Allegato 4a_4

Estimation of temporal nutritional changes of alpine grassland from NDVI data

Luigi Ranghetti¹

¹ Department of Earth and Environmental Sciences (DSTA), University of Pavia, Via Ferrata 9, 27100 Pavia, Italy. E-mail adress: luigi.ranghetti@unipv.it

Introduction

The effects of climate change appear to be especially pronounced in mountain areas. In the Alps, winter is getting warmer and snow-cover duration is decreasing, with sensible effects on plants, which are supposed to advance the beginning of their growing season and speed up their maturation. These alterations can affect the population dynamics of herbivores, such as the Alpine ibex (*Capra ibex*). In particular, ibex females give birth in early summer, in order to fit the best environmental conditions in their most sensitive phase (kid birth and early growth phases): a shift of plant phenology is expected to affect this coupled dynamics.

This changes could concur to explain the strong decrease of ibex populations during the decades 1992-2008, so their experimental confirmations appear to be very important.

Due to the lack of direct field measures useful for this purpose, we need to use remote estimators, as remotely sensed imagery is. In particular, NDVI (Normalised Difference Vegetation Index) is a synthetic index derived from red and near-infrared (NIR) reflectance whose use is rapidly increasing to monitor vegetation and plant responses to environmental change (Pettorelli, Vik et al. 2005). A study about phenological changes within grasslands of Gran Paradiso using an NDVI derived measure (maximum NDVI slope) already exists (Pettorelli, Pelletier et al. 2007); with our study we want to use metrics which are field validated and, when possible, use a finer spatial resolution.

To calculate this index, MODIS derived datasets (MODerate-Resolution Imaging Spectroradiometer) appear to be the best freely available data. This instrument takes images daily, with a spatial resolution of 250 m for near and NIR bands: even if spatial resolution is insufficient to have pure values in a heterogeneous environment such as the alpine grassland, fine temporal resolution allow to construct a good seasonal profile for each pixel. Unfortunately this product exists only since 2000, so to investigate the years before we have to use other data. AVHRR (Advanced Very High Resolution Radiometer) is a sensor which collects continuously daily data since 1981. A dataset of ready to use bimonthly

composite NDVI data is available from 1982 to 2006 (Pinzon et al. 2005; Tucker et al. 2005), but its coarse spatial resolution (near to 4 km) does not allow to distinguish pixels that refer to grassland; so it is possible to obtain only a mean temporal trend.

The aim of this study is to check if alterations in phenology and quality of alpine grasslands have actually occurred in the last 30 years.

To do it, first we test the usability of MODIS data as estimators of the characteristics of grasslands listed before. Another necessary test is the analysis of the possibility to use the two satellite data sets MOD09Q1 (MODIS) and GIMMS (AVHRR) in conjunction into our field area, to verify that an overlap of the temporal series obtained from them could be done without errors.

Then, we tested if the growing season has been shifting in advance during the last 30 years: we did it comparing some phenological metrics (day of beginning of growing season, maximum NDVI value, maximum NDVI slope, and others explained in the next chapter) obtained from the two data set.

Finally, we used these satellite data to make predictions of some nutritional properties of grassland (biomass, protein content, available protein and digestibility). Once we had obtained the temporal series for them, we extract some seasonal synthetic values to better understand how these properties have been changing in the last three decades.

Methods

Field data

All the required field data have been collected within the Gran Paradiso National Park grasslands. For all the details about the field survey, refer to the report of the 4.A.4 action.

With this experimental design we obtained, for each plot, from 4 to 10 samples, for a total of 142 records. A supplementary set has been collected in 2013 in the early part of the season (June and July) within 112 experimental plots. In this case, each sample has been cut within a different plot, to improve spatial distribution of points. Points have been randomly generated on a 250×250 m grid with stratification (5 points in each cell) within alpine grasslands.

Collected samples have been weighted (wet weight), dried in a ventilated oven at 60°C for 48 hours and weighted again (dry weight). Bromatologic analysis have been performed in winter 2012-2013, to obtain, for each sample, relative contents of crude protein (AOAC 1990), neutral detergent fiber (Mertens et al. 2002), acid deterse fiber and lignin (AOAC 2000), digestibility after 24 and 240 hours (Goering and Van Soest 1970).

