of April – 30th of June) between 2016 and 2021. From maps of VIs and yield maps, Spearman rank correlations (rho) were calculated. Annual VIs correlations distributions are displayed in **Fig. 4**.

Regardless the VI considered, correlation levels changed from year to year. In 2017 and 2020, correlations were the highest with in average rho_{2017} = 0.49 and rho_{2020} = 0.37. Also, the year 2020 gathered the greatest number of satellite images over the season (**Table 3**). In 2016, 2018, 2019 and 2021, Spearman correlations ranged on average from 0.19 to 0.37.Among VIs, EVI, MCARI, MCARI1 and MCARI 2 systematically underperformed compared to NDVI (**Fig. 4**). From the above-mentioned VIs, MCARI2 correlations were the closest to those of NDVI. However, interannual median correlations of MCARI2 (rho_{median} = 0.33) remained lower than the median correlations of NDVI (rho_{median} = 0.39). Only CI_{green} and $CI_{rededge}$ succeeded to reach as high correlations as NDVI. All year included, the median correlations of CI_{green} and $CI_{rededge}$ were respectively 0.39 and 0.36, comparable with NDVI ($rho_{cmedian}$ = 0.39). Their 3rd quantiles correlation (CI_{green} : rho_{q3} = 0.54 ; $CI_{rededge}$: rho_{q3} = 0.53) were slightly higher or equivalent to those of NDVI (rho_{q3} = 0.53). In 2016, 2017 and 2020, CI_{green} obtained slightly higher median correlations with yields (CI_{green} : rho_{2016} = 0.32 ; rho_{2017} = 0.60 ; rho_{2019} = 0. 49) than NDVI (NDVI

: rho_{2016} = 0.28 ; rho_{2017} = 0.57 ; rho_{2019} = 0.46). This trend was only visible at each date along the sensing period in 2017 whereas the superiority of CI green on NDVI largely depended on the sensing dates in 2016 and 2020 (**Fig. S3**).

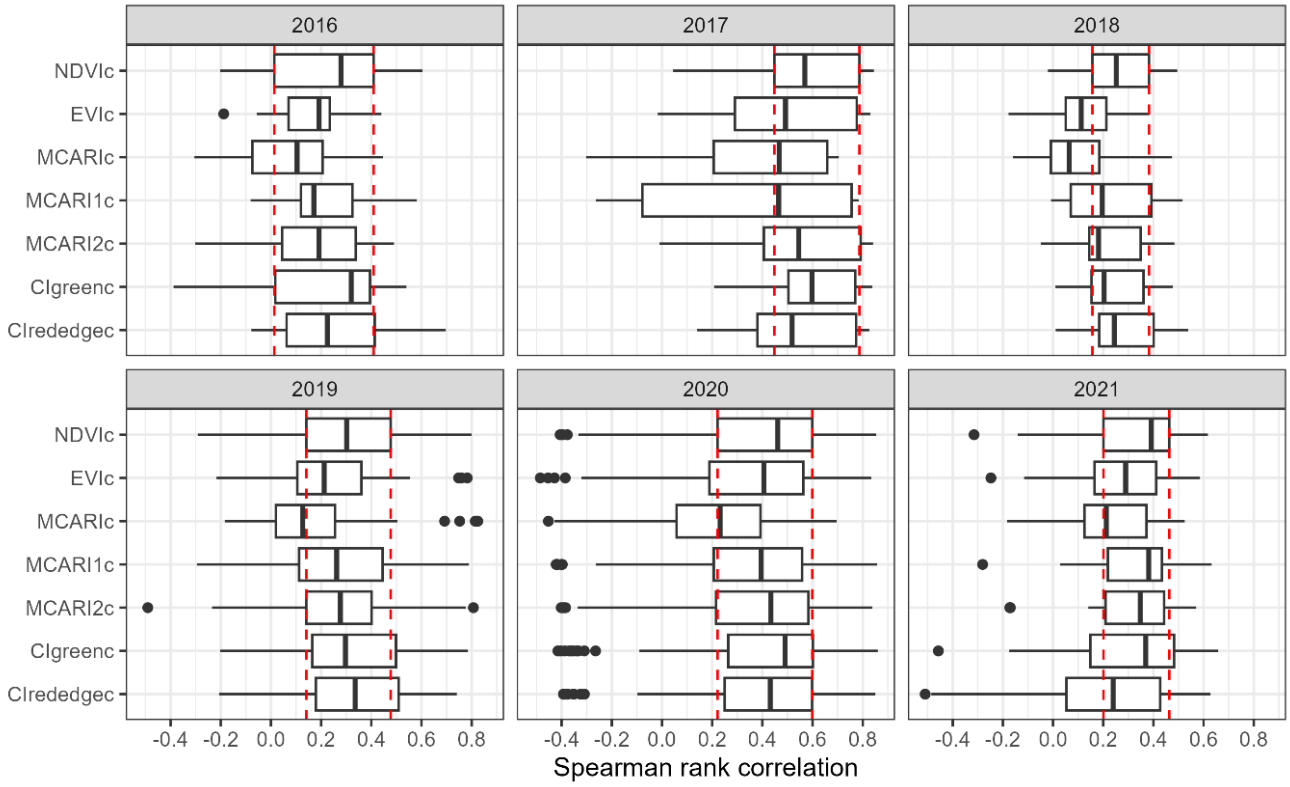

Fig. 4: Spearman correlation distribution computed on each sensing date between yield and vegetation indices on the 35 fields. Vertical dotted lines represent the first and third quantile of NDVI correlation distributions observed each year.

$CI_{rededge}$ obtained higher median correlation than NDVI in 2019 only ($CI_{rededge}$: rho_{2019} = 0.33; NDVI: rho_{2019} = 0.29). Again, this global observation was nuanced at specific dates along the season (Fig S3). Also, in 2021, 25% of CI_{green} correlations were lower than the 1st quantile distribution of NDVI correlations (CI_{green} rho_{q1} = 0.14 ; NDVI : rho_{q1} = 0.20). $CI_{rededge}$ underperformed the 1st quantile correlation of NDVI in 2017 and 2021 ($CI_{rededge}$: rho_{q1} = 0.05 and rho_{q1} = 0.38 ; NDVI : rho_{q1} = 0.20 ; rho_{q1} = 0.44). Overall, the three indices obtained very similar correlations (**Fig.4**).

On the 302 correlations performed between the different yield maps and NDVI computed at the sensing dates, 35 were negative and occurred every year except 2017. These negative correlations mainly appeared at the end of the sensing period; few fields also obtained negative correlations during

most of the period in 2020 (**Fig. S3**).

The autocorrelation analysis performed on the yield maps of the different stacks is displayed on **Fig. 5** for years 2016, 2019 and 2020 (complete figure is available in **Fig. S1**). Spatial structures did not show any irregularities for year 2017, 2018 and 2021.

