NDVI prediction of Mediterranean permanent grassland using soil moisture products

Filippo Milazzo, Luca Brocca , Tom Vanwalleghem, Andres Peñuela

Abstract— Vegetation indexes are widely used as a proxy of vegetation status, they are often used to monitor and assess qualitatively and quantitatively the growing season. The Normalized Vegetation Index (NDVI) is the most widely used in agriculture, frequently as a proxy for different physiological and agronomical aspects, such as drought stress and crop yield losses evaluation. NDVI forecast is usually correlated to precipitation however, in Mediterranean and arid climates, it is not well correlated due to prolonged dry periods and sparse precipitation events. In this study, we forecast Mediterranean permanent grassland NDVI at 7 and 30 days ahead using machine learning and two soil moisture products as predictors, simulated soil moisture values and satellite-based Soil Water Index (SWI) values. Results show that both products can be used as reliable predictors of permanent grassland in Mediterranean areas. Predictions at 7 days are more accurate and better forecast the negative effect of drought on vegetation dynamics than 30 days. This study shows the potential of using a simple methodology and readily available data to predict the grassland growth dynamic in the Mediterranean area.

Index Terms— Soil moisture, grassland, SWI, NDVI

Filippo Milazzo, Phd Studen, Area of Hydraulic Engineering, Department of Agronomy, Leonardo da Vinci Building, Rabanales Campus, University of Cordoba.,14071-Córdoba, z62mimif@uco.es.
Luca Brocca Director of Research Institute for Geo-Hydrological Protection, National Research Council, Perugia, Italy luca.brocca@irpi.cnr.it
Tom Vanwalleghem and Andres Peñuela are professor and researcher at Area of Hydraulic Engineering, Department of Agronomy, Leonardo da Vinci Building, Rabanales Campus, University of Cordoba.,14071-Córdoba, ag2vavat@uco.es and apenuela@uco.es .

I. INTRODUCTION

Savanna-like agroforestry systems cover about 3.5 million hectares in Mediterranean Europe and about 1 million hectares in North Africa [1], [2]. This land use is composed of scattered oak trees (Quercus rotundifolia), and permanent grassland (unrenewed herbage layer of 5 or more years) for livestock grazing [2]. In Iberian peninsula, savanna-like systems are called Dehesa in Spain and Montado in Portugal, these names refer to the fact that the land is divided into large plots bordered by stone walls, where rotational grazing (rangeland) is practised [2]. Grassland ecosystem plays an important role in sustaining the economy of marginal lands in the Mediterranean through livestock production and in preserving their endemic biodiversity and cultural heritage [3]–[5]. In the Mediterranean climate, grassland production is limited due to frequent long dry summers, which cause a severe crop yield drops and hence, important economic losses. This susceptibility of Mediterranean grassland to drought is increasingly exacerbated by climate change [6], [7]. Therefore, accurate forecasts of grassland yield, in particular during dry periods, are crucial for both farmers and policymakers to apply mitigation measures and hence, ensure food security [8].

Different methods have been developed for predicting grassland dynamics and yield using weather forecasting. McDonnel et al. [9], attempted 1 and 6 days forecasts of the grassland dynamics, using management inputs, such as fertiliser application, and weather inputs, such as temperature and precipitation. Trnka et al.[10], developed an accurate grassland growth model, including as inputs, not only the weather and the fertiliser application, but also the soil moisture balance. The benefit of using soil moisture information is that it integrates weather, evapotranspiration, plant available water and hence, vegetation state information. Indeed, soil moisture is an essential driver of grassland dynamics, it can explain up to 60% of the grassland yield variability and when considered in productivity models greatly improve model performance [11]. In the literature, vegetation dynamics are commonly monitored by remote sensing techniques, and in particular by using the Normalized Difference Vegetation Index (NDVI). NDVI is a simple indicator of the vegetation greenness and it is widely applied to estimate vegetation density and crop yields [12], [13]. Notably, in Spain NDVI is used for agricultural insurance purposes to quantify grassland yield losses during the growing

season (Oct-Jun) and compensations due to drought or extreme weather conditions [14]. Previous studies have shown that NDVI can be estimated from soil conditions, in particular, they show a good correlation in dry climates [15]. For instance, Chen et al. [16], found a good correlation between soil moisture and NDVI in Australia's mainland. In climates characterized by prolonged dry seasons, such as South Spain, NDVI has shown to be better correlated to soil moisture than precipitation [17]. In the particular case of grassland areas, Wang et al. [18] found a better correlation in semi-arid than humid regions of the USA.

Previous studies have shown the potential of using NDVI to support grazing and harvesting planning and in particular, NDVI predictions to anticipate water deficiencies and hence, yield losses [19]–[21]. One approach to predict the NDVI is using autoregressive models, i.e. forecast future NDVI values using a linear combination of past NDVI values. This approach has shown high reliability in forestry land uses mainly thanks to the plant growth seasonality [22]. Another approach not based on the use of past data, is the use of seasonal weather forecasts. Iwasaki [23] in an arid climate, tried to predict NDVI distribution for 1-3 months using a seasonal weather forecast. They showed a weak prediction efficiency especially and advised against the use of precipitation forecasts for NDVI prediction in dry regions. Considering the NDVI as a proxy of vegetation growth, some studies have also used parametric crop growth models to forecast NDVI values. However, parametric crop growth models, have shown a low NDVI prediction accuracy [24] and a worse performance than an increasingly popular approach, machine-learning based methods [25].

With the advance of remote sensing methods and the informatization of the agriculture operation, machine-learning algorithms provide the possibility to develop forecasting or decision tools for land managers, farmers and other agro-forestry stakeholders [26]–[28]. Machine-learning approaches provide powerful tools that are applied in different fields [29], [30] such as weed detection [31], soil analysis [32], management zone clustering [33], irrigation and yield prediction and stress prediction [34], [35]. However, predicting vegetation development remains a current challenge because several ecosystem processes affect vegetation dynamics [36], [37]. Currently, process-based models are not able to predict accurately the vegetation dynamic interrelating the multiple ecosystem processes that impact vegetation growth [37]. For this reason, the use of machine-learning, due to its high performance, and multifold applicability quickly increases worldwide [38]. Different approaches have been widely applied to predict the vegetation dynamic, such as artificial neural networks, support vector regression, random forest and regression trees [39]. These methods are characterized of the independence of the relationship between the predictors and predictive variable, particularly if compared to the traditional models as linear regression, which imply a Gaussian distribution for the input variables [40]. Roy [41] compared the performance of some of the most used machine

learning algorithms to forecast large-area average of NDVI in Bangladesh and showed that the Random Forest algorithm had the best performance.

In this study, we present an innovative NDVI forecasting model based on the application of the Random Forest machine learning algorithm and the use of past and present temperature and soil moisture information as predictors. Soil moisture information consists of two products: modelled daily soil moisture values and satellite-derived values of Soil Water Index (SWI) at a point and single-pixel scale respectively. Using each soil moisture product, we create two versions of the NDVI forecasting model that we tested and compared for 7-day and 30-day lead times in a Mediterranean permanent grassland.
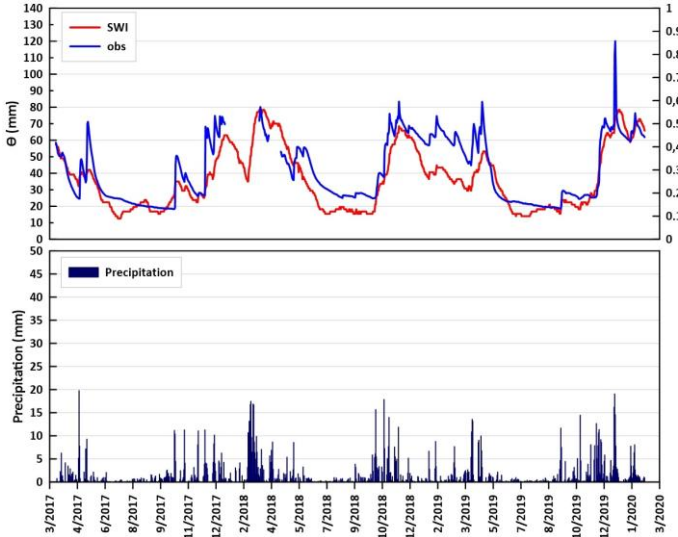
II. Material and Methods

A. Study area

The study was carried out in Santa Clotilde commercial farm located in the north of Córdoba province, Southern Spain (38.2° N; 4. 17° W, 700 m a.m.s.l.). The main activity of the Santa Clotilde farm is the extensive livestock production in the Dehesa agroforestry system, Fig. 1; bovines and swine are grazing rotationally the whole year. Soil texture is sandy-loam (6,7% clay, 64% sand, 29,3% loam), due to rotational grazing, the first 30 cm of soil profile holds 70% of the total carbon stock [42]. According to the Köppen-Geiger classification, the climate is Mediterranean, with an average annual rainfall of 878 mm, cold-dry winters seasons, long summers and a mean temperature of 25.4 °C [43]. Since 2017 five soil moisture sensors (Campbell Scientific CS655) were installed in grassland open-field, at 3 soil depths at 5, 15, 25 cm depth) monitoring the grassland soil moisture dynamics.

