CORRELATION BETWEEN VEGETATION INDEX AND SOIL MOISTURE INDEX USING SENTINEL-2

Carmelo Alonso[1]*, Pilar López[2], Rosa M. Benito[3] and Ana M. Tarquis[3,4]

[1] Earth Observation Systems, Indra Sistemas S.A., Madrid, Spain e-mail: calonso@indra.es
[2] Dept. of Applied Mathematics, Universidad Complutense de Madrid (UCM), 28040 Madrid, Spain. e-mail: maplopez@bio.ucm.es
[3] Grupo de Sistemas Complejos (GSC), E.T.S.I.A.A.B., Universidad Politécnica de Madrid (UPM), 28040 Madrid, Spain e-mail: rosamaria.benito@upm.es
[4] CEIGRAM, E.T.S.I.A.A.B., UPM, 28040 Madrid, Spain. e-mail: anamaria.tarquis@upm.es, web: http:// www.ceigram.upm.es/

RESUMEN. La dinámica de pastos es el resultado de la interacción entre vegetación, suelo, clima y manejo del terreno. En este trabajo, se estudia la correlación entre un índice de vegetación y otro de humedad del suelo a distintos niveles de agregación. Para ello, imágenes mensuales del Sentinel-2A, desde 7/2015 hasta 8/2016, fueron procesadas para extraer el índice de vegetación de diferencia normalizada (*NDVI*) y el índice normalizado de humedad del suelo (*NSMI*). El área de estudio está en una zona de pastos al norte de la Comunidad de Madrid (España). Los valores del *NDVI* están relacionados con la proporción de vegetación en el píxel. Los píxeles fueron clasificados en: suelo desnudo, cubierto por vegetación y una mezcla de ambos. Las mayores correlaciones se encontraron en suelo desnudo, siendo no significativas en las otras clases. Cuando los datos se agregaron a nivel mensual y estacional, el coeficiente de correlación aumentó significativamente.

ABSTRACT. The study of the dynamics of pasture is the result of a complex interaction between vegetation, soil, climate and man activity. In the present work we study the correlation of vegetation index and soil moisture in a pasture area at different aggregation levels. In order to do so, monthly Sentinel-2A images, from July 2015 till August 2016, were processed to extract Normalized Difference Vegetation Index (*NDVI*) and Normalized Soil Moisture Index (*NSMI*). The area of study is located in a pasture landscape at the north of the Community of Madrid (Spain) *NDVI* positive values are sensitive to the proportion of vegetation into the pixel. Based on this, the pixels where classified in: bare soil, a mixture of vegetation and full vegetated. The highest correlations were found in bare soil being the other two non-statistically significant. When data aggregation was made at month and season scale the R^2 increased significantly.

1.- Introduction

The normalized difference vegetation index (*NDVI*), which is the normalized reflectance difference between the near infrared (NIR) and visible red bands (Tucker, 1979), has been used in drought monitoring and assessment during the last decade (Kogan, 1995; Yang et al., 1998; McVicar and Bierwirth, 2001; Ji and Peters, 2003; Wan et al., 2004; Gu et al., 2007). However, several authors have found that there is a time lag between a rainfall deficit and *NDVI* response (Reed et al., 1994; Di et al., 1994; Rundquist and Harrington, 2000; Wang et al., 2001).

Remote sensing has provided measurement of soil moisture content (SMC) with a consistent spatial and time resolution (Cashion et al., 2005). It is expected that the relation between *NDVI* and SMC are closer in time than with any precipitation index. Root-zone soil moisture controls surface vegetation health conditions and coverage, especially in arid and semi-arid areas, where water is the main controlling factor for vegetation growth (Magagi and Kerr, 2001).

Our aim in this work is to analyse the correlation between *NDVI* and *NSMI* taking in account the season and the fractional vegetation cover.