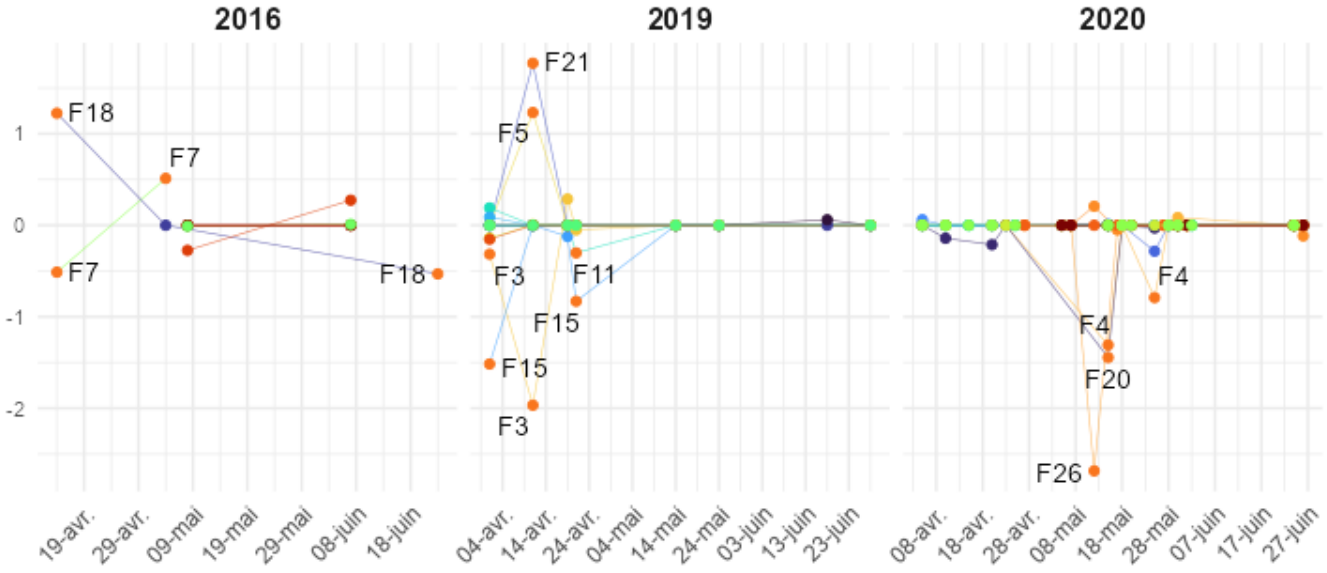

Fig. 5: Moran's I computed on yield maps filtered from the corresponding satellite images and their available pixels. F1 to F35 are the different fields of the study.

Due to cloud coverage, the yield maps filtering implemented from missing pixels of the different satellite images changed the yield map spatial structures on 20 combinations of field and date. In total, 12 fields were affected by these detection issues (**Fig. 5**). Changes in the spatial structure occurred in 2016 but only few satellite images were available due to global high cloud cover all along the sensing period. Early April 2019, from 11th to 21th, cloud coverage strongly affected the maps spatial structures which involved six fields . In 2020, sensing was affected during the middle two-weeks of May (F5, F23, F30). In 2019 and 2020, affected satellite images were excluded from the rest of the analysis.

From the remaining satellites images, a correlation analysis completed by a local bivariate autocorrelation analysis were performed between each yield map and their related maps of NDVI. The correlation analysis is summarized at **Table 6**. Minimum, median, and maximum correlations obtained from the different fields were reported per month's quarters (approximately a week). Periods when

no correlations were provided correspond to cloud cover superior to the 70 % threshold. From 2016 to 2021, every week was at least once affected by cloud cover with no available image. It is important to note that the reported correlations in each subperiod correspond to fields where the number of available images changed each year. This would influence the conclusion derived from **Table 6**.

Table 6: Spearman correlations performed between yield maps and their related maps of NDVI summarized per month quarters over the sensing period and from 2016 and 2021. Bold values are the maximal median correlations along the sensing periods. Minimum (min), median (med) and maximum (max) values are displayed for each subperiod. The average correlation is the average median correlation for each subperiod.

		April				Mai				June				Number of fields
		01-07	08 -14	15-21	22-30	01-07	08 -14	15-21	22-30	01-07	08 -14	15-21	22-30	
Correlations														
2016	Min		0.35			0.39	0.01			-0.20				
2016	Med		0.35			**0.47**	0.23			-0.07				8
2016	Max		0.35			0.54	0.49			0.47				
2017	Min		0.55	0.46			0.26		0.40		0.04			
2017	Med		0.57	0.58			0.53		**0.64**		0.46			2
2017	Max		0.58	0.68			0.77		0.84		0.84			
2018	Min		0.18	0.14		-0.02		0.16		0.19			0.04	
2018	Med		0.34	**0.36**		0.15		0.27		0.32			0.36	3
2018	Max		0.39	0.47		0.29		0.35		0.49			0.46	
2019	Min	0.02	-0.23	0.08	0.09		0.20		-0.13			0.09	-0.78	
2019	Med	0.23	0.34	0.34	0.32		0.38		**0.40**			0.19	0.30	13
2019	Max	0.72	0.73	0.80	0.76		0.66		0.58			0.29	0.62	
2020	Min	-0.31	-0.38	-0.40	-0.41	0.13	-0.31	-0.33	-0.19	0.18			0.08	
2020	Med	0.11	0.19	0.22	0.28	0.50	0.36	0.41	0.47	**0.55**			0.53	16
2020	Max	0.45	0.50	0.60	0.65	0.81	0.83	0.76	0.85	0.85			0.81	
2021	Min			0.45	0.13				0.19	0.21	0.05	-0.31	0.51	
2021	Med			**0.51**	0.29				0.27	0.29	0.25	0.18	0.51	5
2021	Max			0.59	0.62				0.48	0.57	0.43	0.25	0.51	
Average		0.20	0.32	0.39	0.30	0.36	0.37	0.27	**0.40**	0.32	0.35	0.11	0.33	

To overcome this limitation, median correlations of each field at each sensing date were normalized per field and year and displayed in **Fig. S2| top panel.** The years 2019 and 2020 gathered the highest number of yield maps and available sensing dates while in average, only two satellite images were available per field in 2016. Overall, in 2017, correlation levels were the highest but relied on two fields only (**Table 6**). The highest median correlations, marked in bold on **Table 6**, were usually observed at two subperiods: during the third week of April and during May, mainly at the end of the month. While a global drop of the median correlations was observed after the end of May, this decrease was not observed on all fields. The maximum observed correlations remained high during June from 2017 to

2020 while very low or negative correlations were also reported, indicating large fields differences within the trend (**Table 6**). The correlation drop was also visible from normalized correlations (**Fig S2| top panel**).

The bivariate autocorrelation analysis was implemented on the same variables as the correlation analysis. It resulted in Moran's I value ranging between -1 and 1. Under Moran's statistics, values the closest to "1" indicate a similar spatial structure between the two analyzed variables, here subfield yield and NDVI distributions. On each field, Moran's I bivariate were computed along each sensing period and averaged per two-weeks period. Results are reported in **Table 7**. Similarly, to the correlation analysis, the best autocorrelations were found in the second part of May.

When considering solely the first rank, the percentages of highest autocorrelations were closed to each other in May (23% and 27%), and the second part of June (24) (**Table 7**). Strong differences arose for the second and third ranks for which the bivariate spatial autocorrelation reported during the second half of May outperformed the other periods. Moran's I bivariate results were normalized per field and year and reported in **Fig. S2| bottom panel**.