Satellite data

We used data taken from MODIS TERRA MOD09Q1 250 m data set, validated version V005 (NASA Land Processes Distributed Active Archive Center 2013). This product provides an 8 days composite image with two spectral bands, RED (red band, 620 - 670 nm) and NIR (near infrared band, 841 - 846 nm). NDVI maps have been obtained computing, for each pixel, the value $\frac{\text{RED-NIR}}{\text{RED+NIR}}$. We made correction to each pixel time series in order to avoid cloud noise. In this study we used only the pixels that refer to grasslands surface.

AVHRR data have been taken from GIMMS (Pinzon et al. 2005; Tucker et al. 2005). This freely available data set provides a bimonthly composite image with NDVI values already computed; see Tucker et al. (2005) for details. Each pixel side measures 4'21.81", which at our latitude corresponds to a square of $8082 \times (5674 \pm 22)$ m. We used the 14 pixels which refer to the Park surface.

We calculated also some annual ancillary information to use in substitution to *plot* random effect (its use is not possible, since we want to make spatial interpolations outside the plots). This information have been used also in the comparison between MODIS and AVHRR data.

- *elevation* has been extracted from the 10 m TINITALY digital elevation model (Tarquini, Isola et al. 2007; Tarquini, Vinci et al. 2012);
- North-South aspect $(-\cos(\operatorname{aspect}))$ has been calculated from the elevation;
- the date of beginning of growing season (**BGS**) has been extimated as the day when each seasonal serie reached a thresold relative value t. This approach has been used by Fontana et al. (2008), finding a value $t \approx 0.75$ to detect the day of start of growth. Since this value can not be used with our data, we recomputed t with our data, using a similar approach: we consider as "start of growth" (**BGS**) the day when the mean grass height (measured as described before) exceeds the value of 5 cm, and we compute this value with our 2012 field values, selecting the t which better fit **BGS**. We obtained a value of 0.51. We use the same method also to calculate the threshold for AVHRR data, obtaining a value of 0.57.
- *NDVI maximum value* (**NDVImax**), that is the maximum value between all the daily ones, could provide information about the maximum seasonal productivity.
- The *day of NDVI maximum value* (**Dmax**) represents the moment in which grass stops to increment its biomass and begins to dry.
- The *length of the growing phase* (LGS) is the number of days between BGS and Dmax, and represents the speed of growth of the vegetation.



Figure 1: Surface of the Park covered by APEX data (main figure). The two subfigures show two zoom levels to appreciate the spatial resolution.

• The maximum NDVI slope (MaxSlope) is the maximum first derivative of each seasonal NDVI series, and is another measure of the speed of growth of the vegetation (Pettorelli, Pelletier et al. 2007).

Hyperspectral APEX aerial data

In addition to satellite data, an hyperspectral fine-scale resolution imagery (spatial resolution of 7 m, 284 spectral bands between 0.3993 and $2.4202 \,\mu\text{m}$) has been collected at August 31 with APEX (Airborne Prism EXperiment) sensor on a surface that contains the most of the grasslands of the Park (see figures 1 and 2). Even if this is only a "snapshot" (it is impossible to make phenological studies on it), it is potentially very useful to try a direct correlation with nutritional properties of grassland. Indeed, NDVI (index built on only two spectral bands) correlates with grass"greenness", which correlates with nutritional variables; using all the spectral reflectance curve it is possible to make correlation with chemical content of grass like nitrogen or carbon (Beeri et al. 2007; Mutanga et al. 2004; Starks et al. 2006).

Unfortunately, it has not been possible to make this analysis because APEX data have been delivered



Figure 2: Spectral resolution of APEX data at one sample pixel (NIV-BOR plot).

in considerable delay; however, aerial and field samples are available to do it in the future.

Estimation of nutritional properties from NDVI

All the data analyses have been performed using free software. Statistical analysis have been done with R 3.1.0 (R Core Team 2014) with packages sp (Pebesma and Bivand 2005), rgdal (Bivand et al. 2014), gam (Hastie 2014), nlme (Pinheiro et al. 2014) and MuMln (Bartoń 2014); GDAL 1.11.0 (Warmerdam 2008) and QGIS 2.2.0 (QGIS Development Team 2014) have also been used for input corrections and creation of output maps.

