Grassland biomass assessment with remote sensing tools and open source software

Alessandro Cimbelli, Valerio Vitale, Stefano Tersigni
Istat, DCAT – Environmental and Territorial Statistics Directorate
Viale Liegi 13, 00198
Rome, Italy
cimbelli@istat.it, vitale@istat.it, sttersig@istat.it

DOI: 10.1481/icasVII.2016.f30d

ABSTRACT

Images collected by optical and radar satellite sensors represent a most viable solution for the extraction of biophysical parameters of the earth surface. The mid-resolution dataset acquired by Landsat and Sentinel satellites have recently become available free of charge for all users. At the same time, some software of image processing and GIS, like *QGIS*, *R*, and *ImageJ*, have reached a high level of maturity and a large community of users, thanks to their open source license. In this project free satellite images and open source software have been used for the assessment of the grassland biomass. The overall goal is the enhancement of the statistics of grassland production and dried fodder for the animal breeding. Currently the National Institute of Statistics collect this kind of dataset at the province level.

The project, still in progress, consists in some "in situ" surveys in a specific site in central Italy and in the building of a regression model between the grassland heights and the corresponding radiometric values of the most relevant image bands.

Keywords: biomass, Sentinel, Landsat, grassland

1. Study areas

During the planning phase, three areas have been identified for the collection of training and test samples, in the North, Central and Southern Italy. A specific mission in the National park of Murge, in Puglia, South of Italy, has shown however the unsuitableness of the local grassland for the model, due to the presence of stones and shrubs over the ground and the low number of pastures in the small, scattered and flatten areas. The collected samples of the grass height have also demonstrated the diversity of vegetation classes and the need to build a complex regression model considering the height of each class, or each group of classes, separately.

The foreseen campaign of measures in Northern Italy has not taken place until now due to limited time and resources. Therefore, the model has only been trained with the two-year campaigns carried out in Central Italy, in a site named Pian Grande.

Pian Grande is an upland karstic plain in the national park of Monti Sibillini, located between Umbria and Marche, in the Appennini mountains, in central Italy. It is a quite suitable site for the grassland production because of the lack of trees and shrubs over the flat plan, located at 1300 mt above sea level. The plain is the bottom of an ancient mountain lake, now dried up, and has a rectangular shape with an area of around 15 Km^2. From the vegetational point of view, it is possible to consider the existence of 4 types of homogeneous areas:

- areas with agricultural crops (grass meadows, barley and lentils) located in the nearby of Castelluccio village and below the slopes of Vettore Mt., corresponding to the class 2.1.1 (Non-irrigated arable land) of CORINE[1] Land Cover 2012 map;
- areas with natural grassland, equivalent to the class 3.2.1;
- areas with pastures, that occupies the central portion of the plain, corresponding to the class 2.3.1 (Pastures);
- bare rocks (3.3.2) or sparsely vegetated areas (3.3.3), located in the nearby of Vettore Mt.

