AJLP & GS African Journal on Land Policy and Geospatial Sciences

e-ISSN : 2657-2664, Vol. 6, Issue 1, January 2023

https://revues.imist.ma/index.php/AJLP-GS/index

https://doi.org/10.48346/IMIST.PRSM/ajlp-gs.v6i1.35303 *Category of the manuscript : Articles*

***Received in**: 25 October 2022* **Revised in:** 27 December 2022 ***Accepted in**: 28 December 2022*

MODELING WHEAT YIELD BY USING PHENOLOGYCAL METRICS DERIVED FROM SENTINEL2 IN ARID AND SEMI-ARID REGIONS
A case study in MOROCCO

[1]Adra Idrissi, [2] Abdelaziz Htitiou ,[3]Samir Nadem, [4]Abdelghani Boudhar, [5] Youssef Lebrini, [6] Tarik Benabdelouahab

[1] Georesources and Environment Laboratory and Remote Sensing team, i_adra2000@yahoo.fr
Faculty of Sciences and Technology, Sultan Moulay Slimane, University Beni Mellal. Morocco

[2] Water Resources Management and Valorization and Remote Sensing team, Faculty of Sciences and Technology, Sultan Moulay Slimane, University BeniMellal

[3] Georesources and Environment Laboratory, Faculty of Sciences and Technology, Sultan Moulay Slimane, University Beni Mellal. Morocco
samirmellal@hotmail.com

[4] Center for Remote Sensing Applications (CRSA) , Mohammed VI , Polytechnic University , Ben Guerir, Morocco
ab.boudhar@usms.ma

[5] Georessources and Environment Laboratory and Remote Sensing team, Faculty of Sciences and Technology, Sultan Moulay Slimane, University BeniMellal

[6] Natural Resources and Environment Department, National Institute of Agronomic Research, Rabat, Morocco
tarik.benabdelouahab@gmail.com

ABSTRACT

Context and background

Wheat is one of the oldest cultivated plants in the world and has always been one of the most important staples for millions of people around the world and , especially in North Africa, and especially in Morocco where wheat occupies an important place in the agricultural sector. Thus, an operational crop production system is needed to help decision makers make early estimates of potential food . availability Yield estimation using remote sensing data has been widely studied, but such information is generally scarce in arid and semi-arid regions such as North Africa, where interannual variations in climatic factors, and spatial variability in particular, are major risks to food security.

Goal and Objectives:

The aim of this study is to develop a model to estimate wheat yield based on phenological metrics derived from SENTINEL-2 NDVI images in order to generalize a spatial model to estimate wheat yields in Morocco's semi-arid conditions

Methodology:

The 10 m NDVI time series was integrated into TIMESAT software to extract wheat phenology-related metrics during the 2018-2019 agricultural season, the period in which ground truth data was collected. Through the multiple stepwise regression method, all phenological metrics were used to predict wheat yield. Moreover, the accuracy and stability of produced models were evaluated using a K-fold cross-validation (K-fold CV) method.

Results:

The results of the obtained models indicated a good linear correlation between predicted yield and field observations (R2 = 0.75 and RMSE of 7.08q/ha). The obtained method could be a good tool for decision makers to orient their actions under different climatic conditions

Keywords

Sentinel-2, Phenological metrics, estimate wheat yield, NDVI, time series

1. INTRODUCTION

Wheat is the second most important cereal crop in the world and is a major source of protein and calories. It is cultivated on more than 240 million hectares with a current production of around 700 million tons (CIMMYT, 1996; GCRAD, 2012; FAO, 2020).

In Morocco, this crop occupies an important position in the national economy and agricultural sector. It covers about 5.3 million hectares and produces an average of 50 million quintals of grain per year (MAPMDREF, 2010). However, national cereal production is strongly influenced by climate change, especially in arid and semi-arid areas, where agricultural production is highly dependent on the amount and spatial-temporal distribution of rainfall. (Lionboui, 2020; Benabdelouahab, 2019).

Thus, spatio-temporal crop yield monitoring systems are widely considered necessary to support agricultural policies, and yield estimating represents an important tool for optimizing crop yield and evaluating crop-area insurance contracts.

Over the last few decades, a wide range of crop yield estimation and prediction techniques have been used with varying degrees of accuracy:

In many countries, especially developing countries such as Morocco, crop yield estimation still relies on traditional approaches based on field data collection from segment samples (crop cutting experiments). These data are often time-consuming to obtain, costly, and subject to large errors due to incomplete ground observations, leading to an uncertain estimate of crop area (Reynolds, 2000).

The use of statistical or agronomic models based on historical weather and production data has been the subject of much research. Indeed, in some countries, weather data have been used to monitor and forecast agricultural production (Andarzian, 2008). However, the lack of continuity of these weather data and the poor spatial distribution of weather stations for a wide variety of crops limit the usefulness of these approaches.

With the development of satellite sensors, the interest in using remote sensing data for crop monitoring has increased. Indeed, remote sensing data can be used for crop monitoring and crop production forecasting because of its ability to provide data in a synoptic manner with greater spatial coverage, potentially on a global scale.

In addition, remote sensing can provide timely (and potentially real-time) objective crop growth data on a relatively small scale and objective crop growth data at a relatively low cost. In this regard, NDVI has long been used to monitor crop conditions and estimate crop yield (Doraiswamy, 2004).

Remotely sensed data can also be used as an input to crop simulation models. Such an approach involves biophysical crop simulation models, which are calibrated and driven by remotely sensed information on crop characteristics throughout the season.

Several studies have shown a strong correlation between the Normalized Difference Vegetation Index (NDVI) and cereal yields in Morocco due to the semi-arid climate. The estimation of cereal production based on this index has been found to be more efficient (Balaghi, Tychon et al. 2008).

Thus, the first system for estimating cereal yields in Morocco, called (CGMS_MA), was implemented based on a combined approach between satellite and meteorological data. The estimates were based on 10-daily NDVI/AVHRR, 10-day rainfall amounts and average monthly air temperatures (Balaghi, Tychon et al. 2008).

However, this system has a managerial aspect and is not adapted to estimate yield at the field level. Because the image used has a lower resolution of 1.1 kilometers at nadir, it's difficult to estimate the yield at the level plot.

Other studies have used MODIS images, which are characterized by high spectral and temporal resolution and data availability since 2000. This product offers an excellent opportunity to estimate wheat yield at the regional level and also allows us to follow spatial evolution (Benabdelouahab, 2019).

Since 2015, and with the introduction of new satellite sensors with spatial resolutions of 10 m or more have the possibility to track fields of small farmers:

In addition, the improved temporal resolution of Sentinel-2 images (5 days) since the beginning of 2017 and of Planet scope (daily) since the end of 2017, now makes it possible to observe the fields of smallholders at a much higher frequency

On the other hand, several studies have shown that plant development, stress and yield capacity are expressed in phenological profiles which are derived from time series of vegetation index (Jayawardhana and Chathurange 2016; Mkhabela, Bullock et al. 2011; Doraiswamy, Hatfield et al. 2004; Quarmby, Milnes et al. 1993; Rasmussen 1992; Rasmussen 1992).

Crop-specific phenological metrics (such as biomass accumulation, peak green and leaf development period) provide important information on agricultural management and monitoring and can be used as an indicator of crop productivity and can be used to improve yield prediction models (Benabdelouahab, Lebrini et al. 2019, Htitiou, Boudhar et al. 2019).

Thus, in this paper, Sentinel 2 derived phenological metrics (A and B) were exploited to improve wheat yield estimation in a semi-arid region of Morocco.