The highest spatial autocorrelations found in 2017, 2019 and 2020 at the end of May matched the trends of the correlations (**Fig. S2 upper panel**). In 2018 only, the higher autocorrelation expressed later than the higher correlation level. Overall, Spearman correlation and spatial autocorrelation both revealed an optimal sensing window for NDVI at the end of May. In the next part, the satellite-based delineation was computed using these results.

Table 7 : Occurrence (percentage per rank) of the three optimal bivariate spatial autocorrelations over the sensing period per two-weeks.

Months	April		May		June	
Rank of Moran's I bivariate	01 - 14	15 - 30	01 - 14	15 - 31	01 - 14	15 - 30
1	5	14	24	27	5	24
2	5	19	16	35	16	5
3	11	11	11	30	11	5
Average percentage	7	14	17	31	11	12

Adapting the satellite-based delineation to NDVI features

Characterization of the delineation based on yield-maps

Yield map-based delineation was computed using Blackmore's methodology over the six fields having at least three yield maps available (F8, F21, F22, F32, F33 and F38).This initial output was considered as the delineation reference; results are displayed in **Fig. 6 – left panel**.

Among the six fields, yield records from the yield maps varied more spatially than temporally. The annual difference between extreme yields recorded at the field scale was on average 8.9 Mg ha^{-1} while the interannual gap between highest and lowest yield median was on average 3.2 Mg ha^{-1} per field (**Table S1**).

The delineation from yield maps (**Fig. 6 – left panel**) showed scattered management zones. This noise was explained by unequal quality in yield maps records, despite the filtering process (**Table 2**): isolated pixels corresponding to a different management zone than the surrounding area represent between one to twenty pixels per field (on average thirteen pixels). Globally speaking, on average over the six fields, 50 % of the surface was occupied by the "HS" management zone, 48% by the zone "LS" and 2% by the zone "U" (**Table S2**).

Assessing the accuracy of the satellite-based delineation methods

Fig. 7 displays for each field, the accuracies computed when varying the NDVI value threshold used in Blackmore's methodology, with variations around the mean (left panel) or the median (right panel) NDVI used as reference threshold. **Table 8** reports, among else, accuracies recorded when using the mean and the median NDVI as threshold in Blackmore's methodology. Accuracies retrieved from both ranges of variations around the mean and median NDVI ranged in similar intervals, between 63% and 79% from the mean and 65% to 83% from the median (**Fig 7 | Table 8**). On average, 69.4% of fields pixels were correctly classified when using the mean NDVI as threshold in the methodology (**Table 8**). Fields F8 and F33 obtained the lowest accuracy (resp. 63% and 67%) and concomitantly had the lowest surface. F22 exhibited the highest accuracy at mean NDVI, with a value of 79.8 %.

Maximal accuracies obtained from variations around the mean threshold were mostly observed for different increments [-0.04 ; 0.04], depending upon the field considered (Fig. 6).

Average accuracy computed from the median NDVI reached 71.2% (table 7). When accuracies were computed from variations around the median NDVI thresholds, maximum accuracies were observed five times out of six at the exact median NDVI, where accuracies ranged between 65% and 83% (**Fig 7** and **Table 8**). For field F38, the maximum accuracy (77%) was obtained with a shift of 0.01 from median value (**Fig 7**).

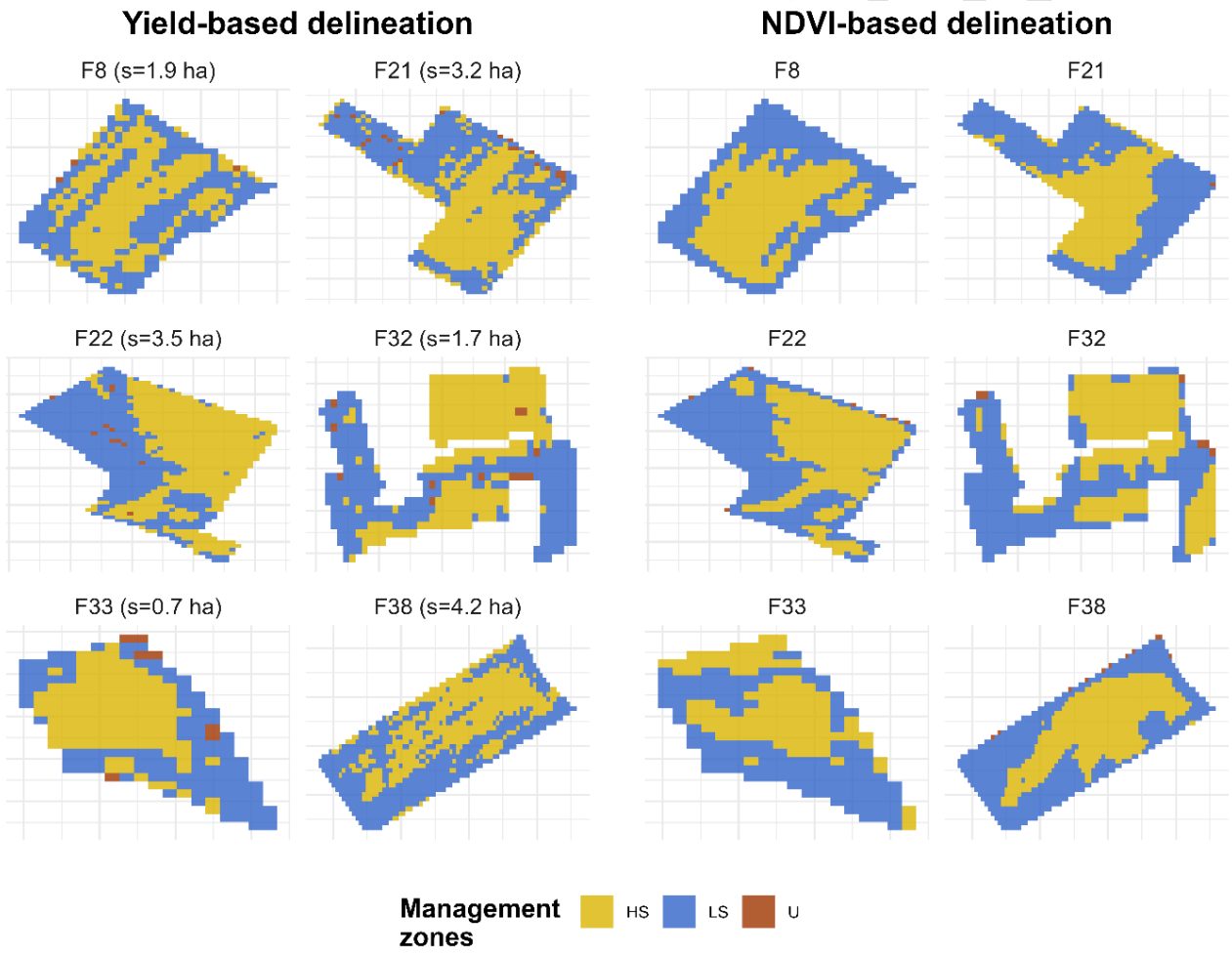

Fig. 6: Comparison of yield map-based delineation and adapted NDVI-based delineation using the median NDVI as a threshold within Blackmore's methodology.