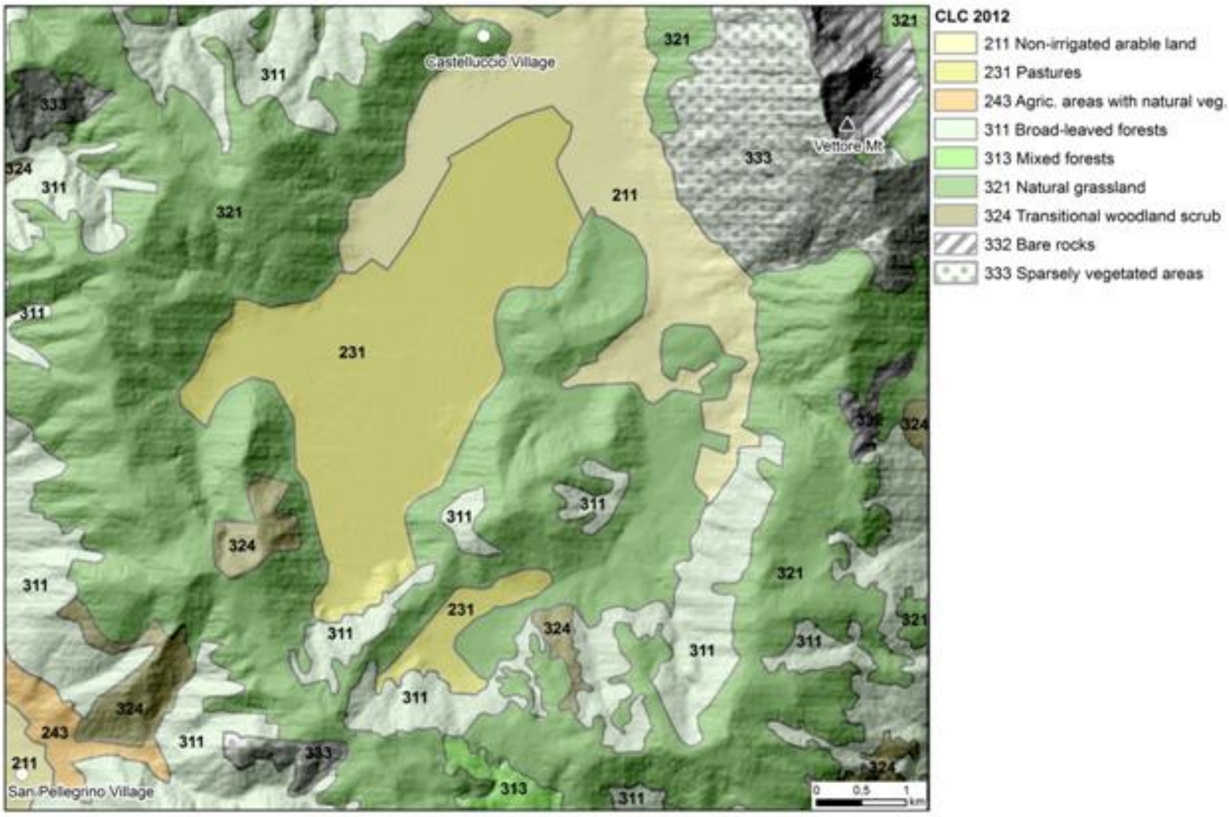

Figure 1: *CORINE Land Cover 2012 map*

The pasture reaches its highest growth at the end of June, depending on the temperatures, after a suggestive blossom, and produces high quality bales of hay. The best period for the field surveys is between the end of June and the beginning of July, just before the cut of the grassland. Two missions have been carried out in the past two years (2015-16) during the same period of the year.

[1] http://www.eea.europa.eu/publications/COR0-landcover

2. Sampling methodology

The need of collecting many samples as possible of the grassland height in a limited time and with the most possible accuracy, taking also into consideration the pixel dimension of the satellite images used for the model, has suggested to the use of the camera and to postprocess the images in order to derive the height with some image processing tools.

The assessment of the height has been executed with a white cardboard of 300 x 70 cm placed on the grass and used as a background for the photograph in order to derive just a linear section of the grass height. Every picture has been linked to the GPS position of the same sample point.

Figure 2: *Survey of a height of a section of grass*

The high contrast of the green vegetation over the white background has helped in the postprocessing of the photographs that has consisted in the extraction, rescaling to a standard width and measure of just the white area (c section in the figure 3) in terms of number of pixels. The division of that area by the fixed width produces its mean height and, by difference with the total height of the cardboard, the mean height of the grass.

The pictures have been analysed with an open source image processing software: *ImageJ*[2]. The software is written in Java and is available for Windows, Linux and Mac OS and OS X. Some of the software tools used in the software for the process are the *Colour Threshold*, for the selection of the white pixels, and the *Colour Pixel Counter*, for the determination of their number. In the 2015 survey a simpler brown cardboard has been used, but sometimes the colour separation and the extraction of grass has been difficult, above all in case of dry hay. Another lesson learnt during the first year of the project is to avoid measures in the second half of July, when the grass is usually cut and compacted in bales.

This method has allowed a very fast survey and a high number of samples collected during each day of measurements. 88 samples have been acquired in a two-day campaign on June 2016, while just 20 were the points surveyed in 2015 because of the partly mown pasture.

[2] https://imagej.nih.gov/ij/index.html , Wayne Rasband, U.S. National Institutes of Health