2.- Material and Methods

2.1.- Site description

A selected area, approximately 6.55 Km^2 (2.56 Km x 2.56 Km), is located in a pasture landscape at the north of the Community of Madrid (Spain) between the municipalities of *Soto del Real* and *Colmenar Viejo* (see Fig. 1). The study area is located between meridians 3º 46' 40'' and 3º 44' 44'' W and parallels 40º 43' 12'' and 40º 42' 36''

The average annual temperature ranges during study period was from 13.8 to 12.7°C, and mean precipitation ranges from 360 to 781 mm. The stations studied were identified semi-arid according to the global aridity index developed by the United-Nations Convention to Combat Desertification (UNEP, 1997). It presented an annual ratio of precipitation and evapotranspiration (P/ETo) between 0.2 and 0.5.

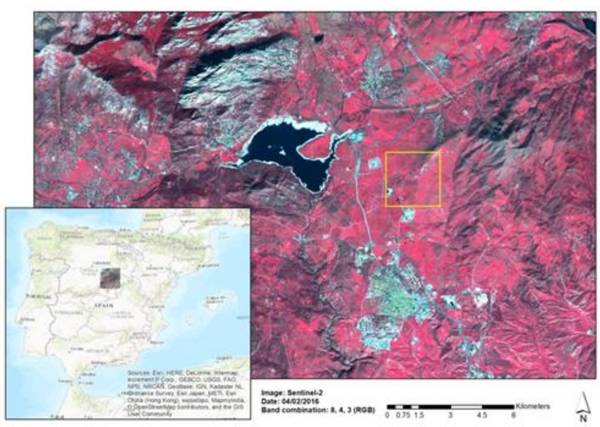

Fig. 1. Area selected in this study with a yellow box between the municipalities of *Soto del Real* and *Colmenar Viejo*.

2.2.- Remote sensing images

In this work we have used images acquired by Sentinel-2A satellite. The Sentinel-2 mission is part of the Copernicus programme of the European Commission, which has been developed and it is operated by the European Space Agency (ESA).

Sentinel-2A satellite was put in orbit in June 2015. It follows a sun-synchronous orbit at 786 km of altitude, with an equatorial crossing time at 10:30 a.m. descending node. Their main sensor for Earth observation is the Multi-Spectral Instrument (MSI) that it is a filter-based push-broom imager and acquires data for thirteen spectral bands in the VNIR and SWIR, with three spatial resolutions. The Table 1 summarize the main characteristics of the MSI spectral bands for Sentinel-2A satellite. The radiometric resolution of the MSI instrument is 12 bits per pixel or 4096 grey levels for the pixel digital value.

Table 1. Main characteristics of the Multi-Spectral Instrument (MSI) spectral bands of Sentinel-2A satellite. The bands marked in bold are the ones used for NDVI (#4 and #8) and for NSMI (#11 and #12).

Band number	Central wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
1	443.9	27	60
2	496.6	98	10
3	560.0	45	10
4	**664.5**	**38**	**10**
5	703.9	19	20
6	740.2	18	20
7	782.5	28	20
8	**835.1**	**145**	**10**
8A	864.8	33	20
9	945.0	26	60
10	1373.5	75	60
11	**1613.7**	**143**	**20**
12	**2202.4**	**242**	**20**

The processing level of the images that we have used is L1C (Gascon et al. 2017). This product corresponds to the Top-Of-Atmosphere (TOA) normalized reflectance in cartographic geometry for each spectral bands. That is, the images are projected in Universal Transverse Mercator (UTM). These images are publicly disseminated by ESA through the Copernicus Open Access Hub (https://scihub.copernicus.eu/dhus/#/home).

Monthly Sentinel-2A images, from July 2015 till August 2016, were processed to extract Normalized Difference Vegetation Index (*NDVI*), with a resolution of 10mx10m, and Normalized Soil Moisture Index (*NSMI*), with a resolution of 20mx20m.

2.2.- Normalized Index Vegetation Index (*NDVI*)

One of the parameter most commonly used to extract the vegetation cover from remote sensing data is the Normalized Difference Vegetation Index or *NDVI* (Tucker, 1979). This index is defined by

$$NDVI = \frac{\rho_{\#8} - \rho_{\#4}}{\rho_{\#8} + \rho_{\#4}} \qquad (1)$$

where $\rho_{\#8}$ is the 8 band (NIR) reflectance and $\rho_{\#4}$ is the 4 band (Red) reflectance of Sentinel-2 data. The original *NDVI* matrix was then passed to a resolution of 20mx20m calculating an average each 2x2 values. For this reason, from a matrix of 256x256 *NDVI* values we passed to 128x128 matrix, each one representing the *NDVI* value of 20mx20m. In this way *NDVI* and *NSMI* presented the same spatial resolution.