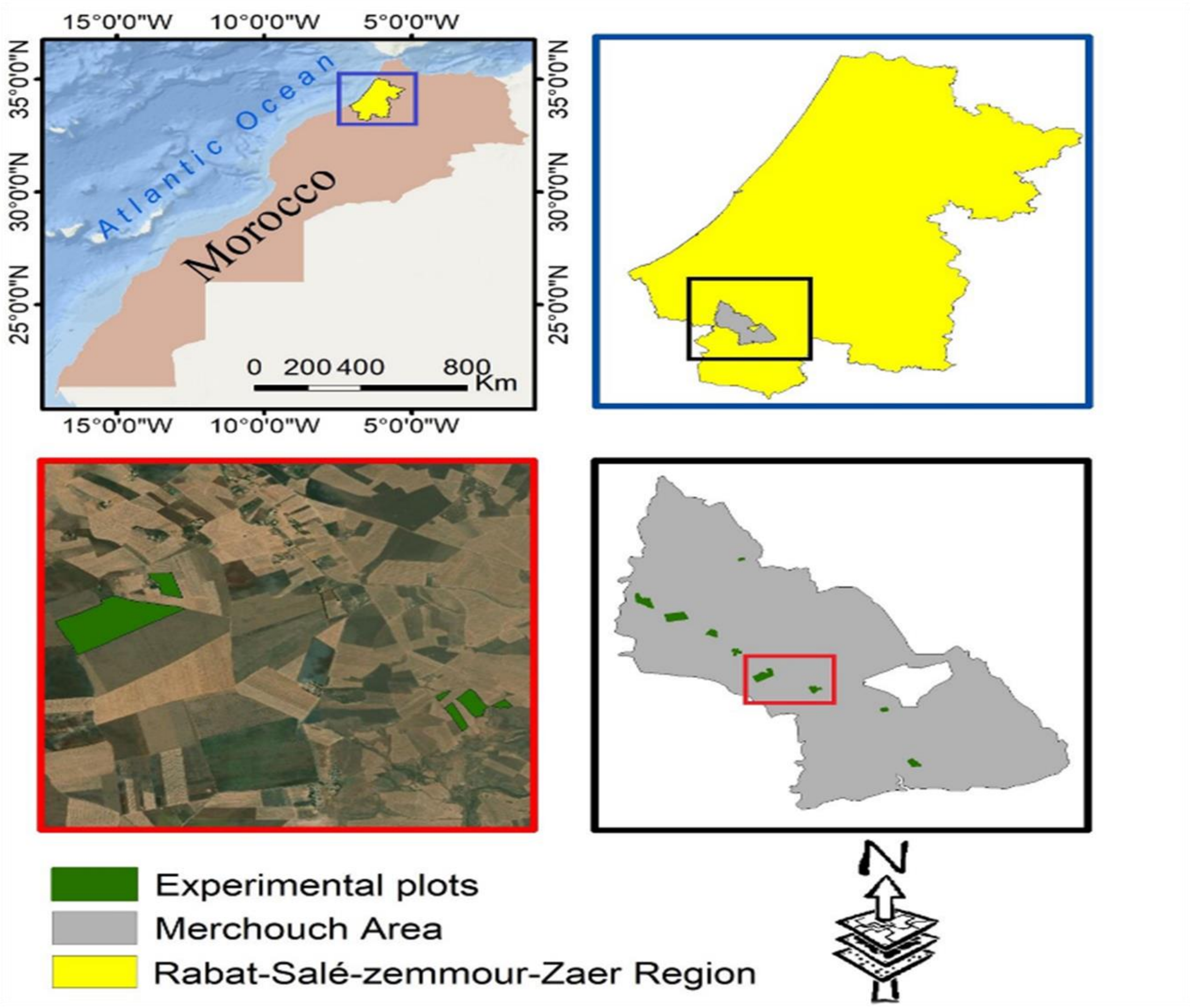
The current study aims to develop a model to estimate wheat yield based on phenological metrics derived from SENTINEL-2 NDVI images in order to generalize a spatial model to estimate wheat yields in Morocco's semi-arid conditions.

2. STUDY SITE AND SATELLITE DATA

2.1 Study Area

The study area took place in the rural commune of Merchouch (MERCHOUCH) belonging to the region of Rabat – Salé-Kenitra in Morocco (Figure 1). It extends between latitudes of 33°26' and 33°40' North and longitudes of 6°44' and 6°30 "West. The Rural Municipality of Merchouch is bounded by SidiBettach town North, the municipality of Ezzhiliga South, the Brachoua commune East, and Had Ghoualem West. The soil in the study area is clay and sub-acid, with low organic matter and potassium content and adequate phosphorous availability. The MERCHOUCH zone is influenced by a semi-arid Mediterranean climate with mean annual temperatures of around 18°C, and minimum monthly temperatures in the range of 10° C to 12°C. The rainfall pattern of the whole area is Mediterranean. Rainfall is highest in November and December and almost zero in July and August, but with some summer showers. The mean annual precipitation is about 300 mm. Finally, according to the rainfall quotient of Emberger, the area is under the influence of a semi-arid climate with warm and temperate winters (Ouharba, El et al. 2019).

Fig. 1. Location of the study area (upper left inset shows the map of Morocco, the study area is on the right and the experimental plot is shown in green)



2.2. Satellite and ground data

- ***Satellite data***

In this study, 62 Sentinel-2 MSI (S2) images, acquired between 01/06/2018 and 01/07/2019, were processed. The Sentinel-2 mission is based on a combination of two satellites, Sentinel-2A and Sentinel-2B, equipped with identical Multispectral Instruments (MSI) capable of collecting multispectral images in 13 bands at different spatial resolutions (between 10 m and 60 m) with a return cycle of five days. The reflectance images used in this study were freely downloaded from the Theia land data center's website (https://www.theia-land.fr/en) for the studied site and period. The Theia center provides S2 corrected products (level 2A) generated with the MAJA software developed in coordination between CNES/CESBIO and DLR. The characteristics of Sentinel 2-A are presented in table1.

Table 1. Characteristics of SENTINEL 2-A (MSI)

Bandes	SENTINEL -2 A(MSI)	
	Spectral range (μm)	Spatial resolution (m)
Coastal/aerosol	0.43-0.45	60
Blue	0.46-0.52	10
Green	0.54-0.58	10
Red	0.65-0.86	10
VRE-1	0.70-0.71	20
VRE-2	0.73-0.74	20
VRE-3	0.77-0.79	20
NIR	0.78-0.90	10
NIR narrow	0.85-0.87	20
Water vapor	0.93-0.95	60
Cirrus	1.37-1.39	60
SWIR-1	1.57-1.66	20
SWIR-2	2.10-2.28	20
Pan	--	--

The NDVI images were generated from downloaded reflectance images based on the near infrared (NIR) and red (RED) spectral bands according to the following equation (Rouse Jr, Haas et al. 1974).

$$\mathbf{NDVI= NIR- RED/NIR+RED} \qquad (1)$$

In this study, NDVI was chosen for monitoring phenological metrics. The choice of this index is based on the fact that it is sensitive to variations in vegetation cover in areas characterized by low vegetation density, unlike other indices such as the EVI (Enhanced Vegetation Index) (Ji, 2007)

- ***Training site (Ground data)***

The training samples used in this study were provided by the department of statistics in the Ministry of Agriculture (DSS). A survey was carried out to estimate the objective yields per municipality. The samples are composed of 5 sample points, called segment-points, and are drawn from the sampling frame using a systematic stratified random sample. The methodology used to identify the sample that will be used for the objective communal performance survey consists of drawing a large number of samples and choosing among them the one that allows making an estimate that is closest to the sampling frame (MAPMEFDR/DSS/DS). We used 123 locations in this study that cover the Merchouch area(Fig1).