Fig. 1: Study area. Santa Clotilte Dehesa farm and sensors location.

In this study, we used the soil moisture readings at 25 cm depth to have comparable results to Soil Water Index, which is representative of the first 20 cm of soil profile Fig.2. Precipitation data is obtained using the SM2RAIN-ASCAT satellite-based method. SM2RAIN-ASCAT is a global product obtained from Advanced SCATterometer (ASCAT) satellite through the SM2RAIN algorithm developed by Brocca et al. [44]. The SM2RAIN algorithm allows calculating rainfall using the inverse equation of water balance to calculate rainfall using in situ or satellite-based soil moisture data [45], [46]. We also estimated the satellite-based Soil Water Index (SWI) at the study area. SWI of Copernicus Global Land Service [47] is acquired from measurements of near-surface soil moisture supplied by ASCAT by means of an algorithm which summarizes and exponentially weights past measurements according to the time length T, which ranges between 001 and 100 [48]. The T factor indicates how many past observations of surface soil moisture affect the current value of SWI. Conceptually, higher delay and the increasing smoothing signal detected at the soil surface, from a higher T value, is comparable to the effect of the soil water infiltration. Thus SWI is a reliable proxy of soil moisture content at 20 cm depth [44], [49]. Specifically, this study has been selected with a value of the T-parameter equal to 20 days.

Fig. 2: Top panel – daily soil moisture information observed in the study area: the blue line is the soil moisture measured by field sensors; the red line is the Soil Water Index (SWI). Bottom panel – daily precipitation obtained using the SM2RAIN-ASCAT method.

B. Soil moisture model

To estimate the soil moisture dynamic of the Mediterranean permanent grassland, we modelled soil moisture using as water input the daily satellite rainfall data from ASCAT data [46]. The soil moisture dynamic model is conceptually based on the BEACH model [50], which divides the soil moisture reservoir in two layers: the top layer, which depth is delimited by the root zone, water balance is determined by rainfall, evapotranspiration, runoff and deep percolation; and the passive layer where soil moisture is mainly driven by the deep percolation. In this study, we simplify the model to only represent the top layer of 25 cm depth. Irrigation input is not considered because it is a rainfed grassland.

The soil moisture model (SM25) calculates the volume of water stored in the soil (St; in mm) considering the stored volume from the previous day (St-1):

$$S_t = S_{t-1} + Rf - ET_{act} - DP \tag{1}$$

where St is the daily soil moisture (mm); St-1 is the soil moisture of the antecedent day (mm); Rf is the net precipitation (mm); ETact is the actual evapotranspiration (mm); Dp is the deep percolation (mm).

To compute Rf that reaches the soil surface we apply a the formula proposed by Morgan and Duzant [51]:

$$Rf = R(1 - PI) \tag{2}$$

Where R is the total daily precipitation (mm) and PI is the plant interception (mm). To calculate PI we applied the empirical equation proposed by Braden [52] as a function of the Leaf area index (LAI), the Canopy cover CC and the daily precipitation R (mm):

$$PI = aLAI\left(1 - \frac{1}{1+\frac{CCR}{aLAI}}\right) \tag{3}$$

where PI daily plant interception (mm); a is an empirical coefficient that ranges between 0,3 (before senescence) and 0,6 (end of the senescence period) [53]; CC is canopy cover; LAI is the Leaf area index. Similar to the DREAM model [54] and the SWAP model [55] the actual evapotranspiration (ETact) is calculated as a combination of the reference evapotranspiration (ET0) from the vegetated fraction (CC) and the actual evaporation from the bare soil fraction (1–CC):

$$ET_{act} = ET_{veg} CC + E_{soil}(1 - CC) \tag{4}$$

where ETveg is the actual daily evapotranspiration from the vegetated fraction (in millimetres) and Esoil the actual daily evaporation of the bare soil fraction (in millimetres), shown in (5) and (6). Both ETveg and Esoil depend on the degree of water availability in the soil. The degree of water availability is expressed by actual soil moisture divided by field capacity soil moisture. This approach is based on the following assumptions [56] (5):

If St-1 > of water stored in the soil at field capacity:

$$ET_{veg} = ETo - PI \tag{5}$$

$$E_{soil} = ETo \tag{6}$$

therefore:

$$ET_{act} = (ETo - PI)CC + ETo(1 - CC) \tag{7}$$

If St-1 < of water stored in the soil at field capacity (Sfc) and higher than at the wilting point (Swp), ETact is equal to the potential plant evapotranspiration in mm, plus the actual soil daily evaporation of the bare soil fraction in mm, (8) and (9):

$$ET_{veg} = (ETo - PI)\left(\frac{S_{t-1}-S_{wp}}{S_{fc}-S_{wp}}\right) \qquad (8)$$

$$E_{soil} = ETo\left(\frac{S_{t-1}-S_{wp}}{S_{fc}-S_{wp}}\right) \qquad (9)$$

Therefore (10):

$$ET_{act} = ET_{veg}\ CC + E_{soil}(1 - CC) \qquad (10)$$

Deep percolation Dp was simulated by applying the same BUDGET model method [57], (11):

$$Dp = d_s\,\tau\,(\theta_{sat} - \theta_{fc})\left(\frac{e^{\theta - \theta_{fc}}-1}{e^{(\theta_{sat}-\theta_{fc})}-1}\right) \qquad (11)$$

Where ds is the depth of the soil A-horizon (mm); θ is the soil moisture expressed as millimetres of water depth per millimetre of soil depth; θsat is the soil moisture at saturation; θfc is the soil moisture at field capacity; τ is a drainage parameter, given by the equation (12):

$$0 \le \tau = 0.0866^{0.8063\log_{10}(K_{sat})} \le 1 \qquad (12)$$

where K_{sat} is the saturated hydraulic conductivity (mm d^{-1}).

C. Soil moisture model calibration and validation

For the soil moisture model calibration and validation, we used the observed daily soil moisture values, from 17/03/2017 to 01/02/2021. From 17/03/2017 to 23/06/2019 for the model calibration and from 24/04/2019 to 12/02/2020 for the model validation.

Model performance is evaluated using the Nash-Sutcliffe efficiency (NSE). NSE determines the relative magnitude of the residual variance compared to the observed data variance [58]. For the model calibration computation we used the Non-dominated Sorting Genetic Algorithm (NSGA-II) and considered the following parameters: canopy cover (CC); saturated hydraulic conductivity in mm/day (K$_{sat}$); soil moisture ratio at wilting point in mm/mm (Swp); soil moisture ratio at field capacity in mm/mm (Sfc); soil moisture ratio at wilting saturation in mm/mm (theta_sat). NSGA-II was set to maximize NSE.