Each properties of grassland has been tested separately with the procedure described below; during this description we refer to each of them generically as "experimental variable" (or Y). We tested these experimental variables: biomass (m), crude protein (CP), neutral detergent fiber (NDF), acid detergent fiber (ADF), lignin (ADL), digestibility at 24 hours (dNDF24), digestibility at 240 hours (dNDF240) and available protein¹ (aCP). m and aCP are absolute measures (expressed in grams), others are relative measures (total content percentage).

To build the model calibrations we perform a model selection from the models derived from

$$Y \sim elevation \cdot aspectNS + BGS + Dmax + NDVImax + NDVI \cdot f(DOS)$$
, random = plot

where **DOS** is hte progressive day from **BGS**, and f(DOS) is the best one between **DOS**, $\text{DOS} + \text{DOS}^2$ and a loess function of **DOS**. The random factor has been removed from the selected model (it has been inserted to prevent problems deriving from the pseudoreplication of the ancillary information).

To validate these models we used two different methods: a leave-by-one cross-validation and a validation with data collected during season 2013.

¹Where $\mathbf{aCP} = \mathbf{m} \cdot \mathbf{CP}$

To perform the validation we use RMSE (Root-Mean-Square Error) and MAE (Mean-Average Error, Willmott and Matsuura 2005), computed as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}$$
 $MAE = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \hat{Y}_i|$

We consider also the residual distributions (mean and standard error) to detect the presence of bias in the predictions.

Finally, we used the validated models to make the predictions at MODIS detail (2000-2013, 250 m spatial resolution, one image every 8 days).

MODIS-AVHRR comparison

Previous models can not be used to make prevision for the whole period of interest (1982-2013), since AVHRR data are not available in a spatial resolution sufficient to directly recognize grasslands. For this reason an indirect correlation (which does not contain elevation and aspect as predictors, but only **BGS**, **Dmax** and **NDVImax**) has been used.

For each year we computed the averaged annual ancillary data within the MODIS grassland pixels, and then we compared these values with the equivalent ones computed on averages of AVHRR selected pixels. Since the two data set are available together only from 2000 to 2006, we obtained a data set of 7 values for each variable.

Correlation between each variable has been tested using pairwise Spearman's rank correlation ρ test, and then grassland values have been computed (where possible) from AVHRR ones with univariate linear models.

Temporal predictions

To obtain the final output we performed the framework below.

Extrapolation of annual synthetic information about nutritional variables. Weekly data are not prompt to snatch temporal variations, so one or more annual synthetic values have been extracted from seasonal series of each variable of interest. First we computed mean values of each variable (**CP.mean, NDF.mean, dNDF24.mean** etc.) between July 1 and September 30 (to avoid temporal extrapolations, we considered only values whose day from the **BGS** was included in the calibration interval). To consider also more punctual values, we extracted values of July 1 (**CP.07, NDF.07** etc.), August 1 (suffix **.08**) and September 1 (suffix **.09**) to have an idea of the situation at the beginning, the central part and the end of the summer. In the case of biomass, we considered also the maximum annual value.

Estimation of averaged annual nutritional variables from averaged phenological metrics Predictability of these annual synthetic metrics from averaged **BGS**, **Dmax** and **NDVImax** has been tested on MODIS data. Models have been chosen selecting the best model (using AIC) between the ones derived from $Y \sim BGS + Dmax + NDVImax$. To check the predictability we used adjusted R² values.

These models have then been used to estimate these metrics from **BGS**, **Dmax** and **NDVImax** derived from AVHRR data.

Analysis of temporal trends Finally, time trends of both phenological predictors than nutritional information have been analysed with a univariate ANOVA. Also general additive models have been built, to check the presence of trends limited to some years.

Output

For the years covered by MODIS data (2000-2013) we provided 8-days maps of each nutritional property with MODIS resolution (250 m). Also, at this spatial resolution 14 annual synthetic maps of phenological (**BGS**, **Dmax** and **NDVImax**) and nutritional data (listed before) have been generated.