3. METHODOLOGY

Firstly, we reconstructed the smoothed NDVI time series from the S2 images using TIMESAT software. Then, we extracted Phenological Metrics from the smoothed curve. Therefore, we exploited the STEEPWISE algorithm to test the different combinations of all variables (PHM).

Then, we developed an empirical model for wheat yield estimation. Finally, the overall accuracy of the model was assessed based on the cross validation procedure.

A workflow of the methodology applied in this study is presented in **Figure2**.

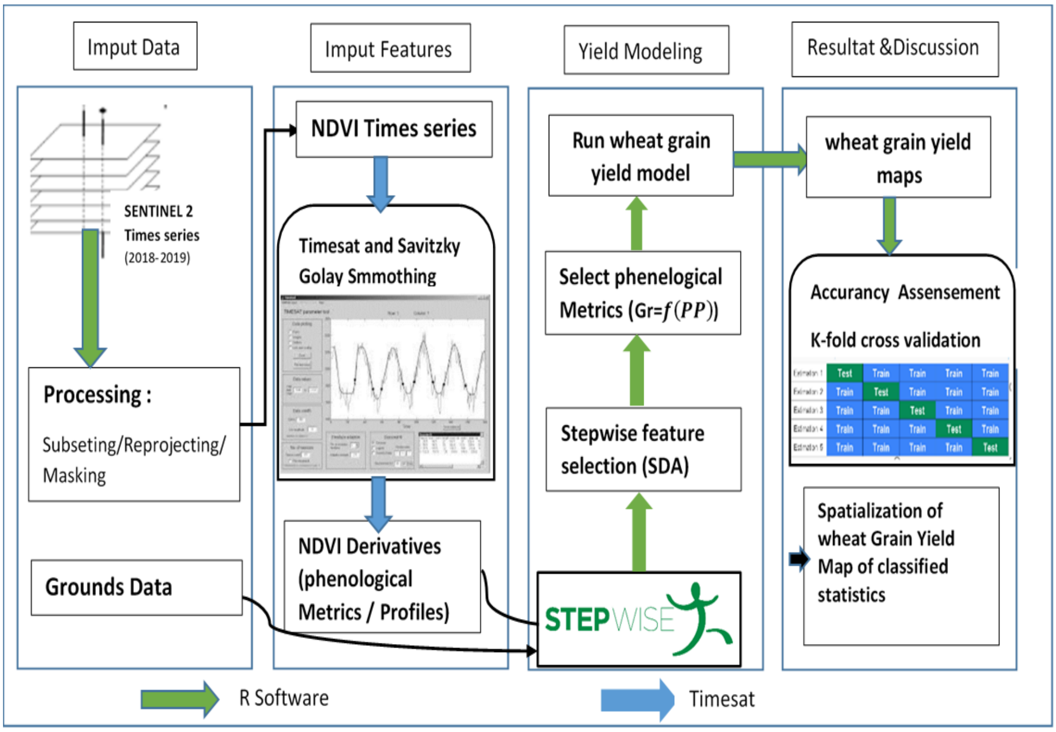


Fig. 2. Workflow of the applied methodology

3.1. Data Time series analysis.

A combination of 62 Sentinel-2 A/B images, acquired between 01/06/2018 and 01/07/2019, were downloaded from their site, then processed by R software and then calculated NDVI of each cell from the two spectral bands red and near infrared, these NDVI images are organized in a time series of 62 NDVI-Sentinels 2 images.

The NDVI values retrieved from satellite images are often subject to noise caused by clouds and atmospheric disturbances. For that reason, various methods have been applied to filter out persistent noise and to get better quality of the NDVI curve. Among these methods, the SavitzkyGolay filter is one of the most widely used methods for NDVI time series curve smoothing due to its ability to maintain the original NDVI profile while eliminating irregular variations and removing invalid noise (Htitiou et al., 2019).

Indeed, Geng, Ma et al. (2014) compared 8 NDVI signal smoothing methods for a study area in China for 3 different products (AVHRR GIMMS, AVHRR Pathfinder, MODIS, and SPOT VEGETATION) and for 5 vegetation types. The results of this study show overall that all the smoothing methods tested are able to correctly convert the temporal profile of NDVI, but with notable differences in the ability of these methods to correctly maintain the NDVI amplitude (difference between the minimum and maximum values) as well as the smoothness. The analysis by Geng, Ma et al. (2014) showed that among the 8 methods tested, the Savitzky-Golay method and the Whittaker method are the most appropriate for obtaining smoothing that best preserves the temporal NDVI profile while eliminating outliers.

The general equation (Equation (2)) of the S – G filter is:

$$g_i = \frac{\sum_{n=-nL}^{N_r} C_n f_{i+n}}{\text{n}} \qquad (2)$$

Where fi represents the original NDVI value in the time series, gi is the filtered NDVI value, and n is the width of the filter window, while nL and nR correspond respectively to the left and right edge of the signal component (Htitiou et al., 2019)

The SG filter used in this study was implemented in TIMESAT software, developed by Eklundh and Jönsson (2015), for processing and smoothing NDVI Image time series data to drive seasonal phenological metrics.

3.2. Extracted Phenological metrics.

After smoothing the NDVI time series with the S-G filter, thirteen phenological metrics values were extracted for each pixel in the studied area over the course of the 2018-2019 season (Figure 1). The values were analyzed using a statistical method to display the distribution of data based on the extremes, first quartile, medium and third quartile (boxplot) (McGill, Tukey et al. 1978))

The start (SOS) and the end (EOS) of the season are the times when the NDVI increases or decreases to a dynamic threshold (10% of the seasonal amplitude), and the length of the season (LOS) is the difference between EOS and SOS. The Large integral (LINTG) is an estimate of total vegetation production from the zero level. The Small integral (SINGTG) is a measure of vegetation production, calculated above the base value (BVAL).

The thirteen phenological metrics that were extracted, and their definitions, are listed in Table 2 and depicted in Fig 3

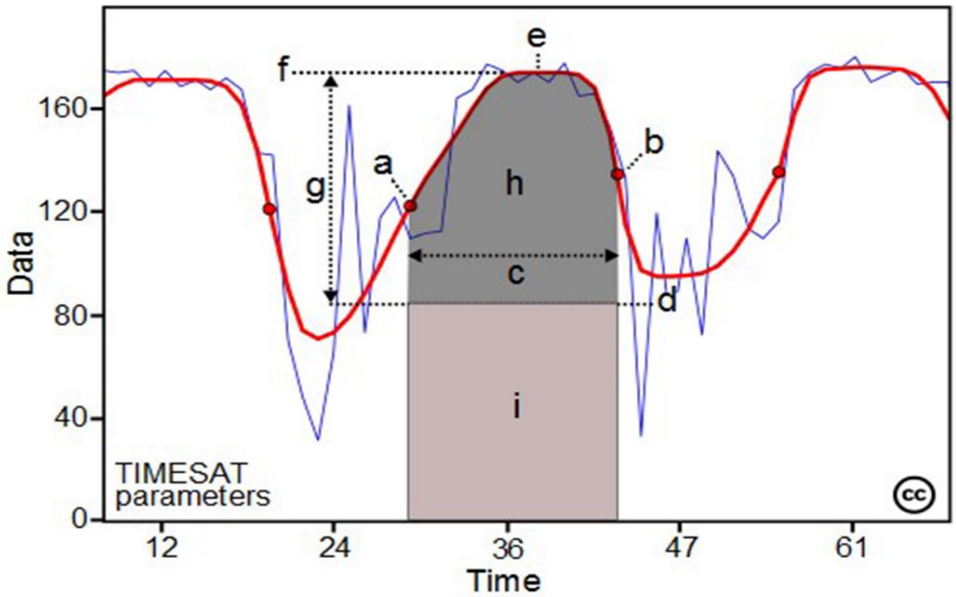


Fig. 3. Phenological metrics generated by TIMESAT: (a) beginning of season, (b) end of season, (c) length of season, (d) base value, (e) time of middle of season, (f) maximum value, (g) amplitude, (h) small integrated value, (h+i) large integrated value. The red and blue lines represent the filtered and the original data, respectively. Source: TIMESAT