Overall, using median NDVI as threshold in the methodology allowed to systematically reach the maximum accuracy. In the frame of the deploying Blackmore's delineation at territorial scale, this

inherently made of such threshold a more robust estimator than the use of mean NDVI. No relationship was found between the accuracy levels and the field surfaces on the other fields (**Table 8 | Fig. 7**).

Table 8: Metrics of NDVI-based delineation accuracy computed on each field according to (1) the mean threshold and (2) the median threshold

	Fields	Accuracy	Sensitivity			Specificity		
			HS	LS	U	HS	LS	U
Mean threshold	F8	0.632	0.825	0.447	0.000	0.452	0.817	1.000
	F21	0.693	0.773	0.643	0.000	0.645	0.758	0.998
	F22	0.799	0.887	0.717	0.000	0.719	0.881	0.997
	F32	0.691	0.749	0.665	0.000	0.691	0.729	0.987
	F33	0.671	0.759	0.624	0.000	0.595	0.765	1.000
	F38	0.677	0.850	0.532	0.000	0.546	0.866	0.985
	Average	0.694	0.807	0.605	0.000	0.608	0.803	0.994
Median threshold	F8	0.647	0.647	0.653	0.000	0.656	0.641	1.000
	F21	0.725	0.660	0.820	0.000	0.821	0.640	0.999
	F22	0.830	0.791	0.889	0.091	0.889	0.790	0.995
	F32	0.689	0.687	0.723	0.000	0.742	0.670	0.989
	F33	0.649	0.607	0.734	0.000	0.716	0.605	1.000
	F38	0.755	0.735	0.772	0.000	0.777	0.746	0.992
	Average	0.716	0.688	0.765	0.015	0.767	0.682	0.996

Diagnosis of the improved satellite-based approach to perform zone management

Complementary to the global accuracies, **Table 8** displays metrics related to the sensitivity and specificity of each zone. Using the mean or the median as a threshold changed the distribution of management zones. Globally speaking, whatever the threshold (using mean or median values) and its global accuracy level, NDVI-based delineation methods failed to retrieve unstable zones. The sensitivities of U zones were always equal to zero (**Table 8**). Though, both yield map-based and NDVI-based delineations identified similar percentages of unstable zones (specificity close to 1) (**Table 8 | Table S2**).

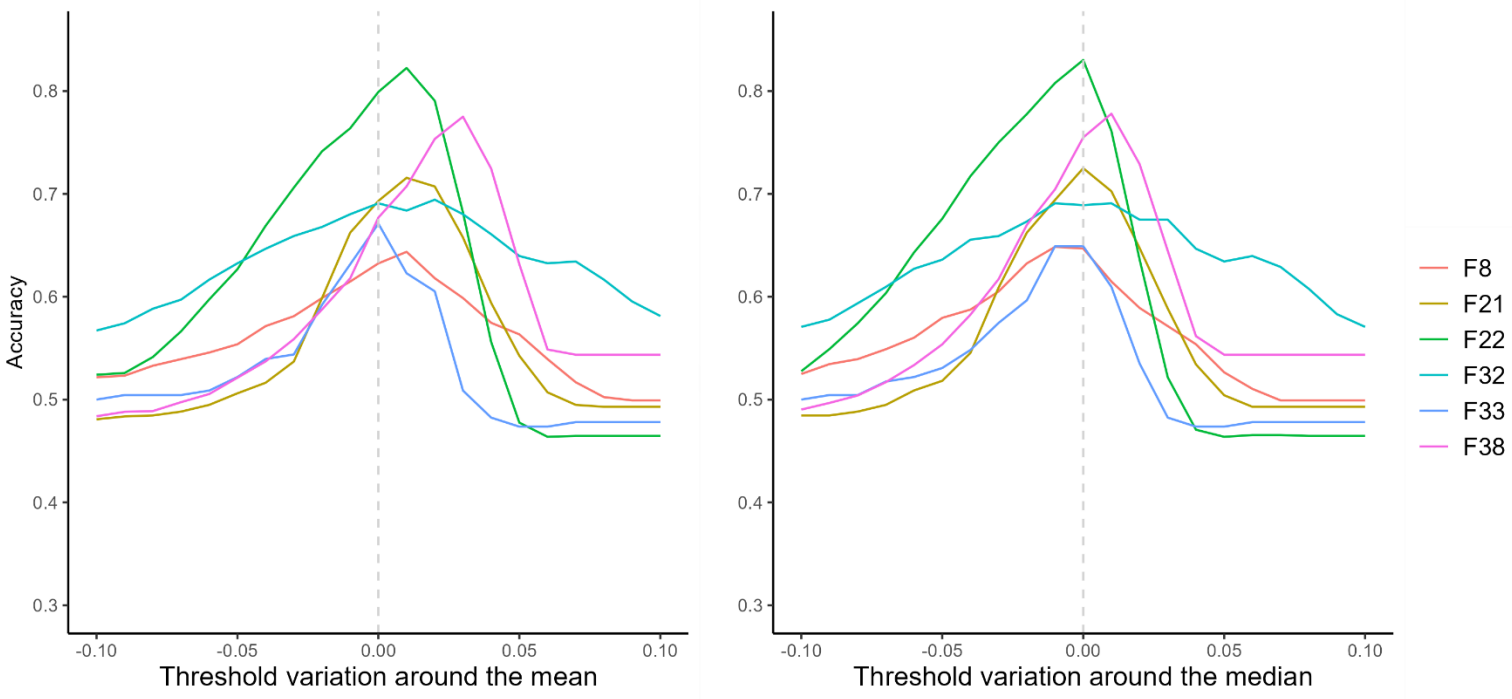

Fig.7: Accuracy variations of NDVI-based delineation when NDVI threshold was incremented from the mean (left panel) or from the median (right panel). Dotted lines are the initial thresholds, the mean NDVI or median NDVI. F8 to F38 refers to the six fields where the delineation was implemented.

Regarding the stable zones, systematic higher sensitivities of HS over LS were observed with the mean threshold, indicating an over-detection of HS at the field scale (**Table 8**). The use of median NDVI as threshold induced a shift in the detection, with sensitivities of LS mostly superior to the ones of HS (**Table 8**).

Discussion

Robustness of NDVI to retrieve annual subfield yield variations on winter wheat

Among the vegetation indices assessed, NDVI, CI_{green} and $CI_{rededge}$ maps obtained the highest correlations with yield maps over the different fields and over the different years (**Fig. 4**). Only few differences of sensitivity were observed between the three indices, both globally (**Fig. 4**) but also along each sensing period (**Fig. S3**). Overall, the correlations levels were consistent with those obtained on NDVI by Kern et al. (2019) and Toscano et al. (2019).