The values of this index are within the range {-1, 1}. Their positive values are sensitive to the proportion of vegetation into the pixel (Carlson and Ripley, 1997). That is, the Fractional Vegetation Cover (*FVC*) is a function of the NDVI

$$FVC = \left[\frac{NDVI - NDVI_0}{NDVI_\infty - NDVI_0}\right]^2 \qquad (2)$$

where $NDVI_0$ and $NDVI_\infty$ correspond to the threshold values of NDVI for bare soil and a surface with a *FVC* of 100%, respectively. For these thresholds we are used the values proposed by Raissouni and Sobrino (2000). In this case:

- $NDVI < 0.2$: the pixel is considered without vegetation or bare soil,
- $0.2 \leq NDVI \leq 0.5$: the pixel is composed by a mixture of bare soil and vegetation. The vegetation proportion is calculated with equation [2].
- $NDVI > 0.5$: the pixel is considered as fully vegetated. The vegetation proportion is 100%.

2.3.- Normalized Soil Moisture Index (*NSMI*)

The measurement of soil moisture from optical remote sensing data was analysed by Musik and Pelletier (1986). Their work was based on the Thematic Mapper bands from Landsat-5 and as a result it was established the correlation between the SWIR bands ratio (TM5/TM7) of Thematic Mapper sensor and the moisture content of soil.

Based on the previous result and using the SWIR bands of Sentinel-2A, the Normalized Soil Moisture Index (NSMI) is defined as (Fabre et al. 2015)

$$NSMI = \frac{\rho_{\#11} - \rho_{\#12}}{\rho_{\#11} + \rho_{\#12}} \qquad (3)$$

where $\rho_{\#11}$ is the 11band (SWIR, 1613 nm) reflectance and $\rho_{\#12}$ is the 12 band (SWIR, 2202 nm) reflectance of Sentinel-2 data.

The *NSMI* represents a dimensionless parameter that can be used to quantify gravimetric soil moisture (Haubrock et al. 2008).

2.4.- Descriptive statistics

The first fourth moments of both index values, *NDVI* and *NSMI*, were calculated for each image: average, variance, kurtosis and asymmetry. In this way we could study their temporal variation and the characteristics of their values distribution.

On the other hand, an average value of both indexes per season and year were calculated to check the existence of a significant correlation between them.

3.- Results and Discussion

3.1.- Statistical distribution of *NDVI* and *NSMI*

The *NDVI* values obtained in each image, once that the matrix was aggregated in 20mx20m, are statistically described in Table2. As expected, the *NDVI* median and average show a cyclic pattern through the seasons. The highest variance is achieved from November to April. The kurtosis is especially high during the summer reducing its value in November and December. The asymmetry is minimum in October and November showing absolute values higher than one in the rest of the months. Both statistical moments point out that the distribution of the *NDVI* values doesn't follow a Gaussian shape in agreement with earlier works by Martin-Sotoca et al. (2019) in pasture.

NDVI average values lower than 0.2 are found in July and August. In September and October this value is between 0.2 and 0.5 being the rest of the months higher than 0.5 in general.

Table 2. The median and the first fourth statistical moments of *NDVI* values for each date: average, variance, kurtosis, asymmetry.

Month	Median	Average	Variance	Kurtosis	Asymmetry
July-15	0.1770	0.1838	0.0015	15.70	2.35
August-15	0.1677	0.1740	0.0013	17.18	2.47
September-15	0.2356	0.2100	0.0029	16.20	1.38
October-15	0.3710	0.3600	0.0069	8.89	0.20
November-15	0.5065	0.4870	0.0108	1.59	-0.98
December-15	0.4889	0.4718	0.0102	2.00	-1.07
January-16	0.5100	0.5300	0.0103	3.75	-1.38
February-16	0.5728	0.5512	0.0107	4.37	-1.74
March-16	0.5848	0.5900	0.0110	4.58	-1.75
April-16	0.6291	0.6037	0.0113	4.70	-1.84
May-16	0.5800	0.6000	0.0081	5.94	-1.61
June-16	0.3062	0.3197	0.0062	4.47	1.26
July-16	0.2157	0.2277	0.0033	8.79	1.89
August-16	0.1755	0.1826	0.0017	12.12	1.58