The same values have been calculated as averaged temporal series for all the other years (1982-2013). On these series the presence of significant trends has been checked with ANOVA.

Results and discussion

Estimation of nutritional properties from NDVI

The built models generally fit quite well the field data: $adj-R^2 \simeq 0.5$ in all the cases, with the positive exception of crude protein ($adj-R^2 = 0.79$) and the negative one of digestibility at 240 hours ($adj-R^2 = 0.33$).

Using 2012 data set for the validation we obtained results that confirmed the calibrative metrics (RMSE are quite similar to MSE, and mean values of ΔY are close to zero). With the exception of digestibility at 240 hours and available protein, with all the other variables $1 - \frac{\Delta RMSE \ 2012^2}{St. \ dev.^2} > 0.40$.

2013 data set shows different results. Globally, we obtained greater RMSE and MAE values than with 2012 data, so this data set worst support the models. In particular, acid detergent fiber and lignin present RMSE values that are even greater than standard deviation ones. The only variable in which results are better than in 2012 is the biomass. Mean differences between measured and predicted values confirm these evidences: in the case of acid detergent fiber, lignin and digestibility at 24 hours, their values are greater than their standard deviations, suggesting the presence of a bias.

Figure 3 shows 2012 and 2013 predicted values. Acid detergent fiber and lignin predicted values for 2013 are clearly biased, while in the other cases that is not evident. Also, it is evident how crude



Figure 3: Predicted values of each variable compared to the measured ones. Values of 2012 data set are marked with "+", 2013 ones with " \times ".



Figure 4: Comparison between MODIS (averaged grassland pixels) and AVHRR (averaged pixels) values of **BGS**, **Dmax**, **NDVImax**, **LGS** and **MaxSlope** (*rho* value of **Dmax** is the one obtained without the 2000).

protein predicted values better fit the measured ones respect to other variables, and how residual variance of digestibility at 240 hours is higher than in the other cases.

For this reason, the variables considered as estimable (and so considered in the next part) are protein and NDF content, biomass and digestibility at 24 hours.

MODIS-AVHRR comparison

Figure 4 shows the relations between considered ancillary data: it is possible to see that only **BGS**, **Dmax** and **NDVImax** are correlated enough to allow the estimation of the values of grassland from AVHRR data. So, **LGS** and **MaxSlope** will not be used to perform forecasts.

In the case of **Dmax**, using all the values a $\rho = 0.39$ would be obtained; however, removing 2000 value it increase up to 0.95 Considering the great increase of ρ obtained and the fact that AVHRR 2000 seasonal series is quite incomplete (it starts from March), in this case – and only in this case – the model has been built without that value and considered suitable.

Spatial predictions

Maps of each phenological and nutritional metric have been produced yearly (from 2000 to 2013) as GeoTiff, both in the original MODIS grid and rescaled in the one in use by the surveillance service of Gran Paradiso National Park. At the end of this document, 2001 maps have been shown as example.

Temporal predictions

Almost all the mean summer nutritional variables have been well estimated from averaged phenological metrics: R² of **CP.mean**, **NDF.mean**, **dNDF24.mean** and **m.mean** are respectively equal to 0.90, 0.98, 0.98 and 0.77. All of them used **BGS** and **NDVImax** as estimator; **Dmax** has never been selected.

Punctual values of protein, fiber and digestibility provided similar results; biomass resulted well estimable in July ($R^2 = 0.97$), less in August ($R^2 = 0.50$) and not estimable in September (null model had a smaller AIC than all the ones of the possible models). Maximum annual biomass resulted quite well estimable ($R^2 = 0.82$).

The analysis of temporal trends without 2013 value (see figure 5) shows that, between predictors, **BGS** is the only one which presents a significant (decreasing) trend (see graphs for $\frac{dY}{dT}$ and t values). Also **Dmax** presents the same mean decrease, but standard deviation is bigger and so this decrease is not statistically significant; however, it suggests that the anticipation is not just about the beginning, but almost all the first half of the growing season. **NDVImax** does not show any temporal trend.