D. NDVI forecasting models

The NDVI forecast model uses available NDVI, temperature and soil moisture data at present to predict NDVI values. We developed two NDVI ahead forecasting models using the selected two soil moisture products:

$NDVI_{SWI}$: NDVIt0+x~ SWI + T20 + NDVIt0

NDVI_SM25: NDVIt0 + x~ SM_25 + T20 + NDVIt0

where NDVIt0+x is the forecasted NDVI at day t0+x (7 or 30); SWI is the Soil Water index (SWI); T20 is the cumulative mean temperature of the previous 20 days; NDVIt0 is the observed NDVI value at present. They were retrieved from the Copernicus Sentinel-2 using the Google Earth Engine. Observed NDVI values are also used as a reference to compare forecasted results. Observed NDVI corresponds to an area of 21 m radius, covering a grid of approximately 3 × 3 pixels with 10-m of spatial resolution, located in the hill-plateau of Santa Clotilde farm, in open grassland avoiding the tree influence; SM_25 is the simulated soil moisture at 25 cm soil depth. Grassland's NDVI was predicted by applying the Random Forest machine learning algorithm [59]. This approach is composed of accumulation of singular decision trees (estimators) that allow an exceptional achievement of prediction accuracy [60]. The training and testing of the NDVI forecast models were performed from 21/07/2015 to 30/12/2021, using 50% of the data for each one. Prediction performance was evaluated using the NSE and the Mean Bias Error (MBE). MBE is used to estimate the bias between the predicted value and the observed [61]. In comparison with NSE, MBE provides a view of how close the forecasts are to the measurements in absolute values, displayed respectively in (13) and (14).

$$NSE = 1 - \frac{\sum_{t=1}^{n}[q_{obs}(t)-q_{sim}(t)]^2}{\sum_{t=1}^{n}[q_{obs}(t)-\bar{q}_{obs}]^2} \qquad (13)$$

$$MBE = \frac{1}{n}\sum_{t=1}^{n}(q_{obs} - q_{sim}) \qquad (14)$$

Moreover, we assess the grassland vegetation response to droughts by filtering the NDVI dataset temporally. To calculate the vegetation response to environmental condition we estimate the anomalies (Z-Score) [62] (15). Conceptually, these anomalies represent the intra-seasonal variations of NDVI in response to the fluctuation of the environmental condition (e.g. drought condition) [63].

$$Z - score = \frac{NDVIt - NDVImean,i}{NDVIstdi} \qquad (15)$$

where $NDVI_t$ is the NDVI observed at time step t; $NDVI_{mean,i}$ is the monthly mean of the NDVI daily values; $NDVI_{std,i}$ is the monthly standard deviation of NDVI daily values. A positive or negative value of Z-Score indicates a period wetter or drier than the average, respectively. This helps us identify exceptionally dry periods which can have an important impact on grass production.

In order to evaluate the correlation between the anomalies (Z-Score) simulated by the NDVI models and the observed ones we applied Pearson's correlation (16):

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (16)$$

where rx,y is the correlation coefficient; n is the length of the time series; i study period (in year); xi and yi are the NDVI anomaly respectively, and x and y are the mean value of NDVI. If the value of rx,y is greater than zero, it is a positive

relationship; if rx,y is a negative value, it is a negative relationship; if rx,y is equal to zero there is no relationship between the two variables [64].
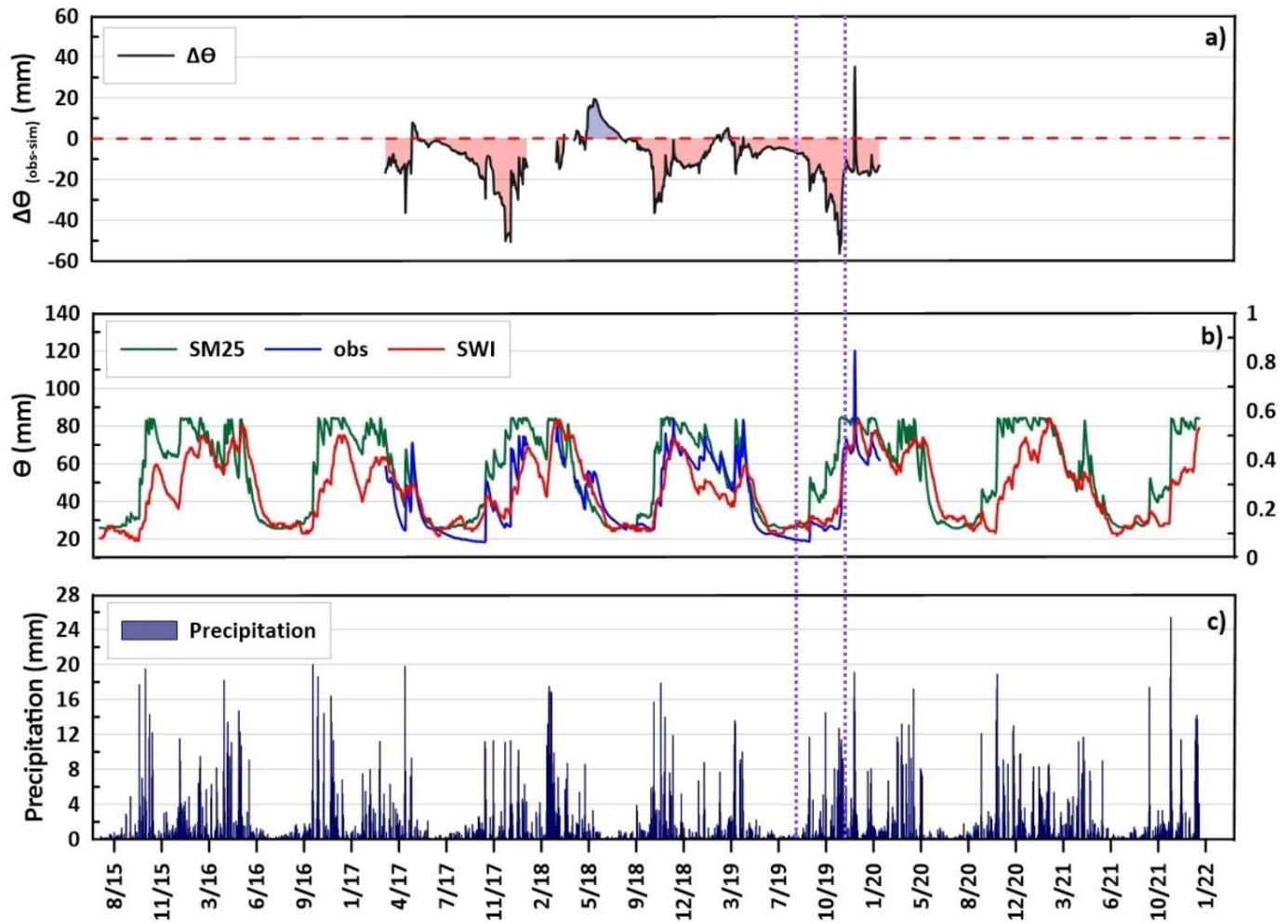
III. Results and Discussion

A .Soil moisture dynamic

The calibration and validation results of the soil moisture model yields NSE values of 0.71 and 0.70 respectively (Supplementary material). This indicates that the model is able to satisfactorily simulate the observed soil moisture. In Fig. 3 we compare the results of the soil moisture dynamic modelled over the study period. Fig. 3a displays the difference between the ground observed soil moisture and the simulated. The results show that the model generally overestimates the

observed values. This can be explained due to the fact that not all the precipitation events are reflected by sensors (e.g. see dotted rectangle in Fig. 3 where wet periods are not reflected in an increase of observed soil moisture). It must be also noted that we are comparing values obtained at different spatial scales, precipitation data and hence model results are pixel values while soil moisture observations are point values.

We observe differences between the two soil moisture products along the study period in how quickly they respond to precipitation. The SM25 reaches higher values of soil moisture (in mm) quicker than SWI, as it is shown in Fig. 3b. This may be due to dissimilarity issue in spatial scales between point (model) and pixel values (satellite observations) [65].