Table 2. Seasonality metrics in TIMESAT

Phenological metrics	Abbreviation	Description
Start of season	SOS	Time for which the left edge has increased to 10% of the seasonal amplitude measured from the left minimum level.
End of season	EOS	Time for which the right edge has decreased to 10% of the seasonal amplitude measured from the right minimum level.
Middle of season	MOS	Mean value of the times for which the left part of the VGI curve has increased to the 90% level and the right part has decreased to the 90% level.
Length of season	LOS	Time from the start to the end of the season.
Base value	BVAL	The average of the left and right minimum values.
Maximum value	PEAK	Maximum VGI value for the fitted function during the season.
Amplitude	AMPL	Difference between the peak value and the base level.
Large integral	LINTG	The area under the smoothed curve between SOS and EOS.
Small integral	SINTG	The area below the base level from the SOS to EOS.
Left derivative	LDERIV	Rate of increase at the SOS between the left 10% and 90% of the amplitude.
Right derivative	RDERIV	Rate of decrease at the EOS between the right 10% and 90% of the amplitude.
Start of season value	SOSV	Start of season value.
End of season value	EOSV	End of season value.

3.3. Regression model development (Stepwise regression)

The phenological metrics were presented in raster format with a 10 m cell size and the pixel values for each corresponding ground measurement were extracted. These metrics were used as input data to develop an empirical model for wheat yield estimation using the STEPWISE regression approach. This method tests different combinations of all variables (thirteen phenological metrics).

The linear regression model was generated for each combination, and the one with the highest R2 values and lowest RMSE values was recognized as the best model to estimate wheat yield (Kerdsueb, Teartisup et al. 2014, An, Zhao et al. 2016).

The general model is shown in Equation (3) where y represents the dependent variable (Wheat Yield), Ci represents the coefficients of regression, Bi represents the spectral bands and b represents the intercept. The general model equation is expressed as follows:

$$Y = a_0 + \sum_1^n C_i * B_i \qquad \textbf{(3)}$$

Where Y, Ci, Bi, and a0 represent the dependent variable (wheat yield), coefficients of regression, variables, and intercept, respectively.

Model performance was assessed by comparing measured and simulated wheat grain yields. The evaluation was performed using the coefficient of determination (R2) for evaluating the linear relationship between the measured and estimated data (Equation. (4)), and the root mean square error (RMSE) to assess the average magnitude of the errors between the measurements (Equation. (5)) (Entekhabi, Reichle et al. 2010).

$$R^2 = \left(\frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2}} \right)^2 \quad \textbf{(4)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n}} \quad \textbf{(5)}$$

$\bar{x}$ and $\bar{y}$ refer to the mean of measured and estimated values of the studied variable respectively;

i is an identifier varying from 1 to the number of measured values, n;

Studies used statistical metrics for validation of models using statistical parameters like RSME, R and R2. Root Mean Square Error-RMSE is the measure of how well a regression line fits the data points. RMSE can also be construed as Standard Deviation in the residuals. In other words, an RMSE value of 0.5 reflects the poor ability of the model to accurately predict the data.

The coefficient of determination (or R square) is a parameter that is calculated when a linear regression is performed. However, understanding this parameter is not easy. All the more so, since some use it in turn to judge the quality of the regression model, while others attach practically no importance to it. The closer the coefficient of determination is to 0, the more the scatterplot disperses around the regression line. On the contrary, the more the R2 tends towards 1, the tighter the scatterplot becomes around the regression line. When the points are exactly aligned on the regression line, then R2 = 1.

3.4. Model validation (K-fold)

The accuracy of the obtained regression model was assessed using the k-fold cross validation (k-fold CV) (Cassel 2007). This approach uses k replicated samples of observation data, builds models with (k-1)/k of data, and tests with the remaining 1/k. Then the average error across all k sets is computed.

Evaluating the performance of the model is an important step in any Machine Learning project. We must be able to measure the generalization capacity of the model without introducing bias or data leaks. It is common to divide the dataset into training, validation, and test data. Unfortunately, one cannot afford this luxury in this case, where data is not sufficient.

Reserving part of the dataset for validation would reduce the already small amount of data available. And even if this sacrifice were made, the validation data would be too small to be representative of the entire dataset. To solve this problem, we are going to use cross-validation, which is an evaluation method that addresses the problems mentioned above. Generally, when we talk about cross-validation (cv), we refer to its most popular variant, the k-fold cross-validation. In this case, we take advantage of all the data available by dividing it into k equal parts (folds) on which a model is trained and tested during k iterations. At each iteration, the model is trained on k-1 folds and tested on the remaining folds (Benabdelouahab et al., 2015)(Hadria et al., 2018).

4. RESULTS AND DISCUSSION

4.1. Estimating phenological metrics for estimating wheat production

The phenological metrics were computed for each pixel on the basis of the NDVI profile.

Figure 4 illustrates the spatialization of the main phenological metrics that explain the yield