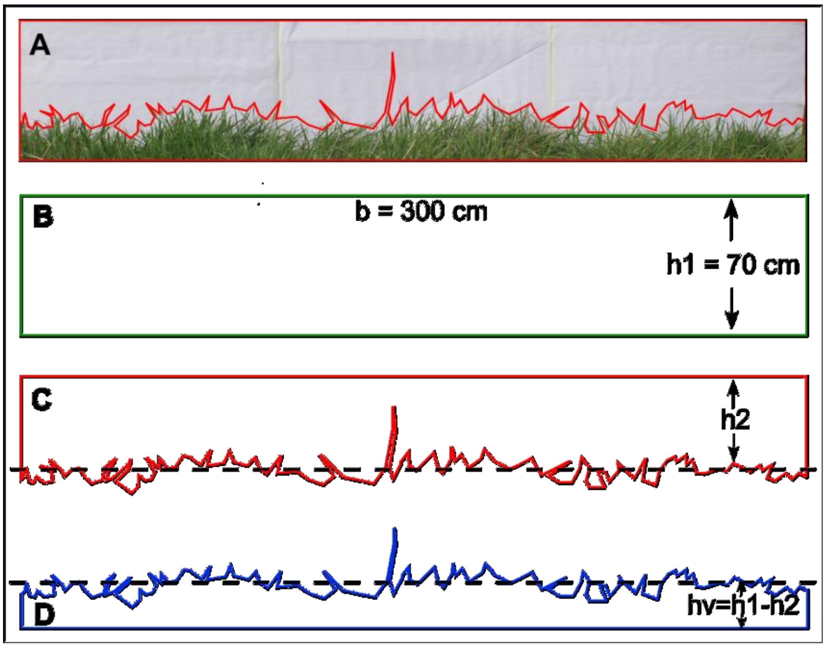

Figure 3: *Postprocessing of the pictures*

3. Satellite images

Currently two mid-resolution satellites acquire images from the earth surface with optical sensors and deliver them free of charge to all users: the Landsat 8 (Nasa) and the Sentinel-2 (Esa).

The OLI sensor of Landsat 8 has replaced the ETM+ one mounted on Landsat 7 but maintains the same bands and the same ground resolution (30 meters) of the previous satellites, allowing in this way the time series analysis. Sentinel-2A is one of the satellites of the Copernicus Esa observation programme and delivers images at a spatial resolution of 10 meters in visible - near infrared spectra. By the end of 2016 the new Sentinel-2B will join the Esa constellation, increasing in this way the revisiting frequency on the same area.

Radar images have also been considered for the correlation model: the Sentinel-1 satellites acquire polarised images in C-band with a spatial resolution of 20 meters. From the first results of the model it seems S-1 images are not so useful for the determination of grass height (see S1_VH and S1_VV in Table 1). Nevertheless, it has been planned to test the model with the high-resolution radar images of the *Cosmo-Skymed* Italian mission.

The values for every image and band have been extracted with the *QGIS*[3] *Zonal Stats* tool with a 20-meters buffer centred on each samples and by extracting basic statistics, like mean and standard deviation, from the overlapping pixel values.

4. Statistical analysis

Descriptive statistics are used to describe the basic features of the data by means of some numeric indexes: mean, standard deviation, standard error of the mean (SEM), coefficient of

[3] http://www.qgis.org/it/site/

variation, skewness and kurtosis. The univariate correlation has been considered too, with the evaluation of the Pearson correlation coefficients and the relative correlation matrix.

Furthermore, a multiple linear regression model has been implemented, based on ordinary least squares method, that is one of the most frequently used statistical approaches to model the correspondence between spectral and field data (Lu, 2006).

The regression model consists of a dependent variable, the vegetation height (Hv) measured in the in-situ survey, and the independent variables, represented by the spectral radiance levels of the bands from Landsat 8 and Sentinel 1-2 satellites. Adjusted R-square are considered as fitting parameters of calculated models.

A backward elimination approaches based on Akaike's Information Criteria (AIC) was used to identify the model that provides the best description of the data using the smallest number of parameters.

4.1. Sentinel

Descriptive statistics and correlation for Sentinel image are shown in Table 1 and Table 2. A positive correlation exists between the S2_NIR and S1_VV while remains negative for all the others bands. Table 3 shows the T-test between the most significant predictors for the model (Red, Green, NIR) and how they differ from each other.

The regression model with the least AIC is:

$$Hv = 71.186 - 0.168 * S2\_G + 0.083 * S2\_R + 0.010 * S2\_NIR$$

The overall significance of the regression model has been estimated with the F-test (F-statistic: 40.62 on 3 and 66 degree of freedom, p-value: 5.42e-15) and the model shows an Adjusted R-square of 0.63.