Fig.3. Soil moisture dynamic. Panel a) shows the difference between the ground-assed and modelled soil moisture dynamic. Panel b) displays with the green line the soil moisture model simulations; in the blue line, the observed soil moisture; in the red line, the Soil Water Index (SWI) dynamic of the study period. Panel c) displays the precipitation events satellite-based.
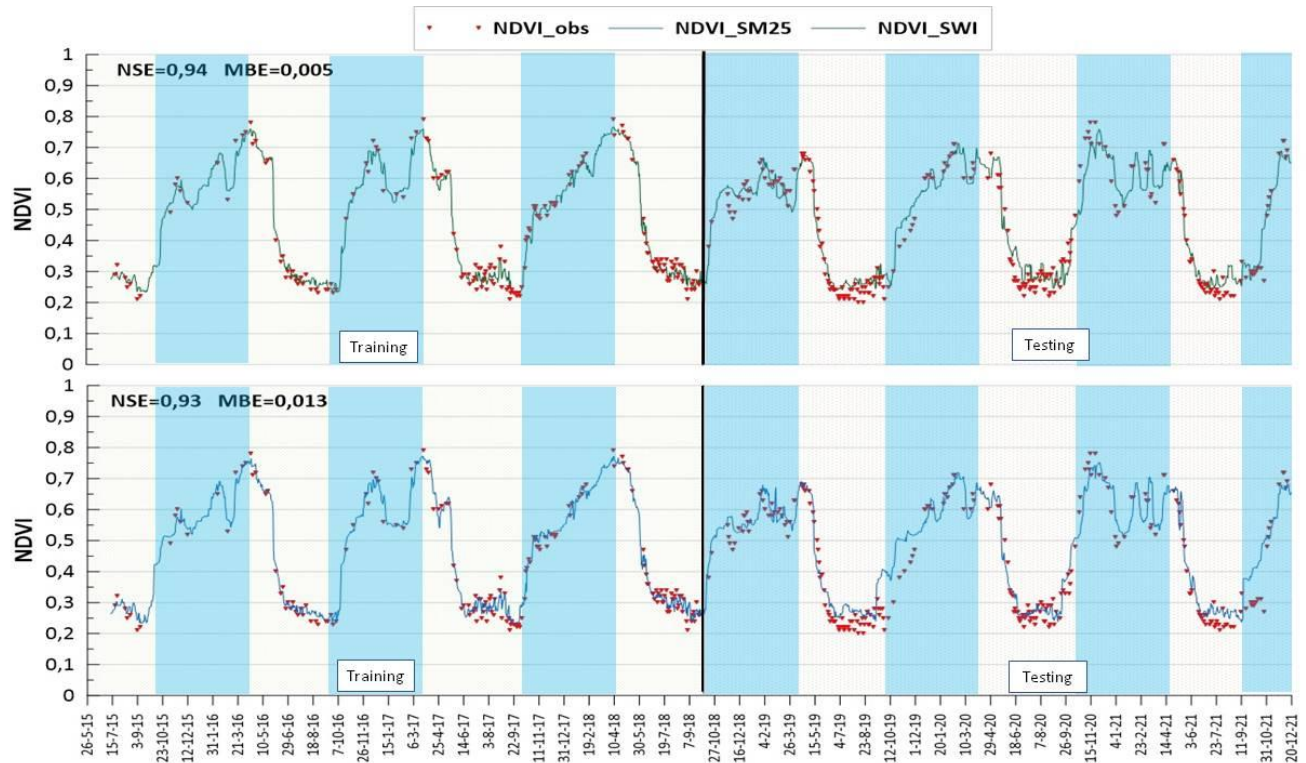
B. NDVI forecasting model
Fig. 4 and 5 show the results of NDVI prediction respectively at 7 and 30 days ahead. For both versions of the NDVI
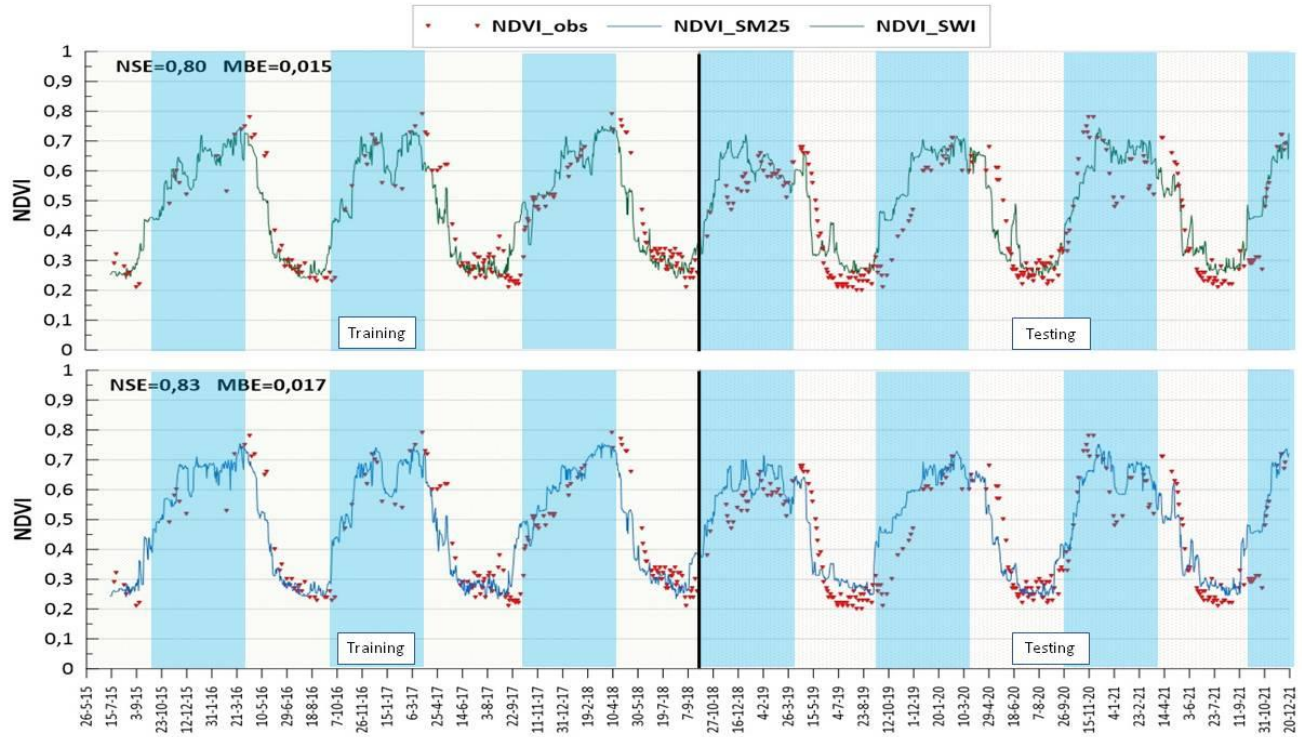
forecast model results are satisfactory. As expected, the 7-day lead time forecasts (NSE over 0.9 and MBE lower than 0.02) are better than the 30-day lead time forecasts (NSE over 0.80 and MBE lower than 0.01). From Fig. 4 and 5, we can break

down the seasonal grassland dynamics into two main stages: the growing season (light blue area in Fig. 4 and 5, when NDVI values raise and reach the highest values and the senescence season when NDVI decreases and reaches the lowest values. In the Mediterranean climate, the grassland growing season is characterized by a fluctuation of NDVI values during the productive season (harvest or grazing season) as a result of dry periods [66], [67], this underline the relationship between the phonology dynamics and soil moisture dynamic [68]. The peak values of NDVI obtained with both forecasting models, at both 7 and 30-day lead times, NDVI between 0.50 and 0.80, are in the order of magnitude of observed values found in the literature in arid and semi-arid climate, which range between 0.53 and 0.78 [20], [69]. These results not only demonstrate the significance of soil moisture as a driver of grassland dynamics in Mediterranean climates but also show the potential use of the two proposed NDVI forecasting models to predict seasonal variations of NDVI. Regarding the intra-seasonal variations or anomalies (z-score), Fig. 6 and 7 show that both NDVI_SM25 and NDVI_SWI for 7-day lead time (r = 0.92, p-value < 0,05, for both models, see Table 1) perform better than for 30-day lead time (NDVI_SWI$_{30}$ r =0,54, NDVI_SM25$_{30}$ r=0.60, p-value < 0,05, Table 1). We observe similar result comparing performances focusing only on the growing season. Indeed, both versions of

the NDVI forecast models at 7-day, NDVI_SWI$_7$ and NDVI_SM25$_7$, showed satisfactory performance recording a high correlation with the observed anomalies (r=0.93 and r=0.92 respectively). Instead, NDVI forecast models at 30-day, do not perform realisable NDVI anomalies during the growing periods. However, the NDVI_SM25$_{30}$ predicts slightly better NDVI anomalies than the NDVI_SWI$_{30}$ (respectively r=0.56 and r= 0.62). Taking as an example the growing season of 2017-2018 with 454 mm of precipitation, and the one from 2018-2019, with 1796 mm (Fig.6 and 7), we can observe how the models perform under particularly dry and wet weather conditions respectively. Under these two conditions, both 7-day forecasting models predict anomalies satisfactorily (r=0.93 for both models), in contrast, forecasting models at 30 days are weakly correlated to the observed anomalies (respectively r=0.59 for NDVI_SWI$_{30}$ and r=0.65 for NDVI_SM25$_{30}$). This shows the limitation of using past and present data to forecast NDVI anomalies in a mid and long term. Future work should explore the use of mid and long-range weather forecast products to improve the performance of this type of NDVI forecasting models. In particular, NDVI_SM25 may benefit by using weather forecast data because to feed the soil moisture model and thus obtain soil moisture forecasts of one or several months.