The statistics of the *NSMI* values obtained in each image are shown in Table 2. It shows, as the *NDVI*, a seasonal pattern the median and average reaching the maximum value during spring. Accordingly, the highest variance is achieved in the same season. The kurtosis and asymmetry show lower values if we compare them with the *NDVI* statistics. The kurtosis is high from May to August and the asymmetry is lower than the unit from September till December. Again, both statistical moments point out that the distribution of the *NSMI* values doesn't follow a Gaussian shape. However, the tails of these *NSMI* distributions in each date are shorter as the kurtosis values are lower compared with the *NDVI* distributions.

In general, *NSMI* average values lower than 0.23 are found in July and August. In September and October this value is between 0.23 and 0.25 being the rest of the months higher than 0.25.

Table 3. The median and the first fourth statistical moments of *NSMI* values for each date: average, variance, kurtosis, asymmetry.

Season	Month	Median	Average	Variance	Kurtosis	Asymmetry
Summer	July-15	0.2242	0.2220	0.0006	6.70	-1.36
	August-15	0.2265	0.2251	0.0006	4.74	-1.12
Autum	September-15	0.2356	0.2370	0.0009	3.11	-0.98
	October-15	0.2410	0.2430	0.0010	2.75	-0.85
	November-15	0.2554	0.2517	0.0014	2.11	-0.75
Winter	December-15	0.2586	0.2550	0.0012	2.54	-0.82
	January-16	0.2500	0.2520	0.0130	2.76	-0.90
	February-16	0.2534	0.2496	0.0013	2.82	-0.97
Spring	March-16	0.2700	0.2800	0.0017	3.55	-1.20
	April-16	0.3013	0.2950	0.0019	3.44	-1.33
	May-16	0.2750	0.2700	0.0017	6.23	-1.31
Summer	June-16	0.2641	0.2614	0.0011	4.63	-1.09
	July-16	0.2476	0.2463	0.0010	6.74	-1.58
	August-16	0.2388	0.2377	0.0008	6.67	-1.22

3.2.- Evolution of classified pixels

As an example, the selection of pixels based on their *NDVI* value is showed in Fig. 2. As it can be observed in February, pixels corresponding to *NDVI* values lower than 0.2 are mainly rural roads and buildings. The ones in the range of 0.2 to 0.5 are the surrounding areas to rural roads and with certain slope.

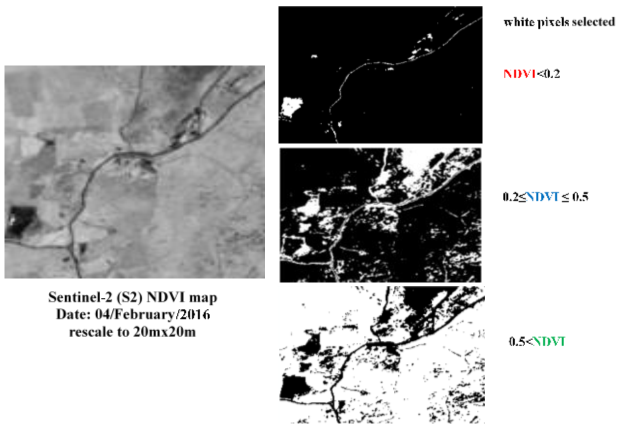

Fig 2. Example of *NDVI* map segmented in three black and white images based on *NDVI* value.

Once that the pixels are segmented based on the *NDVI* value, the *NSMI* values are extracted from the same ones to study their behaviour and correlation.