Each of the three indices relied on two wavelengths and have in common the NIR which is known to be sensitive to changes of LAI and biomass (Ollinger, 2011). As the red wavelength of NDVI is responsible of the saturation effect in high LAI, Gitelson et al. (2003) proposed two indices based respectively on green and red-edge wavelengths, namely CI_{green} and $CI_{rededge}$. Vallentin et al. 2022 reported a saturation effect of NDVI early in the season, since the stem elongation growth stage. Revill et al. 2019 reported the highest sensitivity of red-edge or green, combined with NIR, to detect subfield variations of LAI or chlorophyll content during the wheat growing period. Also, Skakun et al. (2021) reported the interest of these wavelengths for corn yield prediction at the subfield scale. Nonetheless, in the two above-mentioned studies, either red-edge and green were used in the frame of complex approaches, or the related VIs were computed from different sensors having finer resolutions.

Our results did not completely align with these results. In the high yielding context of the current study, it was reasonable to expect a saturation effect in NDVI's ability to detect subfield yield variations. However, implementing CI_{green} or $CI_{rededge}$ provided only slight added value compare to NDVI and this greater performance was not systematic; both CI_{green} and $CI_{redegde}$ underperformed NDVI in 2021 for example. . Furthermore, no conclusion can be made on the saturation effect of NDVI as the three indices obtained similar ranges of correlations all along the different seasons (**Fig. S3**). Still, it must be noted that each index was not computed at the same spatial resolution. While NDVI and CI_{green} were calculated at 10 m resolution, $CI_{rededge}$ was calculated with 20 m resolution wavelengths, inherent to sentinel-2 sensors. This could affect the sensitivity of $CI_{rededge}$ as reported by Vallentin et al. 2022. In

this study, only a simple cubic resampling approach was used to downsample 20 m resolutions bands to 10 m resolution (Schleffer et al. 2017). However, $CI_{rededge}$ might have benefited of more complex downsampling techniques as the one suggested in Kowalecko et al. 2023. Therefore, when using exclusively Sentinel-2 wavelengths and simple statistical approaches, NDVI remained a competitive index to retrieve subfield yield variations as confirmed by Vallentin et al. 2022.

Sensing window that optimally retrieve the subfield yield variations

The highest correlations and maximal spatial autocorrelation between yield maps and maps of NDVI were both optimally expressed at the end of May while high correlations were also found from the end of April and the beginning of May (**Table 6, 7, Fig. S2**).

In our study, yield maps were acquired a posteriori and were assembled from different farms. Unfortunately, precise growth stages records were not available. However, in Belgium, usual winter-wheat growing stage are reported as follows: Stem elongation stage usually occurs by the end of April, flag leaf stage usually occurs by the end of May (Dumont et al., 2015a, 2016) and flowering period is usually observed before mid-June (Gobin, 2018).

Marti et al. (2006) and Vallentin et al. (2022) observed a higher ability of NDVI to predict the yield level from the flowering stage and during later growth stages. Yet, in this study, more contrasted results were obtained in June; Maximal correlations remained in the same level as those observed in May but median correlations were largely lower among the different fields and years (**Table 6, Fig. S2**). Marti et al. (2006) still indicated significant correlations from stem elongation up to flowering stage. The final yield is known to be closely related to the number of grains per surface unit (Miralles and Slafer, 2007), which is also related to changes in biomass occurring few weeks before the flowering stage (Fischer, 1985). In this dataset, the biomass heterogeneity at the subfield scale and its consequences on the final yield variability might explain the higher correlation observed between NDVI maps and yield maps at the beginning of the flowering stage. This assumption is consistent with findings reported by Magney et al. (2016) who observed a positive relationship between the rate of stem elongation and the final

grain yield.

Therefore, it was decided to conserve the period covering the second half of May as the optimal sensing window to retrieve subfield yield variations from NDVI. This sensing window appeared as a best compromise between the maximal expressions of biomass heterogeneity, resulting from the stem elongation phase, and the reported risk of a lack of sensitivity in spatial variations observed in June (partly due to unavailable satellite images) (**Fig. S1**).

A simple use of NDVI to delineate management zones

Correlations levels observed in this study between yield maps and maps of NDVI , while consistent with different research (Kern et al., 2019; Toscano et al. 2019; Vallentin et al. 2022), remained relatively low. Indeed, images sensed during the last two weeks of May obtained on average a median correlation in the range 0.27 to 0.40, with a large disparity among fields and years (**Table 6**).

Considering the unexplained part of yield variability, it was reasonable to question the ability of NDVI to accurately detect management zones. Unlike most studies, that aim at predicting yield on a subfield scale (Bukowiecki et al., 2021; Delloye et al., 2018; Hunt et al., 2019; Vallentin et al., 2022), this study proposes to directly use NDVI to retrieve the subfield variations and try to distinguish high from low yield zones using an adaptation of Blackmore's original methodology (see Materiel and Method). A "corrective effect" can be expected when integrating NDVI maps into a pluri-annual analysis as the one proposed by Blackmore. This correction would increase the ability of the vegetation index to distinguish high and low-yielding parts of the field compare to annual correlation between yield and NDVI maps.

Using the proposed approach (mean NDVI), 69% of the pixels were correctly classified in the different management zones compare to the original yield map-based delineation (**Table 8 |upper part**).

Among else, the misclassified pixels were related to the inability to detect unstable zones. However, unstable zones obtained from yield map delineations were spatially inconsistent as they were dispersed over the field (mostly isolated pixels) and usually not contiguous (Fig. 6). This aspect is

546 further discussed in the following section.

547 Secondly, pixels from the estimated HS were globally better classified than their equivalent from LS
548 (**Table 8 |upper part**), as reported by the higher sensitivities. Overall, this sensitivity imbalance
549 between HS and LS was related to an overprediction of the surface occupied by HS, at the expense of
550 LS. Consequently, estimated HS pixels were more often correctly classified in HS as the number of
551 estimated HS pixels was simply greater than the HS reference. In that way, the number of pixels
552 correctly not classified in HS was lower for HS than for LS (Table. S2).

553 At this stage, the overprediction of HS indicated the unsuitability of the delineation algorithm to
554 correctly distinguish high from low yield zones at the annual scale. The relationship between yield and
555 NDVI is known to be not linear as reported by Baret (2016). The example provided **Fig. 3** illustrate the
556 discrepancies between both yield and NDVI distributions; NDVI follows a Beta distribution while yield
557 is normally distributed (Stoy et al. 2022). In the delineation algorithm proposed and applied to NDVI
558 values, the substitution of the mean by the median NDVI as threshold globally improved the overall
559 accuracy of the delineation (**Fig. 7, Fig S4 and Table 8**). More importantly, the median systematically
560 led to the highest prediction accuracy on each of the six fields. The robustness of the delineation
561 algorithm is a major issue to implement the delineation at larger territorial scale. Such improvement
562 to the methodology resulted in a shift of higher sensitivity from HS to LS zones and an improved
563 specificity of HS at the expense of LS zones.

564 Yet, it has to be noticed that median NDVI based delineation slightly underestimate field performances
565 as it tends to over-represent LS zones over the whole fields (**Fig. 6**). From an environmental
566 perspective, this might be an advantage, as it will globally lead to lower intensity in the management
567 practices (e.g. fertilization recommendations) within those zones.

568

569 **Reliability of the relationship between yield maps and maps of VIs**

570 NDVI-based delineation succeeded to retrieve 71% of the yield map-based delineation. This result is
571 globally very satisfying, in comparison to what was expected from the relationship between yield and

NDVI maps. However, the residual imprecision was as much related to the lack of sensitivity of NDVI to capture the subfield yield variation than to the precision of the different yield maps.