Fig.4. NDVI forecasts model results vs observations (NDVI_obs) for a 7-day lead time. The top panel displays the forecasts obtained using SWI remote sensed observation as soil moisture information (NDVI_SWI). The bottom panel displays the forecasts obtained using the soil water model (NDVI_SM25). The period corresponding to the growing season is shaded in light blue.

Fig.5. NDVI forecasts model results vs observations (NDVI_obs) for a 30-day lead time. The top panel displays the forecasts obtained using SWI remote sensed observation as soil moisture information (NDVI_SWI). The bottom panel displays the forecasts obtained using the soil water model (NDVI_SM25). The period corresponding to the growing season is shaded in light blue

TABLE I

PEARSON CORRELATION (P-VALUE < 0,05) BETWEEN THE OBSERVED NDVI ANOMALIES AND THE FORECASTED AT 7 AND 30 DAYS . THE OVERALL CORRELATION TAKES INTO CONSIDERATION THE ENTIRE STUDY PERIOD; THE GROWING SEASON (GS) CORRELATION TAKES INTO CONSIDERATION ONLY THE GROWING SEASON OF THE STUDY PERIOD; THE 2017-2018 CORRELATION TAKES INTO CONSIDERATION THE DRIEST GROWING SEASON OF OUR STUDY PERIOD
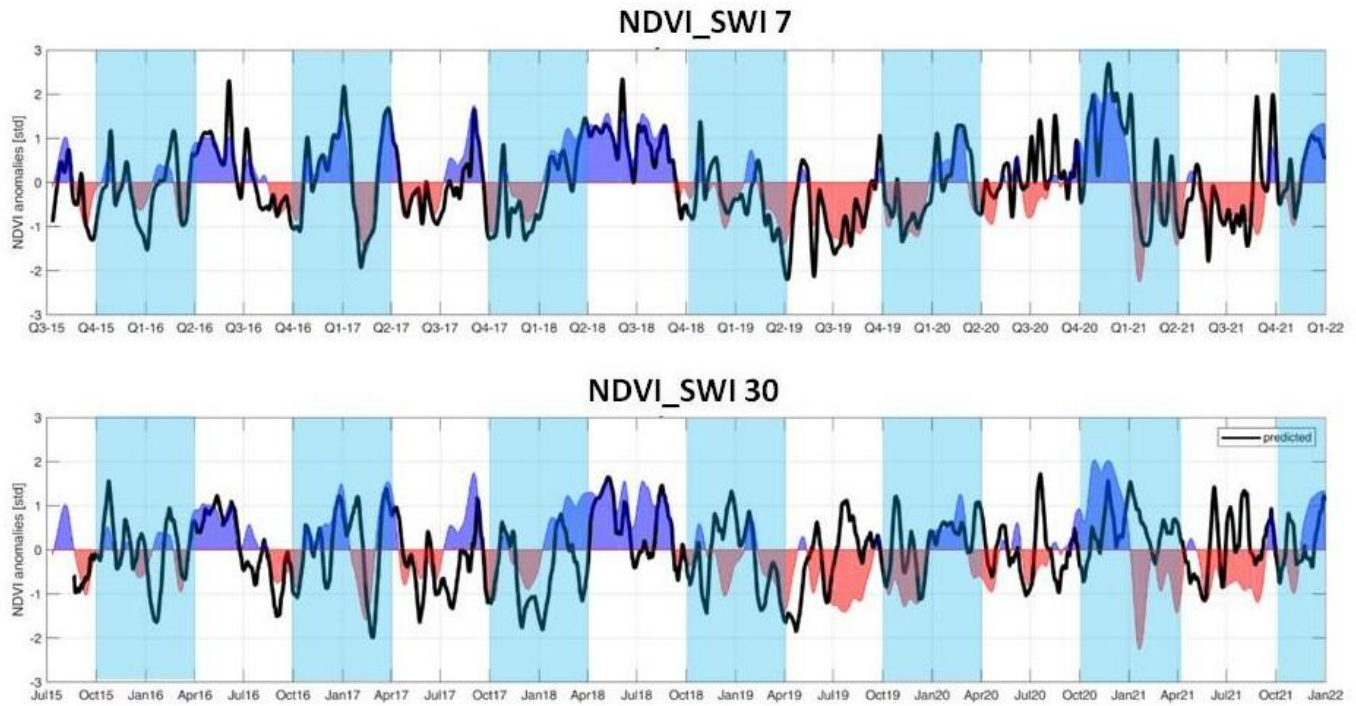
Period	SWI_7	$SM25_7$	SWI_{30}	$SM25_{30}$
Overall	0.92	0.92	0.54	0.60
GS	0.93	0.92	0.56	0.62
GS17-18	0.93	0.93	0.59	0.65

Prediction of NDVI anomalies gains particular importance in the context of agricultural insurance. In Spain, agricultural insurers, under the jurisdiction of the Spanish government, use the NDVI anomaly method to assess grassland yield loss caused by drought or extreme weather events, estimating remotely the production deficit with an NDVI-based indicator called Guaranteed Vegetation Index [14]. It is calculated using data from the last 20 years and during the guaranteed period, which corresponds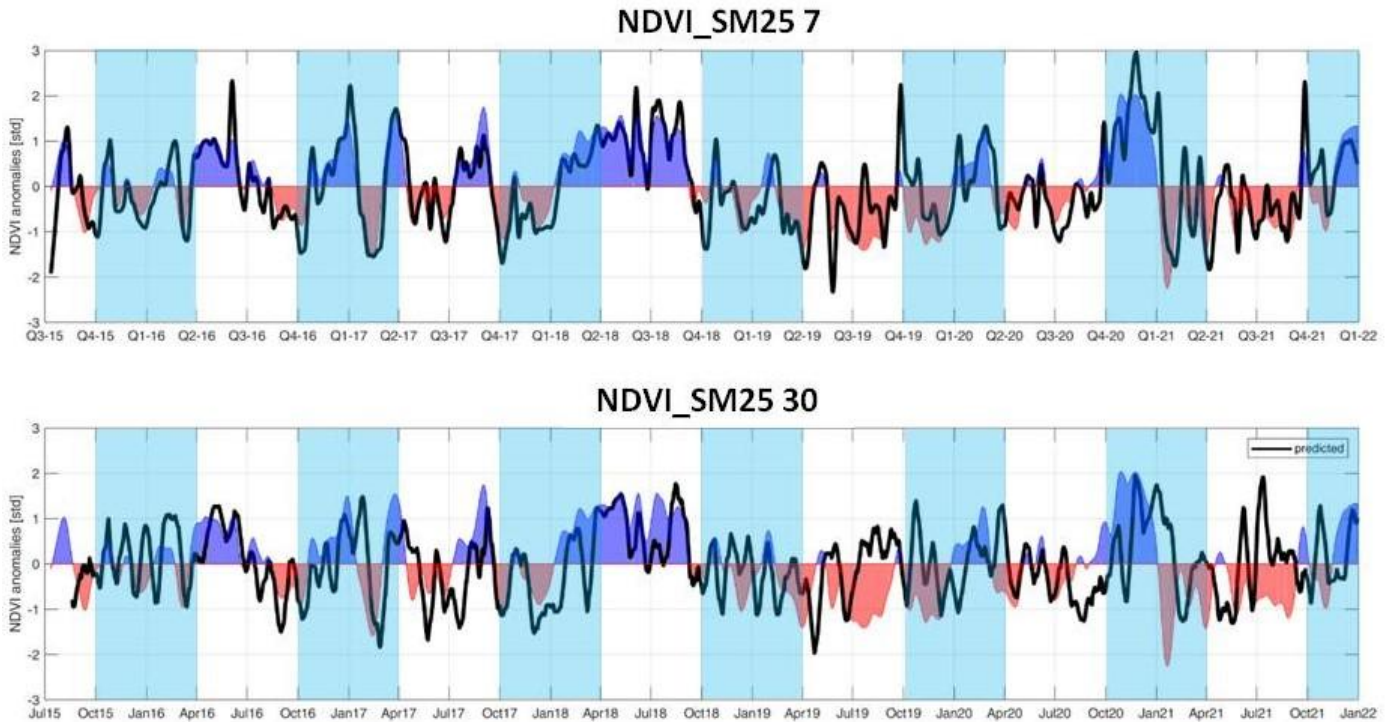 to the growing season, as the 10-day mean NDVI minus 0.5-1.5 times the 10-day standard deviation multiplied by an economical estimator. This model is based on past estimations however the use of the NDVI forecasting models such as the ones presented in this study, can let both farmers and insurers to anticipate production deficits and hence compensations. However, it must be noted that their potential applicability is rather different. In the case of the SWI version, the use of satellite products, increases the potential the scalability of its use, from single pixel scales to larger areas comprising multiple pixels. In the case. The use of the soil moisture model version allows combined with seasonal weather forecast data can potentially increase the temporal scale and potentially obtain better performance for longer lead times than 7 days. While the results of the study are promising, we recognise that there were several limitations, such as one observation point only and the use of historical meteorological data. Considering the limited literature on the use of soil moisture products as NDVI predictors, we advise further investigation into other bioregions and at larger scale. Moreover, the use of stationary weather prediction can be explored to extend the forecasting period to predict anomalies. Field NDVI assessment can be carried out to fit better the models and assess discrepancies with satellite based NDVI observation.