In Fig. 3 the evolution of both indexes in general and the three sets classified are shown. The range of variation for NDVI_1 is small, from 0.2 to 0.1. However, for NSMI_1 the variation is higher being from 0.22 till 0.14, except at the end of 2016 summer that the value achieves 0.24 due mainly to a rain event in July. In the case of NDVI_2, the variation is higher than NDVI_1, moving in a range of 0.23 – 0.40. Contrary, NSMI_2 shows an almost plane evolution with values between 0.22 and 0.26. From July till

December 2015 NSMI_2 value is almost constant.

Finally, NDVI_3 shows a variation from 0.52 till 0.62 and NSMI_3 varies from 0.26 till 0.32. Observing the number of pixels in each of the sets (Fig.3. C) the average of all the values in NDVI and NSMI reflects the dominant class in each month.

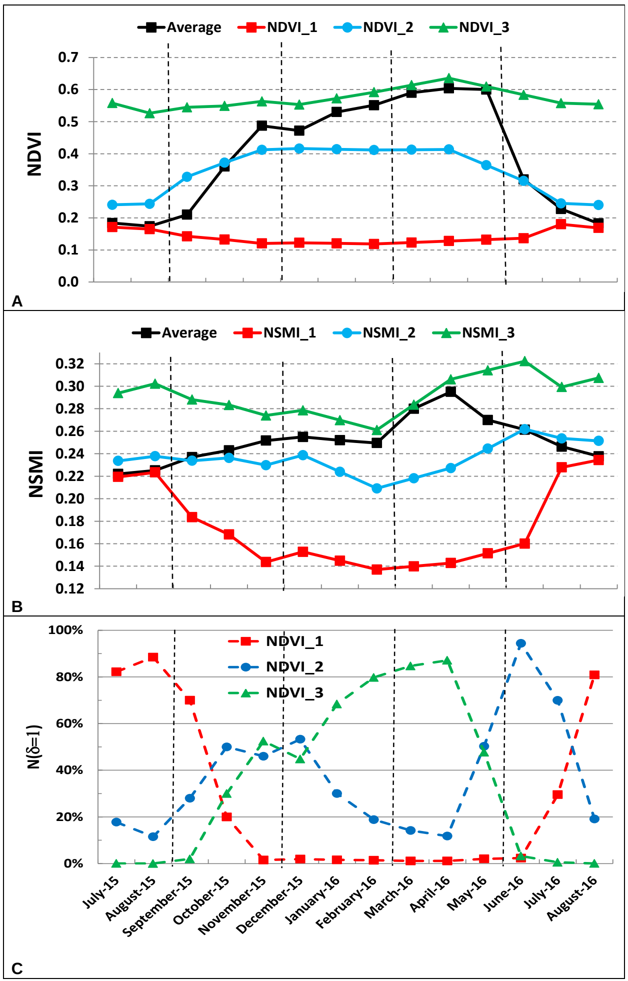

Fig 3. Evolution in time of the average of *NDVI* and *NSMI* maps (A and B respectively). In black the average of all values, in colours the selected pixels based on *NDVI* values: *NDVI*<0.2 (NDVI_1), 0.2≤*NDVI*≤0.5 (NDVI_2) and 0.5<*NDVI* (NDVI_3). (C) Percentage of pixels belonging to each *NDVI* classification.

3.3.- *NDVI* and *NSMI* Correlations

The correlations were made at different levels of aggregation. First, each classified set correlated *NDVI* and *NSMI* values, the next level was at each date and the last level was aggregated the values at each season.

Fig. 4 shows the results obtained at the three levels. When only the pixels with *NDVI*≤0.2 are used (bare soil) the correlation coefficient obtained with *NSMI* is 0.95 and NSMI value is lower than 0.24. In the case that *NDVI*≥0.5, full vegetation cover, there is no correlation and *NSMI* values are over 0.27. In the third case, mix of soil and vegetation, the correlation is not significant and the *NSMI* show a dispersion with values ranging from 0.21 till 0.26.