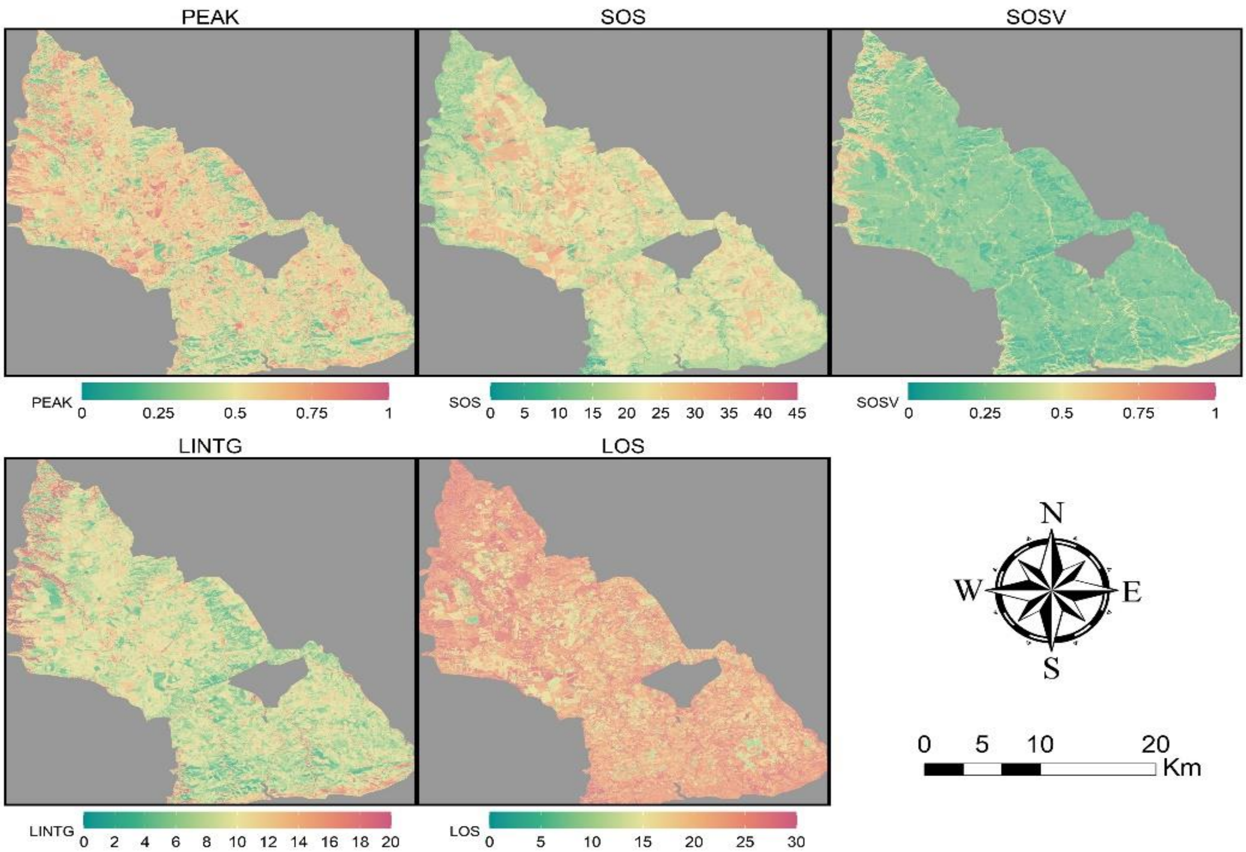


Fig. 4. : Spatialization of the main phenological metrics (PEAK- SOS-SOSV- LINTIG- LOS)

4.2. wheat grain yield estimation model

Based on the stepwise multiple linear regression method function, these metrics were used as dependent variables and the observed grain yield as an independent variable. The developed model is expressed by Equation (6):

Grain Yield (qx) = 0.39* SOS-2.42* LOS-46.68* PEAK+ 4.62* LINTG + 154.78* SOSV (6)

Where GY is the wheat grain yield, SOS is the start of the season, LOS is the length of the season, PEAK is the maximum value, LINTG is the Large Integral, and SOSV is the start of the season value.

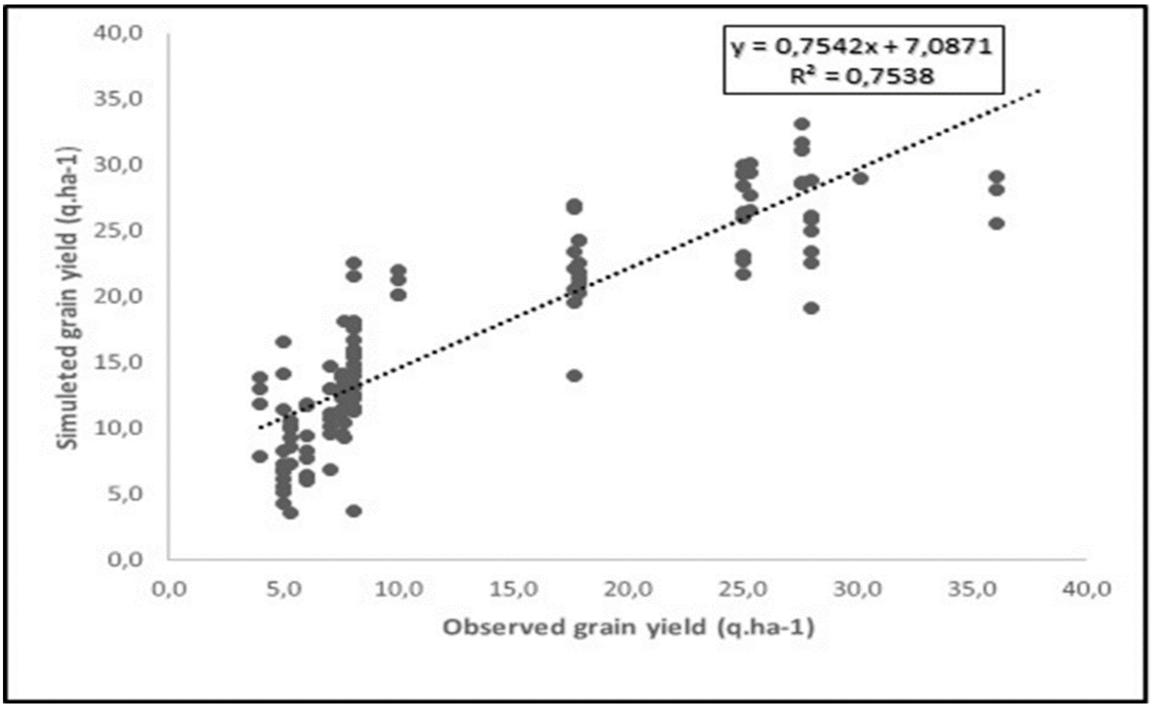


Fig. 5. Relationship between observed grain yield and estimated from phenological metrics

The relationship between the observed and simulated grain yield values was assessed in 123 locations over the study area for the cropping seasons of 2018-2019. The statistical indicators R2 and RMSE were 0.75 (p 0.01) and 7.08 q/ha, respectively. The developed model was validated based on the relationship between observed and predicted values (Figure 5).

We compared the wheat grain yield values predicted using the k-fold CV method with those observed in situ. The statistical indicators obtained from this comparison were R2 = 0.75 (p 0.01) and RMSE = 3.45.

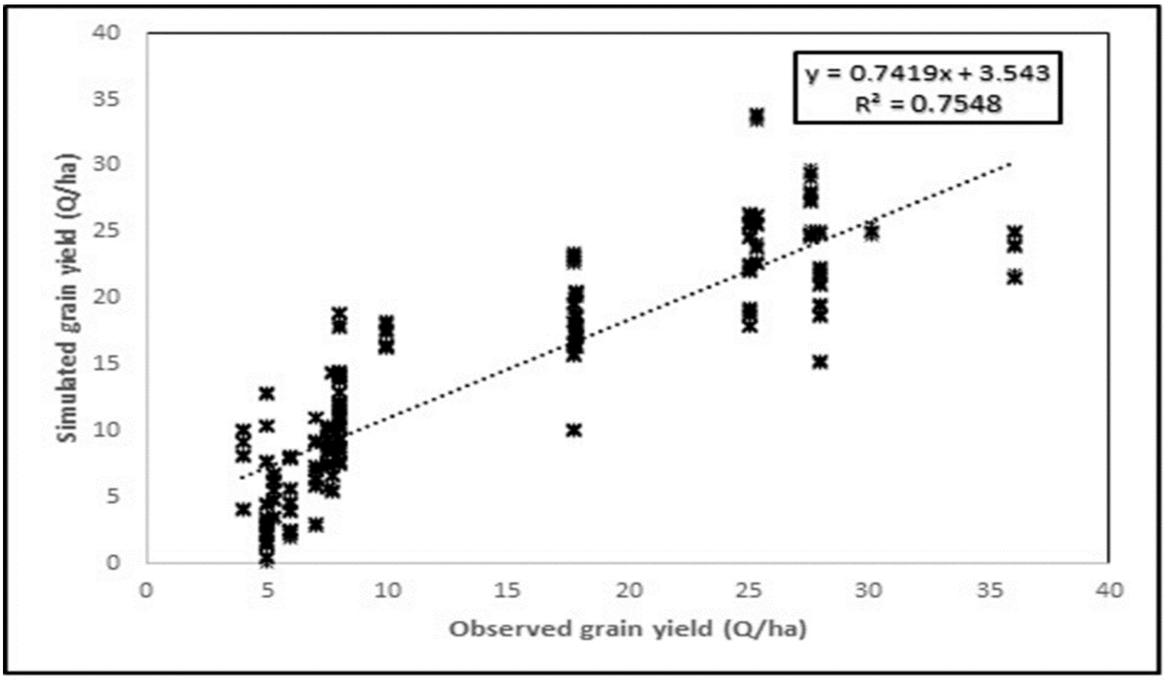


Fig. 6. Comparison between observed and predicted wheat grain yield using the k-fold CV method