Fig.6. Forecasting model's anomalies of NDVI_SWI. Anomalies are calculated using the Z-score. The black line shows the NDVI anomalies predicted at 7 days (upper graph) and at 30 days (bottom graph). The blue shade shows the positive observed NDVI anomalies, the red shade shows the negative observed NDVI anomalies. The background light shade highlight the growing season.

Fig.7. Forecasting model's anomalies of NDVI_SM25. Anomalies are calculated using the Z-score. The black line shows the NDVI anomalies predicted at 7 days (upper graph) and at 30 days (bottom graph). The blue shade shows the positive observed NDVI anomalies, the red shade shows the negative observed NDVI anomalies. The background light shade highlight the growing season.

IV. CONCLUSION

In this study, we present two NDVI forecasting models based on the use of machine learning and past and present weather and soil moisture data as predictors. One model, NDVI_SM25, uses simulated soil moisture values and the other, NDVI_SWI, uses satellite-based Soil Water Index (SWI) values. The performance of both models is evaluated in a Mediterranean permanent grassland in South Spain by comparing forecasted and observed NDVI daily values. Results show high reliability of models, at 7 and 30-day forecast lead times, in predicting seasonal NDVI dynamics and demonstrate the significance of soil moisture dynamics as a driver of grassland phenology in dry climates. In the case of intra-seasonal variations or anomalies, NDVI are significantly better predicted by both models at a 7-day lead time than at a 30-day lead time. These results show the potential of using NDVI forecasting models based on the of soil moisture information and machine learning to help both farmers and insurers anticipate production deficits and apply mitigation measures.

REFERENCES

[1] M. N. Bugalho, M. C. Caldeira, J. S. Pereira, J. Aronson, y J. G. Pausas, «Mediterranean cork oak savannas require human use to sustain biodiversity and ecosystem services», *Frontiers in Ecology and the Environment*, vol. 9, n.º 5, pp. 278-286, 2011, doi: https://doi.org/10.1890/100084.

[2] F. J. Pulido, M. Díaz, y S. J. Hidalgo de Trucios, «Size structure and regeneration of Spanish holm oak Quercus ilex forests and dehesas: effects of agroforestry use on their long-term sustainability», *Forest Ecology and Management*, vol. 146, n.º 1, pp. 1-13, jun. 2001, doi: 10.1016/S0378-1127(00)00443-6.

[3] I. Hadjigeorgiou, K. Osoro, J. P. Fragoso de Almeida, y G. Molle, «Southern European grazing lands: Production, environmental and landscape management aspects», *Livestock Production Science*, vol. 96, n.º 1, pp. 51-59, sep. 2005, doi: 10.1016/j.livprodsci.2005.05.016.

[4] G. Moreno *et al.*, «Exploring the causes of high biodiversity of Iberian dehesas: the importance of wood pastures and marginal habitats», *Agroforest Syst*, vol. 90, n.º 1, pp. 87-105, feb. 2016, doi: 10.1007/s10457-015-9817-7.

[5] R. L. M. Schils *et al.*, «Permanent grasslands in Europe: Land use change and intensification decrease their multifunctionality», *Agriculture, Ecosystems & Environment*, vol. 330, p. 107891, jun. 2022, doi: 10.1016/j.agee.2022.107891.

[6] J. R. Brown, D. Kluck, C. McNutt, y M. Hayes, «Assessing Drought Vulnerability Using a Socioecological Framework», *Rangelands*, vol. 38, n.º 4, pp. 162-168, ago. 2016, doi: 10.1016/j.rala.2016.06.007.

[7] E. Iglesias, K. Báez, y C. H. Diaz-Ambrona, «Assessing drought risk in Mediterranean Dehesa grazing lands», *Agricultural Systems*, vol. 149, pp. 65-74, nov. 2016, doi: 10.1016/j.agsy.2016.07.017.

[8] X. He *et al.*, «Integrated approaches to understanding and reducing drought impact on food security across scales», *Current Opinion in Environmental Sustainability*, vol. 40, pp. 43-54, oct. 2019, doi: 10.1016/j.cosust.2019.09.006.

[9] J. McDonnell, C. Brophy, E. Ruelle, L. Shalloo, K. Lambkin, y D. Hennessy, «Weather forecasts to enhance an Irish grass growth model», *European Journal of Agronomy*, vol. 105, pp. 168-175, abr. 2019, doi: 10.1016/j.eja.2019.02.013.

[10] M. Trnka, J. Eitzinger, G. Gruszczynski, K. Buchgraber, R. Resch, y A. Schaumberger, «A simple statistical model for predicting herbage production from permanent grassland», *Grass and Forage Science*, vol. 61, n.º 3, pp. 253-271, 2006, doi: 10.1111/j.1365-2494.2006.00530.x.

[11] E. S. Krueger, T. E. Ochsner, M. R. Levi, J. B. Basara, G. J. Snitker, y B. M. Wyatt, «Grassland productivity estimates informed by soil moisture measurements: Statistical and mechanistic approaches», *Agronomy Journal*, vol. 113, n.º 4, pp. 3498-3517, 2021, doi: 10.1002/agj2.20709.

[12] J. Wang, P. M. Rich, K. P. Price, y W. D. Kettle, «Relations between NDVI, Grassland Production, and Crop Yield in the Central Great Plains», *Geocarto International*, vol. 20, n.º 3, pp. 5-11, sep. 2005, doi: 10.1080/10106040508542350.

[13] J. Xue y B. Su, «Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications», *Journal of Sensors*, vol. 2017, p. e1353691, may 2017, doi: 10.1155/2017/1353691.

[14] BOE, *BOE.es - BOE-A-2022-6738 Orden APA/355/2022, de 18 de abril*. 2022. Accedido: 15 de noviembre de 2022. [En línea]. Disponible en: https://boe.es/diario_boe/txt.php?id=BOE-A-2022-6738

[15] Y. Han, Y. Wang, y Y. Zhao, «Estimating Soil Moisture Conditions of the Greater Changbai Mountains by Land Surface Temperature and NDVI», *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, n.º 6, pp.

2509-2515, jun. 2010, doi: 10.1109/TGRS.2010.2040830.