Indeed, between nutritional variables it is possible to see that all of them present a significant increasing (biomass and fiber) or decreasing (protein and digestibility) trend, if we consider their mean summer value.

Considering separate trends of July 1, August 1 and September 1, we see that crude protein content decreases in July and August $(-4e - 04 \pm 0.016\% \text{ and } -1e - 04 \pm 0.006\%)$ but increases in September $(1e - 04 \pm 0.039\%)$; as biomass do $(0.437 \pm 0.024 \text{ g}, 0.295 \pm 0.008 \text{ g} \text{ and } 0.016 \pm 0.777 \text{ g})$. Neutral detergent fiber and digestibility present constant trends $(7e - 04 \pm 0.038\%, 7e - 04 \pm 0.038\% \text{ and } 7e - 04 \pm 0.036\%$ the first; $-0.001 \pm 0.022\%, -0.001 \pm 0.016\%$ and $-0.001 \pm 0.018\%$ the second).

However, this trends do not look as constant in all the 3 decades, but concentrated in the period 1991-1997. Also, in all the cases with the exception of biomass it is possible to notice an inversion in trends during the last decade.

Conclusions

This study showed that, with the use of freely available data (MODIS imagery and a digital terrain model), it is possible to estimate the protein content of an alpine grassland, but only computing



Figure 5: Temporal series of three phenological predictors (**BGS**, **Dmax** and **NDVImax**) and the six nutritional synthetic metrics computed. $\frac{dY}{dT}$ and t values refer to models built whitout 2013; dotted lines indicate 95% confidence intervals. Phenological predictors values are indicated with \times when they are real averaged MODIS grassland values, with + if predicted from AVHRR (little ones are predicted values not entered into the models because the real value was available).

ancillary information: a direct correlation between NDVI values and bromatologic parameters have not been shown, so a use of vegetation indexes as a direct estimator of the nutritional content could lead to errors. To improve the predictability of other variables, the integration with more phenological and climatic information appear to be useful.

Analysis of seasonal trends of phenological parameters (beginning of growing season, day of maximum NDVI value and maximum NDVI value) revealed a phenological anticipation of the growing season. Our models can convert this information into different changes in predictable nutritional variables: significant increases in summer biomass (both mean and maximum values) and fiber content are associated to a decrease of the protein content and the digestibility of grassland.

However we have suggestions that in the last decade an inversion of this tendence has been going: future predictions will confirm or deny that.

References

- AOAC (1990), 'Protein (crude) in animal feed (976.06)', Official Methods of Analysis of the Association of Official Analytical Chemists. 15th Ed., p. 72.
- (2000), 'Fiber (Acid Detergent) and Lignin in Animal Feed (973.18)', Official Methods of Analysis of the Association of Official Analytical Chemists. 17th Ed., p. 72.
- Bartoń, K. (2014), *MuMIn: Multi-model inference*, R package version 1.9.26/r268, URL: http://R-Forge.R-project.org/projects/mumin/.
- Beeri, O., R. Phillips, J. Hendrickson, A. B. Frank and S. Kronberg (2007), 'Estimating forage quantity and quality using aerial hyperspectral imagery for northern mixed-grass prairie', *Remote Sensing of Environment*, 110, 2, pp. 216–225,
- Bivand, R., T. Keitt and B. Rowlingson (2014), rgdal: Bindings for the Geospatial Data Abstraction Library, R package version 0.8-16, URL: http://CRAN.R-project.org/package=rgdal.
- Fontana, F., C. Rixen, T. Jonas, G. Aberegg and S. Wunderle (2008), 'Alpine grassland phenology as seen in AVHRR, VEGETATION, and MODIS NDVI time series - A comparison with in situ measurements', English, *Sensors*, 8, 4, pp. 2833–2853,
- Goering, H. and P. Van Soest (1970), 'Forage Fiber Analysis. USDA Agricultural Research Service. Handbook number 379', US Department of Agriculture. Superintendent of Documents, US Government Printing Office, Washington, DC.
- Hastie, T. (2014), gam: Generalized Additive Models, R package version 1.09.1, URL: http://CRAN.R-project.org/package=gam.
- Mertens, D., M. Allen, J. Carmany, J. Clegg, A. Davidowicz, M. Drouches, K. Frank, D. Gambin,M. Garkie, B. Gildemeister, D. Jeffress, C.-S. Jeon, D. Jones, D. Kaplan, G.-N. Kim, S. Kobata,