4.3. The Spatial variations of wheat grain yield

Based on the previous calibrated Grain Yield model, the wheat grain yield was calculated at the level of the surveyed plots and represented in the form of a map showing the spatial variability of the estimated yield at field scale. (Fig 7)

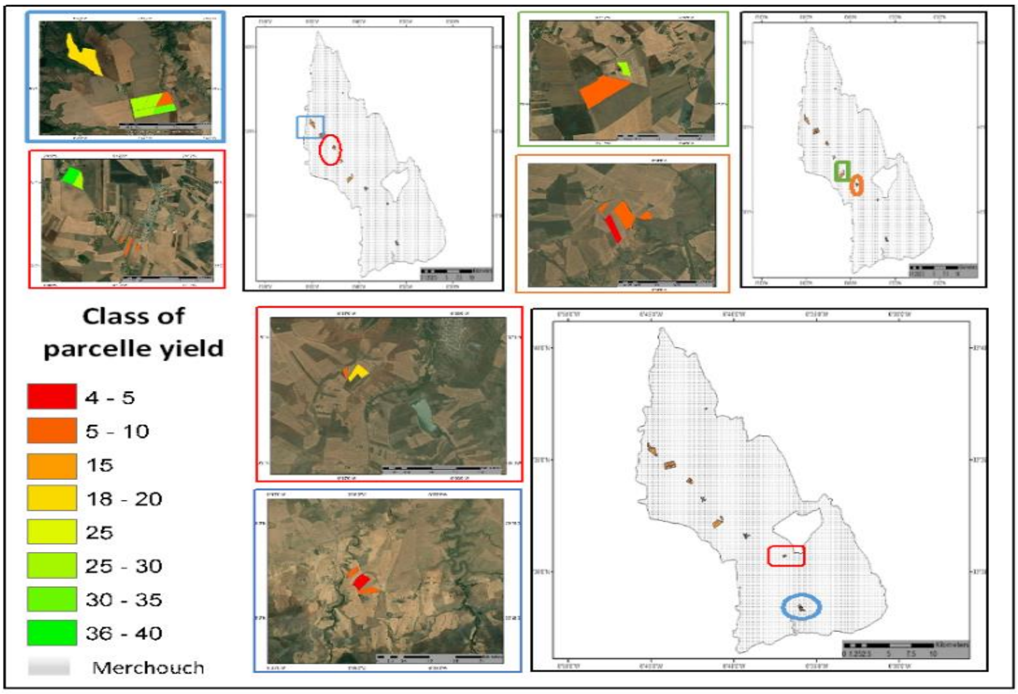


Fig. 7. Spatial variability of wheat yields at field scale for the cropping season 2018/2019

These obtained results clearly show the importance of using high resolution images like SENTINEL 2 and especially extracting phenological information to estimate wheat grain yields at the field scale.

These results are in agreement with previous studies showing that this satellite is the most appropriate for predicting yield (Lambert, Traoré et al. 201), (Toscano, 2019, Hunt, 2019, Escolà, Badia et al. 2017)

4.4. The spatialization of wheat grain yield at the commune level:

The Merchouch commune is a predominantly cereal zone. The calibrated model based on phenological metrics was applied in the study area in order to estimate yield and then we classified the yield by category over the entire Merchouch area. The results are shown in figure 8.

Spatial analysis of wheat yield is an effective tool to assess the suitability of wheat cultivation to different conditions (edaphic, climatic and human) as well as the risks associated with this crop (Benabdelouahab, 2020).

According to our model, the average yield is 15.75qx/ha, but the average yield of the ground sample is 13.17qx/ha. From this result, we can see that more than 30% of the total area has a yield between 10 and 20 qx/ha and more than 25% have less than 10 qx/ha, as well as 30% have a yield greater than 20 qx/ha (table 3).