The two steps filtering algorithm used ahead of the analysis to reduce the noise of yield maps was systematically applied on each of the yield maps. If the algorithm significantly reduced yield maps errors (**Table 2**), the initial quality of the different maps was unequal. Fields F32 and F33 illustrate this limit. Belonging to the same farm and harvested with the same equipment, the two fields obtained largely different correlations at identical sensing dates in 2020. Field size differences would partly explain the differences of yield map quality as reported by Arslan and Colvin (2002) as the yield maps errors are mainly observed at the beginning and the end of harvested rows. In this dataset, the two smallest fields resulted in the less accurate delineation while no linear trend was observed between the size of the field and the accuracy of satellite-based delineation on the other plots (Table 8| Fig.7)

The overall accuracy was also affected by the inability to detect unstable zones. Unstable zones retrieved from yield map-based delineation represented a very low proportion of pixels (**Fig. 6**) while it represented a major originality of Blackmore's method (Blackmore 2000). Given the respective areas concerned, unstable parts of each field were likely to represent yield maps residual errors than truly unstable yields. A specific investigation would be required to reset the stability/instability threshold which was arbitrarily set by Blackmore himself in the initial approach (Blackmore 2000) and later discussed in Blackmore et al. (2003). Adapting the stability threshold to an interannual yield "tolerance threshold" as it is considered in Blackmore et al. 2003 and Basso et al. 2007 seems an interesting solution. However, such a threshold should also take into account the magnitude of interannual yield variations (Ray et al. 2015). In the frame of experimental plots, the yield stability issue is further analyzed in Reckling et al. 2021, but it goes beyond the scope of this study and it was decided to cope with the suggested threshold. A supplementary figure is proposed to visualize the distributions of the temporal coefficient of variation (Eq 10) used to dissociate stable and unstable parts of the field (Fig. S5). Even though it might have an impact on the robustness of these results. Such delineation would also benefit from more complex approaches linking yield values to vegetation indices, such as NDVI,

or directly linking yield to satellite spectral bands. Increasing attention is given to supervised machine learning approaches while improved yield predictions are reported (Revill et al. 2019; Darra et al. 2023). These methods would allow a higher precision in detecting subfield spatial variations (Skakun et al. 2021)

Finally, the high spatial variability of yield maps, partly due to high spatial resolution and low attribute accuracy, led to scattered management zones (not limited to unstable parts as discussed previously). On the contrary, NDVI-based delineation resulted in spatially homogeneous management zones and, in the end, a higher reliability of the resulting delineation is expected (Fig. 6).

Conclusion

Site-specific management is receiving increasing attention due to its economic and environmental interests. Delineation of management zones, such as the one based on yield maps, is a promising but challenging tool that requires well-tuned farm equipment and specific knowledges for data management. Recent availability of high-resolution and more frequent satellite images from Sentinel-2 mission offers new opportunities for simplifying and automatizing the delineation process.

This study showed the potential to delineate management zones from Sentinel-2 images for a winter wheat crop, provided that (1) the derived vegetation index -such as NDVI - was able to capture sub-field variations, and that (2) images sensed at the beginning of the winter wheat flowering stage were used.

The integration of pluri-annual maps, in approaches such as the one proposed by Blackmore in 2000, allows to capture temporal and spatial variations of yields at subfield scale and classify fields in zones according to their potential. Despite variable relationship reported at the annual scale between NDVI and yield maps, the integration of multi-year NDVI maps has been proven to capture yield subfield variations with an accuracy of 69% in comparison to the Blackmore's methodology. Also, the

suggestion to adapt the original delineation method to the shape of NDVI distribution, using median NDVI, increased the accuracy of classification (71%) and, above all, improved the robustness of the delineation, specifically in its ability to differentiate high from low yielding zones.

We believe that this approach, if deployed on a large scale such as a territory, has the potential to support farmers in the adaptation of their practices toward more efficient crop management.

CRediT authorship contribution statement

Arthur Lenoir: Conceptualization, Methodology, Data analysis, Visualization, Writing – original draft. **Bertrand Vandoorne**: Conceptualization, Funding acquisition, Supervision. **Ali Siah**: Supervision. **Benjamin Dumont**: Conceptualization, Methodology, Supervision, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References :

Agreste 2019. Agreste Hauts-de-France – Etudes et analyses (Studies and analyses) – n° 36. Novembre 2019. [https://draaf.hauts-de-france.agriculture.gouv.fr/IMG/pdf/Etudes_et_analyses_No36_-_fertilisation_cle87ac22.pdf] (Last accessed 28 June 2023)

Agreste 2020. Memento 2020 – Hauts-de-France. Agreste, La statistique, l'évaluation et la prospective agricole (The agricultural statistic, evaluation and prospective) and [https://draaf.hauts-de-france.agriculture.gouv.fr/IMG/pdf/Memento2020-Productions-vegetales_cle8b64ff.pdf] (Last accessed 28 June 2023)

Alexander, P., Arneth, A., Henry, R., Maire, J., Rabin, S., Rounsevell, M.D.A., 2023. High energy and fertilizer prices are more damaging than food export curtailment from Ukraine and Russia for food prices, health and the environment. *Nature Food* 4, 84–95. https://doi.org/10.1038/s43016-022-00659-9

Anselin, L., Syabri, I., Kho, Y., 2010. GeoDa: An Introduction to Spatial Data Analysis, in: Fischer, M.M., Getis, A. (Eds.), Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Springer, Berlin, Heidelberg, pp. 73–89. https://doi.org/10.1007/978-3-642-03647-7_5Arslan, S., Colvin, T.S., 2002. Grain yield mapping: Yield sensing, yield reconstruction, and errors. *Precision Agriculture* 3, 135–154.

Azzari, G., Jain, M., Lobell, D.B., 2017. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sensing of the Environment*. 202, 129–141. https://doi.org/10.1016/j.rse.2017.04.014

Baret, F., 2016. 2 - Estimation of Biophysical Variables from Satellite Observations, in: Baghdadi, N., Zribi, M. (Eds.), *Land Surface Remote Sensing in Agriculture and Forest*. Elsevier, pp. 37–80. https://doi.org/10.1016/B978-1-78548-103-1.50002-9

Basso, B., Bertocco, M., Sartori, L., Martin, E.C., 2007. Analyzing the effects of climate variability on spatial pattern of yield in a maize–wheat–soybean rotation. European Journal of Agronomy 26, 82–91. https://doi.org/10.1016/j.eja.2006.08.008

Basso, B., Ritchie, J.T., Cammarano, D., Sartori, L., 2011. A strategic and tactical management approach to select optimal N fertilizer rates for wheat in a spatially variable field. *European Journal of Agronomy*. 35, 215–222. https://doi.org/10.1016/j.eja.2011.06.004

Basso, B., Cammarano, D., Fiorentino, C., Ritchie, J.T., 2013. Wheat yield response to spatially variable nitrogen fertilizer in Mediterranean environment. *European Journal of Agronomy*. 51, 65–70. https://doi.org/10.1016/j.eja.2013.06.007