D. Main, X. Moua, B. Paul, J. Robertson, D. Taysom, N. Thiex, J. Williams and M. Wolf (2002), 'Gravimetric determination of amylase-treated neutral detergent fiber in feeds with refluxing in beakers or crucibles: Collaborative study', English, *Journal of AOAC International*, 85, 6, pp. 1217– 1240,

- Mutanga, O., A. Skidmore and H. Prins (2004), 'Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features', *Remote Sensing of Environment*, 89, 3, pp. 393–408,
- NASA Land Processes Distributed Active Archive Center (2013), *MOD09Q1*, USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, URL: https://lpdaac.usgs. gov/products/modis_products_table/mod09q1.
- Pebesma, E. J. and R. Bivand (2005), 'Classes and methods for spatial data in R', *R News*, 5, 2 (Nov. 2005), pp. 9–13,
- Pettorelli, N., F. Pelletier, A. von Hardenberg, M. Festa-Bianchet and S. D. Côté (2007), 'Early onset of vegetation growth vs. rapid green-up: Impacts on juvenile mountain ungulates', English, *Ecology*, 88, 2, Appendix: http://www.esapubs.org/archive/ecol/E088/023/appendix-A.htm, pp. 381–390,
- Pettorelli, N., J. Vik, A. Mysterud, J.-M. Gaillard, C. Tucker and N. Stenseth (2005), 'Using the satellite-derived NDVI to assess ecological responses to environmental change', English, *Trends in Ecology and Evolution*, 20, 9, pp. 503–510,
- Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar and R Core Team (2014), nlme: Linear and Nonlinear Mixed Effects Models, R package version 3.1-117, URL: http://CRAN.R-project.org/package= nlme.
- Pinzon, J., M. E. Brown and C. J. Tucker (2005), 'Satellite time series correction of orbital drift artifacts using empirical mode decomposition', in *Hilbert-Huang Transform: Introduction and Applications*, ed. by N. Huang, World Scientific, pp. 167–186.
- QGIS Development Team (2014), QGIS Geographic Information System, Open Source Geospatial Foundation, URL: http://qgis.osgeo.org.
- R Core Team (2014), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, URL: http://www.R-project.org/.
- Starks, P., D. Zhao, W. Phillips and S. Coleman (2006), 'Development of canopy reflectance algorithms for real-time prediction of bermudagrass pasture biomass and nutritive values', English, Crop Science, 46, 2, cited By (since 1996)25, pp. 927–934,
- Tarquini, S., I. Isola, M. Favalli, F. Mazzarini, M. Bisson, M. Pareschi and E. Boschi (2007), 'TIN-ITALY/01: A new Triangular Irregular Network of Italy', English, Annals of Geophysics, 50, 3, cited By (since 1996)41, pp. 407–425,

- Tarquini, S., S. Vinci, M. Favalli, F. Doumaz, A. Fornaciai and L. Nannipieri (2012), 'Release of a 10m-resolution DEM for the Italian territory: Comparison with global-coverage DEMs and anaglyphmode exploration via the web', English, *Computers and Geosciences*, 38, 1, cited By (since 1996)10, pp. 168–170,
- Tucker, C., J. Pinzon, M. Brown, D. Slayback, E. Pak, R. Mahoney, E. Vermote and N. El Saleous (2005), 'An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data', English, *International Journal of Remote Sensing*, 26, 20 (Oct. 2005), pp. 4485–4498,
- Warmerdam, F. (2008), 'The Geospatial Data Abstraction Library', in Open Source Approaches in Spatial Data Handling, ed. by G. Brent Hall and M. G. Leahy, Springer Berlin Heidelberg, pp. 87– 104.
- Willmott, C. and K. Matsuura (2005), 'Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance', English, *Climate Research*, 30, 1, pp. 79–82,



Figure 6: Examples of output maps (1/3).



Figure 7: Examples of output maps (2/3).



Figure 8: Examples of output maps (3/3).