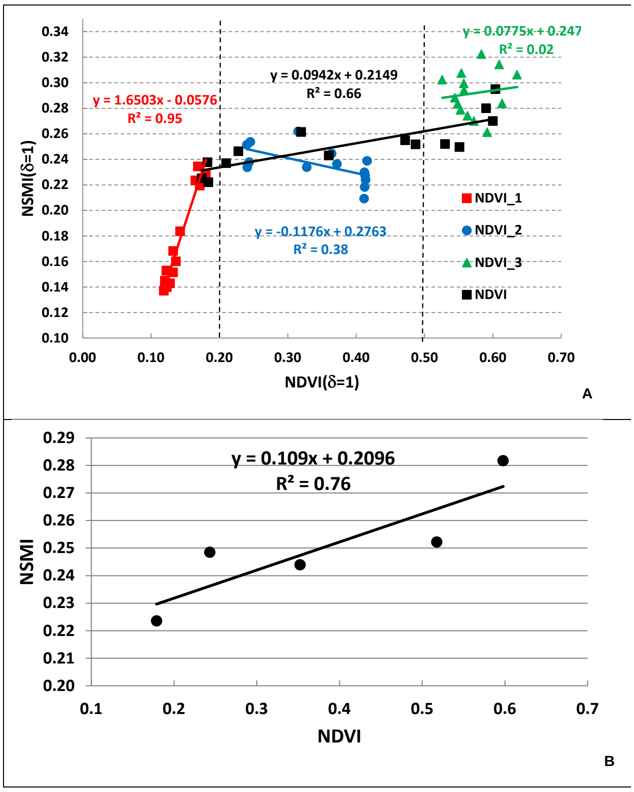

Fig 4. Plot of *NSMI* average versus *NDVI* average. A) For each date, at pixel scale (δ=1), in black the average of all values, in colours the selected pixels based on *NDVI* values: *NDVI*<0.2 (NDVI_1), 0.2≤*NDVI*≤0.5 (NDVI_2) and 0.5<*NDVI* (NDVI_3). (B) *NSMI* and *NDVI* average of all values and per season.

The correlation obtained by date using all the *NDVI* values is significant with an R^2=0.66. This result is obtained thanks to the high correlation of bare soil pixels. When we aggregate the values as season level, see Fig. 4B, the correlation coefficient increases to 0.76.

When a pixel contains a certain coverage by vegetation the correlation between *NDVI* and *NSMI* is lost in a resolution 20 m x 20 m.

6.- Conclusions

The term drought is normally used to refer to deficiency in rainfall, soil moisture, vegetation greenness or ecological conditions. To calculate the impact of a drought is crucial in determining the environmental and agricultural consequences. However, drought intensity varies spatially and temporally pointing out the complexity to study this hazard. Due to this, the use of remote sensing data has been increasingly used as it provides more continuous information in time and space than other approaches.

The *NDVI* has been used in drought monitoring and assessment during the last decade. However, several authors have found that there is a time lag between a rainfall deficit and *NDVI* response and some authors question the correlation between *NDVI* and the soil moisture content measured with the *NSMI*.

In order to establish these comparisons, monthly Sentinel-2A images, from July 2015 till August 2016, were processed to extract *NDVI* and *NSMI*. An area was selected, approximately 6.55 Km^2 (2.56 Km x 2.56 Km), and located in a pasture landscape at the north of the Community of Madrid (Spain) between the municipalities of *Soto del Real* and *Colmenar Viejo*.

NDVI pixels were classified as: bare soil (*NDVI*<0.2), mixture of soil and vegetation and full vegetated (*NDVI*>0.5). Correlations for each set of *NDVI* and the corresponding *NSMI* pixels were calculate with a resolution of 20mx20m as well as without any segmentation. Only a significant correlation was found in pixels of bare soil. When data was aggregated by date and by season using all the pixels the correlation coefficient was again significant obtaining 0.66 and 0.72 respectively.

Acknowledgements. This research has been partially supported by funding from Ministerio de Ciencia, Innovación y Universidades under contract no. PGC2018-093854-B-I00. The authors are also grateful to the Comunidad de Madrid (Spain) and Structural Funds 2014–2020 (ERDF and ESF) for the financial support (project AGRISOSTCM S2018/BAA-4330) and EU project 821964 – BEACON.

7.- References

Carlson, T. N. and D.A. Ripley, 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ.*, 62, 241– 252.

Cashion, J., V. Lakshmi, D. Bosch and T.J. Jackson, 2005. Microwave remote sensing of soil moisture: evaluation of the TRMM microwave imager (TMI) satellite for the Little River Watershed Tifton, Georgia. *J. Hydrol.,* 307, 242–253.