Basso, B., Dumont, B., Cammarano, D., Pezzuolo, A., Marinello, F., Sartori, L., 2016. Environmental and economic benefits of variable rate nitrogen fertilization in a nitrate vulnerable zone. *Science of The Total Environment* 545–546, 227–235. https://doi.org/10.1016/j.scitotenv.2015.12.104

Blackmore, S., 2000. The interpretation of trends from multiple yield maps. *Computers and Electronics in Agriculture*. 26, 37–51. https://doi.org/10.1016/S0168-1699(99)00075-7

Blackmore, S., Godwin, R.J., Fountas, S., 2003. The Analysis of Spatial and Temporal Trends in Yield Map Data over Six Years. Biosystems Engineering 84, 455–466. https://doi.org/10.1016/S1537-5110(03)00038-2

Bukowiecki, J., Rose, T., Kage, H., 2021. Sentinel-2 Data for Precision Agriculture?—A UAV-Based Assessment. *Sensors* 21, 2861. https://doi.org/10.3390/s21082861

Darra, Nicoleta, Evangelos Anastasiou, Olga Kriezi, Erato Lazarou, Dionissios Kalivas, and Spyros Fountas. 2023. "Can Yield Prediction Be Fully Digitilized? A Systematic Review" *Agronomy* 13, no. 9: 2441. https://doi.org/10.3390/agronomy13092441

Daughtry C.S.T, Walthall C.L., Kim M.S., Brown de Colstoun E. and McMurtrey E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*. 74: 229-239.

Delin, S., Lindén, B., Berglund, K., 2005. Yield and protein response to fertilizer nitrogen in different parts of a cereal field: potential of site-specific fertilization. *European Journal of Agronomy*. 22, 325–336. https://doi.org/10.1016/j.eja.2004.05.001

Delloye, C., Weiss, M., Defourny, P., 2018. Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. *Remote Sensing of Environment*. 216, 245–261. https://doi.org/10.1016/j.rse.2018.06.037

Diacono, M., Rubino, P. & Montemurro, F. Precision nitrogen management of wheat. A review. *Agron. Sustain. Dev.* **33**, 219–241 (2013). https://doi.org/10.1007/s13593-012-0111-z

Doraiswamy, P.C., Moulin, S., Cook, P.W., Stern, A., 2003. Crop Yield Assessment from Remote Sensing. *Photogrammetric Engineering & Remote Sensing*. 69, 665–674. https://doi.org/10.14358/PERS.69.6.665

Dumont, B., Basso, B., Bodson, B., Destain, J.-P., Destain, M.-F., 2015. Climatic risk assessment to improve nitrogen fertilisation recommendations: A strategic crop model-based approach. *European Journal of Agronomy* 65, 10–17. https://doi.org/10.1016/j.eja.2015.01.003

EAW, 2022. Froment d'hiver : produits, charges et marge brute. Etat de l'Agriculture Wallonne. (Winter wheat : products, inputs, and gross margin. State of Walloon Agriculture) [https ://etat-agriculture.wallonie.be/contents/indicatorsheets/EAW-C_III_b_2.html#] (Last accessed : 28 June 2023)

Fang, H., Baret, F., Plummer, S., Schaepman-Strub, G., 2019. An Overview of Global Leaf Area Index (LAI): Methods, Products, Validation, and Applications. *Review of Geophysics*. 57, 739–799. https://doi.org/10.1029/2018RG000608

Fowler, D., Coyle, M., Skiba, U., Sutton, M.A., Cape, J.N., Reis, S., *et al.* , 2013. The global nitrogen cycle in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 368, 20130164. https://doi.org/10.1098/rstb.2013.0164

Gastal, F., Lemaire, G., Durand, J.-L., Louarn, G., 2015. Quantifying crop responses to nitrogen and avenues to improve nitrogen-use efficiency, in: *Crop Physiology*. Elsevier, pp. 161–206.

Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B., 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies: Remote estimation of Leaf Area Index. *Geophysical Research Letters*. 30, https://doi.org/10.1029/2002GL016450

Gobin, A., 2018. Weather related risks in Belgian arable agriculture. *Agricultural Systems*. 159, 225–236. https://doi.org/10.1016/j.agsy.2017.06.009

Haboudane, D., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*. 90, 337–352. https://doi.org/10.1016/j.rse.2003.12.013

Hagolle, O., Huc, M., Villa Pascual, D., Dedieu, G., 2015. A Multi-Temporal and Multi-Spectral Method to Estimate Aerosol Optical Thickness over Land, for the Atmospheric Correction of FormoSat-2, LandSat, VENµS and Sentinel-2 Images. *Remote Sensing*. 7, 2668–2691. https://doi.org/10.3390/rs70302668

Hatfield, J.L., Gitelson, A.A., Schepers, J.S., Walthall, C.L., 2008. Application of Spectral Remote Sensing for Agronomic Decisions. *Agronomy Journal*. 100, S-117-S-131. https://doi.org/10.2134/agronj2006.0370c

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric

and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment.* 83, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W., Rowland, C.S., 2019. High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment*. 233, 111410. https://doi.org/10.1016/j.rse.2019.111410

IRM (Meteorological RoyalInstitut), 2021. Bilans climatologiques de 2021. [https://www.meteo.be/en/weather/observations/belgium] (Last accessed 21 March 2023)

Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P., Lichtenberger, J., 2018. Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. *Agricultural and Forest Meteorology* 260–261, 300–320. https://doi.org/10.1016/j.agrformet.2018.06.009

Kowaleczko, P., Tarasiewicz, T., Ziaja, M., Kostrzewa, D., Nalepa, J., Rokita, P., Kawulok, M., 2023. A Real-World Benchmark for Sentinel-2 Multi-Image Super-Resolution. Sci Data 10, 644. https://doi.org/10.1038/s41597-023-02538-9

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*. 28, 1–26. https://doi.org/10.18637/jss.v028.i05

Lachia N, Pichon L, Marcq P., Taylor J.-A. and Tisseyre B., 2021. Why are yield sensors seldom used by farmers? A French case study. *Conference: 13th European Conference on Precision Agriculture*.

Lassaletta, L., Billen, G., Grizzetti, B., Anglade, J., Garnier, J., 2014. 50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland. *Environmental Research Letters*. 9, 105011. https://doi.org/10.1088/1748-9326/9/10/105011

Le Bas C., 2021, "Carte de la profondeur du sol issue de la Base de Données Géographique des Sols de France" (Soil depth map from the Geographical Base of Soils of France), https://doi.org/10.15454/7ZDND6, Data Hub INRAE, V1

Leroux, C., Jones, H., Clenet, A., Dreux, B., Becu, M., Tisseyre, B., (2018). A general method to filter out defective spatial observations from yield mapping datasets. *Precision Agriculture* 19, 789–808. https://doi.org/10.1007/s11119-017-9555-0

Leroux C. (2020). R codes to be used within QGIS for Precision Agriculture applications. https://github.com/Corentin39-aspexit/R-QGIS-Precision-Agriculture/blob/master/R_codes/ [Last accessed December 2021]

Lyle, G., Bryan, B.A., Ostendorf, B., 2014. Post-processing methods to eliminate erroneous grain yield measurements: review and directions for future development. *Precision Agriculture* 15, 377–402. https://doi.org/10.1007/s11119-013-9336-3