Table 1: *Descriptive statistics of Sentinel bands and vegetation height variable*

	Mean	Sd	Sem	Cv	Skewness	Kurtosis	n
Hv	12.53726	4.88430	0.58378	0.38958	0.46653	0.06819	70
S2_R	566.23010	72.28901	8.64018	0.12766	0.03663	-0.93344	70
S2_G	834.29860	38.00711	4.54271	0.04555	0.02921	-0.76980	70
S2_B	816.94480	37.14889	4.44014	0.04547	-0.02553	-1.12067	70
S2_NIR	3447.06900	463.68190	55.42058	0.13451	0.28618	-0.80817	70
S2_SWIR1	1994.56300	144.10530	17.22387	0.07224	-0.06513	-0.81735	70
S2_SWIR2	878.25000	106.24390	12.69857	0.12097	-0.10227	-0.95836	70
S1_VH	0.00696	0.00150	0.00018	0.21648	0.04729	-0.30711	70
S1_VV	0.02249	0.00719	0.00085	0.31961	1.54624	3.85916	70

Table 2: *Correlation matrix of Sentinel bands and vegetation height variable.*

	Hv	S2_R	S2_G	S2_B	S2_NIR	S2_SWIR1	S2_SWIR2	S1_VH	S1_VV
Hv	1.00	-0.50	-0.66	-0.50	0.42	-0.58	-0.54	-0.24	0.16
S2_R	-0.50	1.00	0.77	0.95	-0.77	0.86	0.92	0.19	-0.20
S2_G	-0.66	0.77	1.00	0.85	-0.32	0.84	0.74	0.28	-0.06
S2_B	-0.50	0.954	0.85	1.00	-0.60	0.87	0.88	0.24	-0.14
S2_NIR	0.42	-0.77	-0.32	-0.60	1.00	-0.61	-0.78	-0.07	0.32
S2_SWIR1	-0.58	0.86	0.84	0.87	-0.61	1.00	0.95	0.21	-0.20
S2_SWIR2	-0.54	0.92	0.74	0.88	-0.78	0.95	1.00	0.19	-0.24
S1_VH	-0.24	0.19	0.28	0.24	-0.07	0.21	0.19	1.00	0.16
S1_VV	0.16	-0.20	-0.06	-0.14	0.32	-0.20	-0.24	0.16	1.00

Table 3: *Coefficents of the regression model.*

	Estimate	Std. Error	t value	Pr(>|t|)	Signif. codes
(Intercept)	71.186	9.599	7.415	2.94e-10	***
S2_G	-0.168	0.019	-8.576	2.47e-12	***
S2_R	0.083	0.015	5.436	8.51e-07	***
S2_NIR	0.009	0.001	6.168	4.73e-08	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.2. Landsat 8

Descriptive statistics and the correlation matrix for Landsat 8 image have been evaluated as well (see Table 4 and Table 5). The correlation is similar to Sentinel: positive between the grass height and the NIR band while remains negative for all the other bands. Table 6 shows the coefficients obtained with the T-test.

The resulting model is:

$$Hv = 161.094 - 0.042 * S2\_G + 0.024 * S2\_R + 0.002 * S2\_NIR$$

The F-test reports similar results of the previous case (F-statistic: 33.31 on 3 and 66 degree of freedom, p-value: 3.153e-13) and the model shows an Adjusted R-square of 0.58.

Table 4: *Descriptive statistics of Landsat 8 bands and vegetation height variable.*

	Mean	Sd	Sem	Cv	Skewness	Kurtosis	n
Hv	12.53726	4.88430	0.58378	0.38958	0.46653	0.06819	70
S2_R	7930.85964	313.99845	37.52999	0.03959	0.32972	-0.96813	70
S2_G	8984.19961	195.03900	23.31161	0.02170	0.12753	-1.24466	70
S2_B	9094.65477	168.12788	20.09512	0.01848	0.30531	-1.12506	70
S2_NIR	22488.18104	1755.67616	209.84343	0.07807	0.19165	-0.83607	70
S2_SWIR1	13748.75454	719.60525	86.00927	0.05233	-0.18136	-1.05729	70
S2_SWIR2	8844.71064	462.31373	55.25706	0.05227	0.03363	-1.12945	70

Table 5: *Correlation matrix of Landsat8 bands and vegetation height variable.*

	Hv	S2_R	S2_G	S2_B	S2_NIR	S2_SWIR1	S2_SWIR2
Hv	1.00	-0.47	-0.63	-0.43	0.35	-0.53	-0.51
S2_R	-0.47	1.00	0.88	0.96	-0.70	0.85	0.91
S2_G	-0.63	0.88	1.00	0.91	-0.43	0.87	0.86
S2_B	-0.43	0.96	0.91	1.00	-0.56	0.89	0.92
S2_NIR	0.35	-0.70	-0.43	-0.56	1.00	-0.47	-0.65
S2_SWIR1	-0.53	0.85	0.87	0.89	-0.47	1.00	0.97
S2_SWIR2	-0.51	0.91	0.86	0.92	-0.65	0.97	1.00