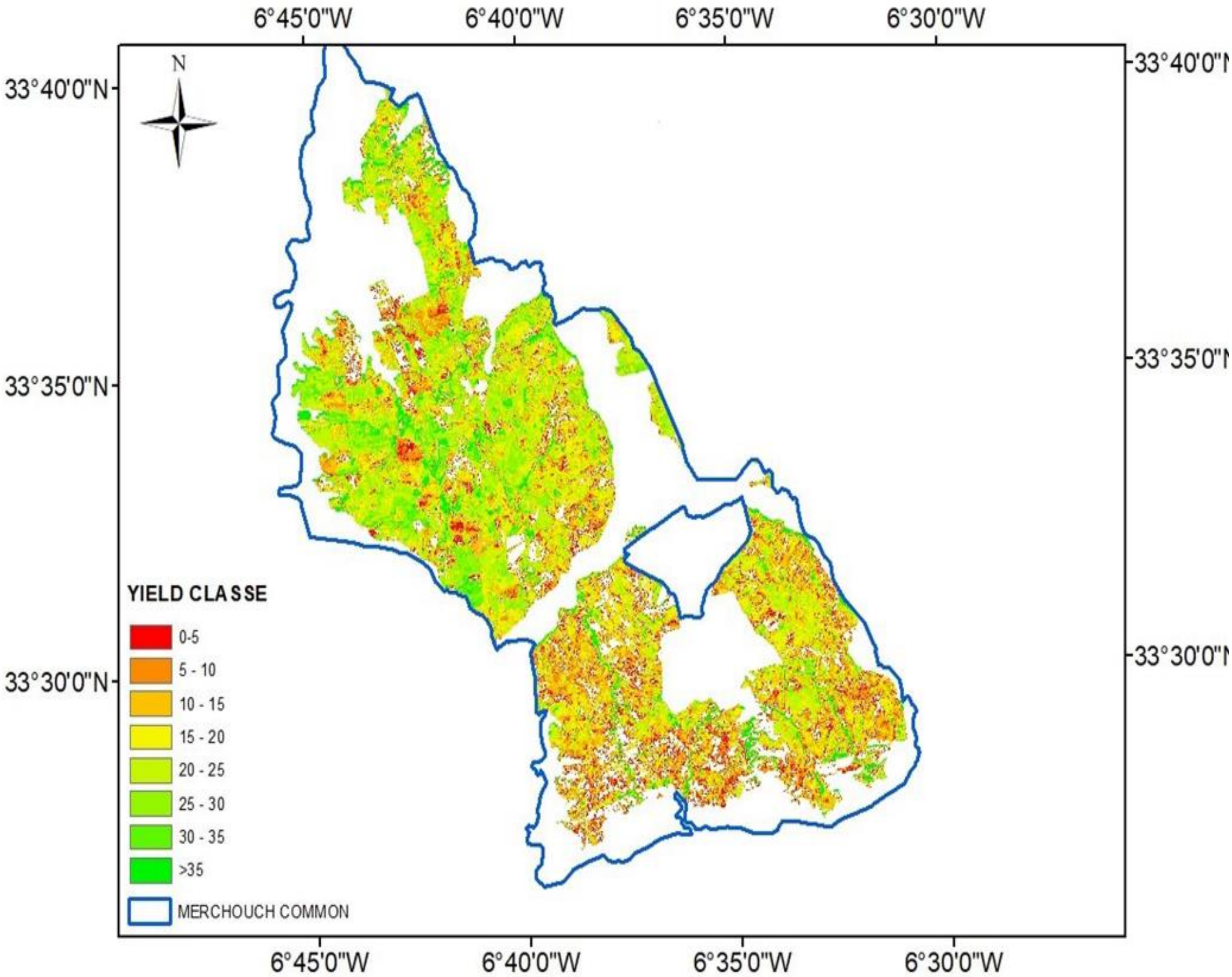


Fig. 8. Spatial wheat grain yield for cropping season 2018-2019 at the Merchouch common

The spatial grain yield distribution revealed a high spatial heterogeneity within the same area, calling into question the practices employed and challenging regional services, specifically the agricultural council, to better supervise agriculture in order to achieve a good yield. Regression analysis models for yield estimation can be a good tool for decision makers to orient their actions.

However, this model needs to be further refined with valid site truths and must be tested over several years and at different scales and under different conditions to be able to use it in agricultural statistics and specially to forecast yields one or two months before harvest.

Table 3. Class grain yield

Classe	Grain yield averge	Area(ha)	Percentage
Low yield	0-5	3475.53	14%
Low yield	5-10	1863,07	11%
Medium yield1	10-15	2408,49	14%
Medium yield	15-20	2992,89	17%
Medium yield	20-25	3103,60	18%
High yield	25-30	1855,28	11%
High yield	30-35	715,83	4%
very high	>35	953,46	5%

The model, based on the selected phenological metrics, can be used as an operational tool for monitoring wheat grain yield on a large scale.

5. CONCLUSIONS AND PERSPECTIVES.

Grain yield estimation is one of the main ministry's concerns in Morocco, especially for wheat, a crop with high economic potential. Remote Sensing offers an efficient and reliable means of collecting and creating information.

The proposed wheat grain yield model explained 75% of its spatial wheat grain yield variation with a root-mean-square error of 0.7 t ha-1. This result showed that the errors were acceptable, confirming the ability of the model to estimate wheat grain yield accurately.

The model developed by Iizumi, Yokozawa et al. (2014) explained only 45–81% of the spatial variation of yields, with root-mean-square errors of 0.5–1.8 t ha-1.

The proposed model, based on phenological metrics derived from NDVI time series, overcomes the problem of lack of meteorological, soil and yield data without losing accuracy while taking advantage of the spatial resolution offered by satellite products. (Bakker, Govers et al. 2005, Anagnostou, Maggioni et al. 2010), Lionboui, Benabdelouahab et al. 2020)

The use of SENTINEL 2 data at 10m leads to a class mixture at the pixel level, which affects the mean NDVI value. This mixture affects the model accuracy, especially with the small farm areas of less than 5 ha, which represents more than 80% of total farm categories.

This model describes the wheat behavior at the pixel level and reflects production conditions, including physical or human factors. The developed model can substitute the empirical models proposed by regions that combine the indices of vegetation, precipitation and temperature.

According to our model, the average yield in Merchouch is 15.75qx/ha, but the average yield of the plots under investigation is 13.17qx/ha. The model, based on the selected phenological metrics, can be used as an operational tool for monitoring wheat grain yield on a large scale and can be used in agriculture statistics. The approach presented can potentially be replicated in other regions of Morocco. Our results prove that this data has the ability to produce a good model for estimating yield at the plot, especially with the launch of new satellites such as Sentinel-2B. The images are now provided on a 5 day5-day basis. Also, they should improve the availability of input data (NDVI Series). This spatial analysis approach allows managers and stakeholders to analyze the agricultural policies' impact, monitor the agronomic potential regression, and optimize the land use choices.

This model is a good tool for managers; we cannot pretend that this model is operational to estimate yields now. The final model needs to be further refined with ground truth data and must be tested over several years at different scales and under different conditions to be able to use it in agricultural statistics and specially to prevent yield one or two months before harvest.

6. ACKNOWLEDGMENT

I would like to thank all the collaborating professors who guided me in this research work and who, thanks to them, were able to model the yield of wheat from the phenological metrics derived from satellite images, this fundamental crop in the agriculture of Morocco.

7. FUNDING

NO FUNDING

8. AUTHOR CONTRIBUTIONS: MAIN Author who writes the paper : Adra IDRISSI /

Contributors to the study: Aziz Htitiou & Youssef Lebrini / **Supervisors of the Study**: Samir NADEM, Abdelghani BOUDHAR and Tarik BENABDELOUAHAB

9. REFERENCES

An, D., G. Zhao, C. Chang, Z. Wang, P. Li, T. Zhang, and J. J. I. J. o. R. S. Jia (2016). "Hyperspectral field estimation and remote-sensing inversion of salt content in coastal saline soils of the Yellow River Delta." 37 (2): 455-470 .

Anagnostou, E. N., V. Maggioni, E. I. Nikolopoulos, T. Meskele, F. Hossain and A. Papadopoulos (2010). "Benchmarking High-Resolution Global Satellite Rainfall Products to Radar and Rain-Gauge Rainfall Estimates." IEEE Transactions on Geoscience and Remote Sensing48 (4): 1667-1683.

Andarzian, B., Bakhshandeh, A. M., Bannayan, M., Emam, Y., Fathi, G., & Alami Saeed, K. (2008). WheatPot: A simple model for spring wheat yield potential using monthly weather data. *Biosystems Engineering*, *99*(4), 487–495. https://doi.org/10.1016/J.BIOSYSTEMSENG.2007.12.008

Bakker, M. M., G. Govers, F. Ewert, M. Rounsevell, and R. Jones (2005). "Variability in regional wheat yields as a function of climate, soil, and economic variables: Assessing the risk of confounding." Agriculture, Ecosystems & Environment110 (3): 195-209.

Balaghi, R., B. Tychon, H. Eerens, and M. Jlibene (2008). "Empirical regression models using NDVI, rainfall, and temperature data for the early prediction of wheat grain yields in Morocco." International Journal of Applied Earth Observation and Geoinformation10 (4): 438-452.

Benabdelouahab, T., Balaghi, R., Hadria, R., Lionboui, H., Minet, J., & Tychon, B. (2015). Monitoring surface water content using visible and short-wave infrared SPOT-5 data of wheat plots in irrigated semi-arid regions. International Journal of Remote Sensing, 36(15), 4018–4036. https://doi.org/10.1080/01431161.2015.1072650

Benabdelouahab, T., Y. Lebrini, A. Boudhar, R. Hadria, A. Htitiou, and H. Lionboui (2019). "Monitoring spatial variability and trends of wheat grain yield over the main cereal regions in Morocco: a remote-based tool for planning and adjusting policies." Geocarto International: 1-20.

Cassel, D. L. (2007). Re-sampling and simulation, the SAS way. Proceedings of the SAS Global Forum 2007 Conference, Cary, NC, SAS Institute Inc.