[16] T. Chen, R. A. M. de Jeu, Y. Y. Liu, G. R. van der Werf, y A. J. Dolman, «Using satellite based soil moisture to quantify the water driven variability in NDVI: A case study over mainland Australia», *Remote Sensing of Environment*, vol. 140, pp. 330-338, ene. 2014, doi: 10.1016/j.rse.2013.08.022.

[17] V. García-Gamero, A. Peña, A. M. Laguna, J. V. Giráldez, y T. Vanwalleghem, «Factors controlling the asymmetry of soil moisture and vegetation dynamics in a hilly Mediterranean catchment», *Journal of Hydrology*, vol. 598, p. 126207, jul. 2021, doi: 10.1016/j.jhydrol.2021.126207.

[18] X. Wang, H. Xie, H. Guan, y X. Zhou, «Different responses of MODIS-derived NDVI to root-zone soil moisture in semi-arid and humid regions», *Journal of Hydrology*, vol. 340, n.º 1, pp. 12-24, jun. 2007, doi: 10.1016/j.jhydrol.2007.03.022.

[19] J. A. Escribano, C. G. H. Díaz-Ambrona, L. Recuero, M. Huesca, A. Palacios, y A. M. Tarquis, *Application of Vegetation Indices to Estimate Acorn Production at Iberian Peninsula*. en Pastos 44.2. 2014.

[20] J. R. Insua, S. A. Utsumi, y B. Basso, «Estimation of spatial and temporal variability of pasture growth and digestibility in grazing rotations coupling unmanned aerial vehicle (UAV) with crop simulation models», *PLOS ONE*, vol. 14, n.º 3, p. e0212773, mar. 2019, doi: 10.1371/journal.pone.0212773.

[21] J. A. E. Rodríguez y C. G. H. D. Ambrona, «Estimación de la producción de pastos en dehesas mediante índices de vegetación», en *Los pastos: Nuevos retos, nuevas oportunidades, 2013, ISBN 978-84-695-6999-3, págs. 465-472*, Sociedad Española para el Estudio de los Pastos, 2013, pp. 465-472. Accedido: 4 de octubre de 2022. [En línea]. Disponible en: https://dialnet.unirioja.es/servlet/articulo?codigo=8327116

[22] O. Bounouh, A. M. Tarquis, y I. R. Farah, «Novel Method for Combining NDVI Time Series Forecasting Models», en *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, jul. 2022, pp. 2355-2357. doi: 10.1109/IGARSS46834.2022.9883375.

[23] H. Iwasaki, «NDVI prediction over Mongolian grassland using GSMaP precipitation data and JRA-25/JCDAS temperature data», *Journal of Arid Environments*, vol. 73, n.º 4, pp. 557-562, abr. 2009, doi: 10.1016/j.jaridenv.2008.12.007.

[24] R. Ahmad, B. Yang, G. Ettlin, A. Berger, y P. Rodríguez-Bocca, «A machine-learning based ConvLSTM architecture for NDVI forecasting», *International Transactions in Operational Research*, 2020, doi: 10.1111/itor.12887.

[25] A. Berger, G. Ettlin, C. Quincke, y P. Rodríguez-Bocca, «Predicting the Normalized Difference Vegetation Index (NDVI) by training a crop growth model with historical data», *Computers and Electronics in Agriculture*, vol. 161, pp. 305-311, jun. 2019, doi: 10.1016/j.compag.2018.04.028.

[26] J. J. Casanova, S. A. O'Shaughnessy, S. R. Evett, y C. M. Rush, «Development of a Wireless Computer Vision Instrument to Detect Biotic Stress in Wheat», *Sensors*, vol. 14, n.º 9, Art. n.º 9, sep. 2014, doi: 10.3390/s140917753.

[27] R. Hadria *et al.*, «Derivation of air temperature of agricultural areas of Morocco from remotely land surface temperature based on the updated Köppen-Geiger climate classification», *Model. Earth Syst. Environ.*, vol. 5, n.º 4, pp. 1883-1892, dic. 2019, doi: 10.1007/s40808-019-00645-4.

[28] A. Htitiou, A. Boudhar, Y. Lebrini, R. Hadria, H. Lionboui, y T. Benabdelouahab, «A comparative analysis of different phenological information retrieved from Sentinel-2 time series images to improve crop classification: a machine learning approach», *Geocarto International*, vol. 37, n.º 5, pp. 1426-1449, mar. 2022, doi: 10.1080/10106049.2020.1768593.

[29] A. T. Nieuwenhuizen, L. Tang, J. W. Hofstee, J. Müller, y E. J. van Henten, «Colour based detection of volunteer potatoes as weeds in sugar beet fields using machine vision», *Precision Agric*, vol. 8, n.º 6, pp. 267-278, dic. 2007, doi: 10.1007/s11119-007-9044-y.

[30] T. U. Rehman, Md. S. Mahmud, Y. K. Chang, J. Jin, y J. Shin, «Current and future applications of statistical machine learning algorithms for agricultural machine vision systems», *Computers and Electronics in Agriculture*, vol. 156, pp. 585-605, ene. 2019, doi: 10.1016/j.compag.2018.12.006.

[31] L. A. M. Pereira, R. Y. M. Nakamura, G. F. S. de Souza, D. Martins, y J. P. Papa, «Aquatic weed automatic classification using machine learning techniques», *Computers and Electronics in Agriculture*, vol. 87, pp. 56-63, sep. 2012, doi: 10.1016/j.compag.2012.05.015.

[32] A. Haghverdi, B. G. Leib, R. A. Washington-Allen, P. D. Ayers, y M. J. Buschermohle, «Perspectives on delineating management zones for variable rate irrigation», *Computers and Electronics in Agriculture*, vol. 117, pp. 154-167, sep. 2015, doi: 10.1016/j.compag.2015.06.019.

[33] B. Boydell y A. B. McBratney, «Identifying Potential Within-Field Management Zones from Cotton-Yield Estimates», *Precision Agriculture*, vol. 3, n.º 1, pp. 9-23, mar. 2002, doi: 10.1023/A:1013318002609.

[34] S. Liu, S. Cossell, J. Tang, G. Dunn, y M. Whitty, «A computer vision system for early stage grape yield estimation based on shoot detection», *Computers and Electronics in Agriculture*, vol. 137, pp. 88-101, may 2017, doi: 10.1016/j.compag.2017.03.013.

[35] J. Y. Park, S. Ale, W. R. Teague, y S. L. Dowhower, «Simulating hydrologic responses to alternate grazing management practices at the ranch and watershed scales», *Journal of Soil and Water Conservation*, vol. 72, n.º 2, pp. 102-121, mar. 2017, doi: 10.2489/jswc.72.2.102.

[36] A. Anav *et al.*, «Spatiotemporal patterns of terrestrial gross primary production: A review», *Reviews of Geophysics*, vol. 53, n.º 3, pp. 785-818, 2015, doi: 10.1002/2015RG000483.

[37] J. Xia *et al.*, «Global Patterns in Net Primary Production Allocation Regulated by Environmental Conditions and Forest Stand Age: A Model-Data Comparison», *Journal of Geophysical Research: Biogeosciences*, vol. 124, n.º 7, pp. 2039-2059, 2019, doi: 10.1029/2018JG004777.

[38] K. Dokic, L. Blaskovic, y D. Mandusic, «From machine learning to deep learning in agriculture – the quantitative review of trends», *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 614, n.º 1, p. 012138, dic. 2020, doi: 10.1088/1755-1315/614/1/012138.

[39] S. Xie *et al.*, «A progressive segmented optimization algorithm for calibrating time-variant parameters of the snowmelt runoff model (SRM)», *Journal of Hydrology*, vol. 566, pp. 470-483, nov. 2018, doi: 10.1016/j.jhydrol.2018.09.030.

[40] X. Li, W. Yuan, y W. Dong, «A Machine Learning Method for Predicting Vegetation Indices in China», *Remote Sensing*, vol. 13, n.º 6, Art. n.º 6, ene. 2021, doi: 10.3390/rs13061147.

[41] B. Roy, «Optimum machine learning algorithm selection for forecasting vegetation indices: MODIS NDVI & EVI», *Remote Sensing Applications: Society and Environment*, vol. 23, p. 100582, ago. 2021, doi: 10.1016/j.rsase.2021.100582.