Machet, J.-M., Dubrulle, P., Damay, N., Duval, R., Julien, J.-L., Recous, S., 2017. A Dynamic Decision-Making Tool for Calculating the Optimal Rates of N Application for 40 Annual Crops While Minimising the Residual Level of Mineral N at Harvest. *Agronomy* 7, 73. https://doi.org/10.3390/agronomy7040073

Maestrini, B., Basso, B., 2018. Drivers of within-field spatial and temporal variability of crop yield across the US Midwest. *Scientific Reports*. 8, 14833. https://doi.org/10.1038/s41598-018-32779-3

Magney, T.S., Eitel, J.U.H., Huggins, D.R., Vierling, L.A., 2016. Proximal NDVI derived phenology improves in-season predictions of wheat quantity and quality. *Agricultural and Forest Meteorology*. 217, 46–60. https://doi.org/10.1016/j.agrformet.2015.11.009

Makowski, D., Tichit, M., Guichard, L., Van Keulen, H., Beaudoin, N., 2009. Measuring the accuracy of agro-environmental indicators. *Journal of Environmental Management*. 90, S139–S146. https://doi.org/10.1016/j.jenvman.2008.11.023

Marti, J., Bort, J., Slafer, G. a., Araus, J. l., 2007. Can wheat yield be assessed by early measurements of Normalized Difference Vegetation Index? *Annals of Applied Biology*. 150, 253–257. https://doi.org/10.1111/j.1744-7348.2007.00126.x

Meynard, J.M., Aubry, C., Justes, E., Le Bail, M. (1997). Nitrogen Diagnosis and Decision Support. In: Lemaire, G. (eds) *Diagnosis of the Nitrogen Status in Crops*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-60684-7_9

Miralles, D.J., Slafer, G.A., 2007. Sink limitations to yield in wheat: how could it be reduced? *Journal of Agricultural Science*, 145, 139–149. https://doi:10.1017/S0021859607006752

Ollinger, S.V., 2011. Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist*. 189, 375–394. https://doi.org/10.1111/j.1469-8137.2010.03536.x

Panek, E., Gozdowski, D., 2020. Analysis of relationship between cereal yield and NDVI for selected regions of Central Europe based on MODIS satellite data. *Remote Sensing Application : Society and Environment*. 17, 100286. https://doi.org/10.1016/j.rsase.2019.100286

Pôle Theia, 2022. Value-Adding Products and Algorithms for Land Surfaces. Retrieved from https://www.theia-land.fr/en/homepage-en/ [Accessed January 2022].

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote sensing of environment 48, 119–126.

Ravier, Clémence, Jeuffroy, M.H., Meynard, Jean-Marc, 2016. Mismatch between a science-based decision tool and its use: The case of the balance-sheet method for nitrogen fertilization in France | Elsevier Enhanced Reader. *NJAS - Wageningen Journal of Life Sciences* 79:1, 31–40. https://doi.org/10.1016/j.njas.2016.10.001

Ray, D.K., Gerber, J.S., MacDonald, G.K., West, P.C., 2015. Climate variation explains a third of global crop yield variability. *Nature Communications*. 6, 5989. https://doi.org/10.1038/ncomms6989

Reckling, M., Ahrends, H., Chen, T.-W., Eugster, W., Hadasch, S., Knapp, S., Laidig, F., Linstädter, A., Macholdt, J., Piepho, H.-P., Schiffers, K., Döring, T.F., 2021. Methods of yield stability analysis in long-term field experiments. A review. Agron. Sustain. Dev. 41, 27. https://doi.org/10.1007/s13593-021-00681-4

Revill, A., Florence, A., MacArthur, A., Hoad, S., Rees, R., Williams, M., 2020. Quantifying Uncertainty and Bridging the Scaling Gap in the Retrieval of Leaf Area Index by Coupling Sentinel-2 and UAV Observations. *Remote Sensing*. 12, 1843. https://doi.org/10.3390/rs12111843

Revill, A., Florence, A., MacArthur, A., Hoad, S.P., Rees, R.M., Williams, M., 2019. The Value of Sentinel-2 Spectral Bands for the Assessment of Winter Wheat Growth and Development. *Remote Sensing*. 11, 2050. https://doi.org/10.3390/rs11172050

Scheffler, D., Hollstein, A., Diedrich, H., Segl, K., Hostert, P., 2017. AROSICS: An Automated and Robust Open-Source Image Co-Registration Software for Multi-Sensor Satellite Data. Remote Sensing 9, 676. https://doi.org/10.3390/rs9070676

Schulte-Uebbing, L.F., Beusen, A.H.W., Bouwman, A.F., de Vries, W., 2022. From planetary to regional boundaries for agricultural nitrogen pollution. *Nature* 610, 507–512. https://doi.org/10.1038/s41586-022-05158-2

Skakun, S., Kalecinski, N.I., Brown, M.G.L., Johnson, D.M., Vermote, E.F., Roger, J.-C., Franch, B., 2021. Assessing within-Field Corn and Soybean Yield Variability from WorldView-3, Planet, Sentinel-2, and Landsat 8 Satellite Imagery. *Remote Sensing*. 13, 872. https://doi.org/10.3390/rs13050872

SPW 2022. Productions Végétales. [WWW Document]. URL: https://etat-agriculture.wallonie.be/contents/indicatorsheets/EAW2.html# [Accessed April 2022]

Stoy, P.C., Khan, A.M., Wipf, A., Silverman, N., Powell, S.L., 2022. The spatial variability of NDVI within a

821 wheat field: Information content and implications for yield and grain protein monitoring. *PLoS*
822 *ONE* 17, e0265243. https://doi.org/10.1371/journal.pone.0265243
823 Tharwat, A., 2021. Classification assessment methods. *Applied Computing and Informatics* 17, 168–192.
824 https://doi.org/10.1016/j.aci.2018.08.003
825 Toscano, P., Castrignanò, A., Di Gennaro, S.F., Vonella, A.V., Ventrella, D., Matese, A., 2019. A Precision
826 Agriculture Approach for Durum Wheat Yield Assessment Using Remote Sensing Data and Yield
827 Mapping. *Agronomy* 9, 437. https://doi.org/10.3390/agronomy9080437
828 Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote*
829 *Sensing of Environment* 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0
830 Vallentin, C., Harfenmeister, K., Itzerott, S., Kleinschmit, B., Conrad, C., Spengler, D., 2022. Suitability
831 of satellite remote sensing data for yield estimation in northeast Germany. *Precision*
832 *Agriculture*. 23, 52–82. https://doi.org/10.1007/s11119-021-09827-6
833 Vannoppen, A., Gobin, A., 2021. Estimating Farm Wheat Yields from NDVI and Meteorological Data.
834 *Agronomy* 11, 946. https://doi.org/10.3390/agronomy11050946
835 Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S. *et al.* 2019. Food in the
836 Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems.
837 *The Lancet* 393, 447–492. https://doi.org/10.1016/S0140-6736(18)31788-4
838 Zhang, X., Davidson, E., Mauzerall, D. et al. Managing nitrogen for sustainable development. *Nature*
839 528, 51–59 (2015). https://doi.org/10.1038/nature15743