Table 6: *Coefficents of the regression model.*

	Estimate	Std. Error	t value	Pr(>|t|)	Signif. codes
(Intercept)	161.0937	22.6334	7.118	1.00e-09	***
S2_G	-0.0421	0.0053	-7.937	3.43e-11	***
S2_R	0.0235	0.0042	5.606	4.38e-07	***
S2_NIR	0.0019	0.0003	5.019	4.18e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5. Statistics on grassland production collected by Istat

As part of official statistics, Istat disseminates several statistics on grassland in Italy. The Istat classification of grassland includes the following classes: grass meadow; alternated grassland; grasses; grazings. The Italian Agricultural Census allows to have an analysis with a high spatial detail of grassland areas (at municipal level), in hectares. The inter census sample survey “Farm Structure” provides the same information at the regional level.

The annual estimates of the areas and production quantities of grassland are recorded by: “Estimate of crop, flower and pot plant production and area”. This statistics are produced using expert information. Data are provided by local authorities that collect experts evaluations on area and yield of different crops. The auxiliary information could be included in expert's estimate, such as verifying the availability of external sources (e.g. professional bodies or associations of producers, administrative sources, auxiliary sources of data related to the cultivation being estimated). Crops under investigation are different for each month and take into account the phenological stage of cultivation. For this reason more than one estimate can be determined for each crop during the year (provisional, temporary or permanent).

As part of the official agricultural statistics, more general information are produced by sample survey “Early estimates for main fields crops”, carried out annually in order to provide preliminary estimates of the areas affected by the most interesting crops during the current crop year.

6. Results

The project is still ongoing but the first results show a poor correlation ($R2 \geq 0.6$) between the VNIR bands and the grass height. Radar Sentinel-1 images, even polarised, seems not influenced by different vegetation biomasses of upland pastures, perhaps due to the limited grassland height. From the statistical results seems the higher grass has a bigger component of red while green and NIR decrease. The validation of the model is still to be performed.
However, the sampling method permitted fast operational activities and has given good results. The model seems to work better for mid-height grassland (10-20 cm), perhaps because in case of very low grass the images receive signals from other visible elements, like small stones, dry grass, and the terrain itself, while the higher grass doesn’t change the same radiometric values acquired for mid-height samples.

REFERENCES

Anniballe R., Casa R., Castaldi F., Fascetti F., Fusilli F., Huang W., Laneve G., Marzialetti P., Palombo A., Pascucci S., Pierdicca N., Pignatti S., Qiaoyun X., Santini F., Silvestro P.C., Yang H., Yang G. (2015) Sinergistic use of radar and optical data for agricultural data products assimilation: a case study in central Italy. *International Geoscience and Remote Sensing Symposium 2015 (Milan).*

Grant K, Siegmund R., Wagner M., Hartmann S. (2015) Satellite-based assessment of grassland yields. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7/W3, 15-18.

Iannetta M., Borfecchia F., Ciucci L., Compagnone L., Dibari C., Pedrotti F., Schino G., Trotta C. (2011) Mapping Real Vegetation in the Sibillini National Park (Central Italy): an Application of Satellite Remote Sensing *Colloques Phytosociologiques*, XXIX.

James G., Witten D., Hastie T., Tibshirani R. (2013) An Introduction to Statistical Learning: with Applications in R. Springer. http://www-bcf.usc.edu/~gareth/ISL/ISLR%20First%20Printing.pdf

Lu D.S. (2006) The potential and challenge of remote sensing based biomass estimation. *International Journal of Remote Sensing*, 27, 1297–328.

Schino G., Borfecchia F., De Cecco L., Dibari C., Iannetta M., Martini S., Pedrotti F.(2003) Satellite estimate of grass biomass in a mountainous range in central Italy, *Agroforestry Systems* 59 (2), 157-162.

Sentinel-1 User Handbook (2013) European Space Agency (ESA). https://sentinels.copernicus.eu/documents/247904/685163/Sentinel-1_User_Handbook

Sentinel-2 User Guide (2015) European Space Agency (ESA). https://sentinels.copernicus.eu/web/sentinel/user-guides/document-library

Yunxiang J., Xiuchun Y. , Jianjun Q., Jinya L., Tian G. , Qiong W., Fen Z., Hailong M., Haida Y., Bin X. (2014) Remote Sensing-Based Biomass Estimation and Its Spatio-Temporal Variations in Temperate Grassland, Northern China. *Remote Sensing*, 6, 1496-1513.