Di, L., D.C. Rundquist and L. Han, 1994. Modelling relationships between NDVI and precipitation during vegetative growth cycles. *Int. J. Remote Sens.*, 15, 2121-2136.

Gascon, F., C. Bouzinac, O. Thépaut, M. Jung, B. Francesconi, J. Louis, V. Lonjou, B. Lafrance, S. Massera, A. Gaudel-Vacaresse, F. Languille, B. Alhammoud, F. Viallefont, B. Pflug, J. Bieniarz, S. Clerc, L. Pessiot, T. Trémas, E. Cadau, R. De Bonis, C. Isola, P. Martimort and V. Fernandez, 2017. Copernicus Sentinel-2A Calibration and Products Validation Status. *Remote Sens*., 9, 584; doi:10.3390/rs9060584.

Gu, Y., J.F. Brown, J.P. Verdin and B. Wardlow, 2007. A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central Great Plains of the United States. *Geophys. Res. Lett.*, 34, L06407.

Haubrock, S.N., S. Chabrillat, C. Lemmnitz and H. Kaufmann, 2008. Surface soil moisture quantification and validation based on hyperspectral data and field measurements. *J. of Applied Remote Sens.*, 2, 023552.

Ji, L. and A.J. Peters, 2003. Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices. *Remote Sens. Environ.*, 87(1), 85-98.

Kogan, F.N., 1995. Droughts of the late 1980s in the United States as derived from NOAA polar-orbiting satellite data. *Bull. Amer. Meteor. Soc.*, 76, 655-668.

Magagi, R.D., and Y.H. Kerr, 2001. Estimating surface soil moisture and soil roughness over semiarid areas from the use of the copolarization ratio. *Remote Sens. Environ.*,75, 432–445.

Martín-Sotoca, J. J., A. Saa-Requejo, R. Moratiel, N. Dalezios, I. Faraslis and A.M. Tarquis, 2019. Statistical analysis for satellite-index-based insurance to define damaged pasture thresholds. *Nat. Hazards Earth Syst. Sci.*, 19, 1685–1702.

McVicar, T. R. and P.N. Bierwirth, 2001. Rapidly assessing the 1997 drought in Papua New Guinea using composite AVHRR imagery. *Int. J. Remote Sens.*, 22(11), 2109-2128.

Musick, H.B. and R.E. Pelletier, (1986). Response of Some Thematic Mapper Band Ratios to Variation in Soil Water Content. *Photogramm. Eng. Remote Sens.* 52, 1661–1668.

Raissouni, N. and J. Sobrino, 2000. Toward Remote Sensing Methods for Land Cover Dynamic Monitoring: Application to Morocco. *Int. J. Remote Sens.*, 21, 353-366.

Reed, B.C., J.F. Brown, D. VanderZee, T.R. Loveland, J.W. Merchant and D.O. Ohlen. 1994. Measuring phenological variability from satellite imagery. *J. of Vegetation Science*, 5, 703-714.

Rundquist, B. C., and J.A. Harrington, 2000. The effects of climatic factors on vegetation dynamics of tallgrass and shortgrass cover. *GeoCarto International*, 15, 33-38.

Tucker, C. J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens Environ.*, 8, 127–150.

UNEP (1997). Word Atlas of Desertification. Second Ed. *United Nations Enviroment Programe*, Nairobi, 182 p.

Wan, Z., P. Wang and X. Li, 2004. Using MODIS land surface temperature and normalized difference vegetation index products for monitoring drought in the southern Great Plains, USA. *Int. J. Remote Sens.*, 25, 61-72.

Wang, Q., and J.D. Tenhunen, 2004. Vegetation mapping with multi temporal NDVI in North Eastern China transect (NECT). *Int. J. Appl. Earth Obs.*, 6, 17-31.

Yang, L., B.K., Wylie, L.L. Tieszen and B.C. Reed, 1998. An analysis of relationships among climate forcing and time-integrated NDVI of grasslands over the US northern and central Great Plains. *Remote Sens Environ.*, 65, 25-37.