[42] A. Román-Sánchez, T. Vanwalleghem, A. Peña, A. Laguna, y J. V. Giráldez, «Controls on soil carbon storage from topography and vegetation in a rocky, semi-arid landscapes», *Geoderma*, vol. 311, pp. 159-166, feb. 2018, doi: 10.1016/j.geoderma.2016.10.013.

[43] M. C. Peel, B. L. Finlayson, y T. A. McMahon, «Updated world map of the Köppen-Geiger climate classification», *Hydrology and Earth System Sciences*, vol. 11, n.º 5, pp. 1633-1644, oct. 2007, doi: 10.5194/hess-11-1633-2007.

[44] L. Brocca *et al.*, «Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe», *Remote Sensing of Environment*, vol. 115, n.º 12, pp. 3390-3408, dic. 2011, doi: 10.1016/j.rse.2011.08.003.

[45] L. Brocca *et al.*, «Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data», *Journal of Geophysical Research: Atmospheres*, vol. 119, n.º 9, pp. 5128-5141, 2014, doi: 10.1002/2014JD021489.

[46] L. Brocca *et al.*, «SM2RAIN–ASCAT (2007–2018): global daily satellite rainfall data from ASCAT soil moisture observations», *Earth System Science Data*, vol. 11, n.º 4, pp. 1583-1601, oct. 2019, doi: 10.5194/essd-11-1583-2019.

[47] B. Bauer-Marschallinger *et al.*, «Soil Moisture from Fusion of Scatterometer and SAR: Closing the Scale Gap with Temporal Filtering», *Remote Sensing*, vol. 1019, n.º 7, Art. n.º 7, 2018, doi: 10.3390/rs10071030.

[48] W. Wagner, G. Lemoine, M. Borgeaud, y H. Rott, «A study of vegetation cover effects on ERS scatterometer data», *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, n.º 2, Art. n.º 2, mar. 1999, doi: 10.1109/36.752212.

[49] C. Paulik, W. Dorigo, W. Wagner, y R. Kidd, «Validation of the ASCAT Soil Water Index using in situ data from the International Soil Moisture Network», *International Journal of Applied Earth Observation and Geoinformation*, vol. 30, pp. 1-8, ago. 2014, doi: 10.1016/j.jag.2014.01.007.

[50] V. Sheikh, S. Visser, y L. Stroosnijder, «A simple model to predict soil moisture: Bridging Event and Continuous Hydrological (BEACH) modelling», *Environmental Modelling & Software*, vol. 24, n.º 4, pp. 542-556, abr. 2009, doi: 10.1016/j.envsoft.2008.10.005.

[51] R. P. C. Morgan y J. H. Duzant, «Modified MMF (Morgan–Morgan–Finney) model for evaluating effects of crops and vegetation cover on soil erosion», *Earth Surface Processes and Landforms*, vol. 33, n.º 1, pp. 90-106, 2008, doi: 10.1002/esp.1530.

[52] H. Braden y W. Deutscher, «The model AMBETI», Offenbach am Main, 1995. Accedido: 4 de marzo de 2023. [En línea]. Disponible en: https://opendata.dwd.de/climate_environment/CDC/grids_germany/daily/soil_temperature_5cm/AMBETI.pdf

[53] J. G. Arnold *et al.*, «SWAT: Model use, calibration, and validation», *Transactions of the ASABE*, vol. 55, n.º 4, pp. 1491-1508, jul. 2012.

[54] S. Manfreda, M. Fiorentino, y V. Iacobellis, «DREAM: a distributed model for runoff, evapotranspiration, and antecedent soil moisture simulation», en *Advances in Geosciences*, Copernicus GmbH, mar. 2005, pp. 31-39. doi: 10.5194/adgeo-2-31-2005.

[55] D. P. Kroese, T. Brereton, T. Taimre, y Z. I. Botev, «Why the Monte Carlo method is so important today», *WIREs Computational Statistics*, vol. 6, n.º 6, Art. n.º 6, 2014, doi: https://doi.org/10.1002/wics.1314.

[56] V. P. Singh, «Computer models of watershed hydrology», Water Resources Publications, 1995. Accedido: 26 de septiembre de 2022. [En línea]. Disponible en: https://scholar.google.com/scholar_lookup?title=Computer+models+of+watershed+hydrology&author=Singh%2C+V.P.&publication_year=1995

[57] D. Raes, S. Geerts, E. Kipkorir, J. Wellens, y A. Sahli, «Simulation of yield decline as a result of water stress with a robust soil water balance model», *Agricultural Water Management*, vol. 81, n.º 3, pp. 335-357, mar. 2006, doi: 10.1016/j.agwat.2005.04.006.

[58] J. E. Nash y J. V. Sutcliffe, «River flow forecasting through conceptual models part I — A discussion of principles», *Journal of Hydrology*, vol. 10, n.º 3, pp. 282-290, abr. 1970, doi: 10.1016/0022-1694(70)90255-6.

[59] L. Breiman, «Random Forests», *Machine Learning*, vol. 45, n.º 1, pp. 5-32, oct. 2001, doi: 10.1023/A:1010933404324.

[60] G. N. Zaimes, D. Gounaridis, y E. Symenonakis, «Assessing the impact of dams on riparian and deltaic vegetation using remotely-sensed vegetation indices and

Random Forests modelling», *Ecological Indicators*, vol. 103, pp. 630-641, ago. 2019, doi: 10.1016/j.ecolind.2019.04.047.

[61] Z. Li, S. M. Rahman, R. Vega, y B. Dong, «A Hierarchical Approach Using Machine Learning Methods in Solar Photovoltaic Energy Production Forecasting», *Energies*, vol. 9, n.º 1, Art. n.º 1, ene. 2016, doi: 10.3390/en9010055.

[62] A. Klisch y C. Atzberger, «Operational Drought Monitoring in Kenya Using MODIS NDVI Time Series», *Remote Sensing*, vol. 8, n.º 4, Art. n.º 4, abr. 2016, doi: 10.3390/rs8040267.

[63] F. Kogan, A. Gitelson, E. Zakarin, L. Spivak, y L. Lebed, «AVHRR-Based Spectral Vegetation Index for Quantitative Assessment of Vegetation State and Productivity», *Photogrammetric Engineering & Remote Sensing*, vol. 69, n.º 8, pp. 899-906, ago. 2003, doi: 10.14358/PERS.69.8.899.

[64] S. Tong *et al.*, «Spatial and temporal variations of vegetation cover and the relationships with climate factors in Inner Mongolia based on GIMMS NDVI3g data», *J. Arid Land*, vol. 9, n.º 3, pp. 394-407, jun. 2017, doi: 10.1007/s40333-017-0016-4.

[65] A. Gruber, W. a. Dorigo, S. Zwieback, A. Xaver, y W. Wagner, «Characterizing Coarse-Scale Representativeness of in situ Soil Moisture Measurements from the International Soil Moisture Network», *Vadose Zone Journal*, vol. 12, n.º 2, p. vzj2012.0170, 2013, doi: 10.2136/vzj2012.0170.

[66] S. Chelli *et al.*, «The response of sub-Mediterranean grasslands to rainfall variation is influenced by early season precipitation», *Applied Vegetation Science*, vol. 19, n.º 4, pp. 611-619, 2016, doi: 10.1111/avsc.12247.

[67] E. S. Zavaleta *et al.*, «Grassland Responses to Three Years of Elevated Temperature, Co2, Precipitation, and N Deposition», *Ecological Monographs*, vol. 73, n.º 4, pp. 585-604, 2003, doi: 10.1890/02-4053.

[68] P. J. Gómez-Giráldez, M. J. Pérez-Palazón, M. J. Polo, y M. P. González-Dugo, «Monitoring Grass Phenology and Hydrological Dynamics of an Oak–Grass Savanna Ecosystem Using Sentinel-2 and Terrestrial Photography», *Remote Sensing*, vol. 12, n.º 4, Art. n.º 4, ene. 2020, doi: 10.3390/rs12040600.

[69] E. S. Flynn, C. T. Dougherty, y O. Wendroth, «Assessment of Pasture Biomass with the Normalized Difference Vegetation Index from Active Ground-Based Sensors», *Agronomy Journal*, vol. 100, n.º 1, pp. 114-121, 2008, doi: 10.2134/agronj2006.0363.