CIMMYT. (2018). New publications: The importance of wheat in the global food supply to a growing population – CIMMYT. Importance of Wheat in the World.

https://www.cimmyt.org/publications/new-publications-the-importance-of-wheat-in-the-global-food-supply-to-a-growing-population

Doraiswamy, P., J. Hatfield, T. Jackson, B. Akhmedov, J. Prueger, and A. J. R. s. o. e. Stern (2004). "Crop condition and yield simulations using Landsat and MODIS." 92 (4): 548-559 .

B., R. Hadria, S. Erraki, G. Boulet, P. Maisongrande, A. Chehbouni, R. Escadafal, J. Ezzahar, J. C. B. Hoedjes, M. H. Kharrou, S. Khabba, B. Mougenot, A. Olioso, J. C. Rodriguez, and V. Simonneaux (2006). "Monitoring wheat phenology and irrigation in Central Morocco: On the use of relationships between evapotranspiration, crop coefficients, leaf area index and remotely-sensed vegetation indices." Agricultural Water Management79 (1): 1-27.

Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. J. J. H. Crow (2010). "Performance metrics for soil moisture retrievals and application requirements." 11 (3): 832-840.

Escolà, A., N. Badia, J. Arnó, and J. A. Martnez-Casasnovas (2017). "Using Sentinel-2 images to implement Precision Agriculture techniques in large arable fields: First results of a case study." Advances in Animal Biosciences8: 377-382.

FAO. (2022). FAO Cereal Supply and Demand Brief | World Food Situation | Food and Agriculture Organization of the United Nations.

Geng, L., M. Ma, X. Wang, W. Yu, S. Jia, and H. J. R. S. Wang (2014). "Comparison of eight techniques for reconstructing multi-satellite sensor time-series NDVI data sets in the Heihe river basin, China." 6 (3): 2024-2049.

Hadria, R., B. Duchemin, L. Jarlan, G. Dedieu, F. Baup, S. Khabba, A. Olioso, and T. Le Toan (2010). "Potentiality of optical and radar satellite data at high spatio-temporal resolution for the monitoring of irrigated wheat crops in Morocco." International Journal of Applied Earth Observation and Geoinformation12: S32-S37.

Hadria, R., Benabdelouahab, T., Mahyou, H., Balaghi, R., Bydekerke, L., El Hairech, T., & Ceccato, P. (2018). Relationships between the three components of air temperature and remotely sensed land surface temperature of agricultural areas in Morocco. International Journal of Remote Sensing, 39(2), 356–373. https://doi.org/10.1080/01431161.2017.1385108

Htitiou, A., A. Boudhar, Y. Lebrini, R. Hadria, H. Lionboui, L. Elmansouri, B. Tychon, and T. Benabdelouahab (2019). "The Performance of Random Forest Classification Based on Phenological Metrics Derived from Sentinel-2 and Landsat 8 to Map Crop Cover in an Irrigated Semi-arid Region." Remote Sensing in Earth Systems Sciences2 (4): 208-224.

Iizumi, T., M. Yokozawa, G. Sakurai, M. I. Travasso, V. Romanenkov, P. Oettli, T. Newby, Y. Ishigooka, and J. Furuya (2014). "Historical changes in global yields: major cereal and legume crops from 1982 to 2006." Global Ecology and Biogeography23 (3): 346-357.

Jayawardhana, W. and V. J. P. f. s. Chathurange (2016). "Extraction of agricultural phenological parameters of Sri Lanka using MODIS and NDVI time series data." 6: 235-241.

Kerdsueb, P., P. J. I. J., E. S. Teartisup and Development (2014). "The use of geoinformatics for estimating soil organic matter in the central plain of Thailand." 5 (3): 282.

Lambert, M.-J., P. C. S. Traoré, X. Blaes, P. Baret and P. Defourny (2018). "Estimating smallholder crop production at the village level from Sentinel-2 time series in Mali's cotton belt." Remote Sensing of the Environment216: 647-657.

Lionboui, H., T. Benabdelouahab, A. Htitiou, Y. Lebrini, A. Boudhar, R. Hadria, and F. Elame (2020). "Spatial assessment of losses in wheat production value: A need for an innovative approach to guide risk management policies." Remote Sensing Applications: Society and Environment18: 100300.

Löw, F., U. Michel, S. Dech, C. J. I. j. o. p. Conrad and r. sensing (2013). "Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines." 85: 102-119.

McGill, R., J. W. Tukey, and W. A. J. T. A. S. Larsen (1978). "Variations of box plots." 32 (1): 12-16.

Mkhabela, M., P. Bullock, S. Raj, S. Wang, Y. J. A. Yang, and F. Meteorology (2011). "Crop yield forecasting on the Canadian Prairies using MODIS NDVI data." 151 (3): 385-393.

Ouharba, E., Z. El, Z. Triqui, and R. Moussadek (2019). "Impact des Changements Climatiques sur la Céréaliculture au Maroc. Etude de Cas:Rommani (Région de Rabat), Centre du Bassin Versant du Bouregreg. Climate Change Impact on Cereal Culture in Morocco. Impact des Changements Climatiques sur la Céréaliculture à Rommani. "

Quarmby, N., M. Milnes, T. Hindle, and N. J. I. J., R. S. Silleos (1993). "The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction." 14 (2): 199-210.

Rasmussen, M. S. J. I. J. o. R. S. (1992). "Assessment of millet yields and production in northern Burkina Faso using integrated NDVI from the AVHRR." 13 (18): 3431-3442.

Reynolds, P. E., Thevathasan, N. V., Simpson, J. A., Gordon, A. M., Lautenschlager, R. A., Bell, W. F., Gresch, D. A., & Buckley, D. A. (2000). Alternative conifer release treatments affect microclimate and soil nitrogen mineralization. *Forest Ecology and Management*, *133*(1–2), 115–125. https://doi.org/10.1016/S0378-1127(99)00302-3

Rouse Jr., J., R. Haas, D. Deering, J. Schell, and J. Harlan (1974). "Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation. [Great Plains Corridor**].** last accessed 2016/11/21.

10. ADDITIONAL READING

Monitoring spatial variability and trends of wheat grain yield over the main cereal regions in Morocco: a remote-based tool for planning and adjusting policies." Geocarto International: 1-20. By Benabdelouahab, T., Y. Lebrini, A. Boudhar, R. Hadria, A. Htitiou, and H. Lionboui (2019)

Crop yield forecasting on the Canadian Prairies using MODIS NDVI data." 151 (3): 385-393 by Mkhabela, M., P. Bullock, S. Raj, S. Wang, Y. J. A. Yang, and F. Meteorology (2011

The Performance of Random Forest Classification Based on Phenological Metrics Derived from Sentinel-2 and Landsat 8 to Map Crop Cover in an Irrigated Semi-arid Region." Remote Sensing in Earth Systems Sciences2 (4): 208-224 by Htitiou, A., A. Boudhar, Y. Lebrini, R. Hadria, H. Lionboui, L. Elmansouri, B. Tychon, and T. Benabdelouahab (2019).

11. KEY TERMS AND DEFINITIONS

SENTINEL 2: The Copernicus Sentinel-2 mission is based on a constellation of two identical satellites in the same orbit. Each satellite carries an innovative wide swath high-resolution multispectral imager with 13 spectral bands for a new perspective of our land and vegetation (source ESA)

Phenological metrics: The phenological metrics describe the stages of the phenology of the plants which are phenophases that follow one another, from the dormancy (or sowing in the agricultural practices) at the beginning of the season, the apogee (or heading), the maturity phase, the flowering, the senescence, the fall of the leaves, and finally, the dormancy