

# Proximal Sensing for modeling development curves and genetic parameter estimation in alfalfa

Ranjita Thapa<sup>1</sup>, Nicholas Santantonio<sup>2</sup>, Nicolas Morales<sup>1</sup>, Julie Hansen<sup>1</sup>, Liam Wickes-Do<sup>1</sup>, Ian Ray<sup>3</sup>, Christopher Pierce<sup>3</sup>, Michael A Gore<sup>1</sup>, Virginia Moore<sup>1</sup>, and Kelly Robbins<sup>1</sup>

<sup>1</sup>Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University

<sup>2</sup>School of Plant and Environmental Sciences, College of Agriculture and Life Sciences, Virginia Tech Blacksburg

<sup>3</sup>College of Agricultural, Consumer, and Environmental Sciences, New Mexico State University

October 25, 2022

## Proximal Sensing for modeling development curves and genetic parameter estimation in alfalfa

Ranjita Thapa<sup>1</sup>, Nicholas Santantonio<sup>2</sup>, Nicolas Morales<sup>1</sup>, Julie Hansen<sup>1</sup>, Liam Wickes-Do<sup>1</sup>, Ian Ray<sup>3</sup>, Christopher Pierce<sup>3</sup>, Michael A. Gore<sup>1</sup>, Virginia Moore<sup>1</sup>, Kelly Robbins<sup>1</sup>

<sup>1</sup>Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853; <sup>2</sup>School of Plant and Environmental Sciences, College of Agriculture and Life Sciences, Virginia Tech Blacksburg, VA 24060; <sup>3</sup>College of Agricultural, Consumer, and Environmental Sciences, New Mexico State University

### Abstract

Vegetative indices (VIs) collected from an unoccupied aerial vehicle (UAV) equipped with a multi-spectral camera can be used to study growth and development of alfalfa throughout each growth cycle. Random regression models are well suited to fit longitudinal phenotypes such as VIs collected over time to estimate growth curves using covariance functions. Using these functions genetic variation in growth through time can be estimated and the relationships between VIs and end-use traits, like forage yield and quality, can be assessed. The main objectives of this project are (1) to incorporate aerial high-throughput phenotyping to predict performance and genetic merit of the breeding materials, (2) to fit longitudinal random regression models to estimate genotype-specific growth curves and estimate the heritability of key growth parameters. Univariate and multivariate models were used to estimate heritability of image features for alfalfa trial of Helfer, 2020 and 2021. The heritability of different image features in alfalfa ranged from 0 - 0.78. The preliminary results showed the strongest correlation for Green NDVI and biomass yield (0.4053, 0.7875, and 0.6779), followed by Red edge NDVI and biomass yield (0.417, 0.7898, and 0.6417) for the first, second and third cuttings respectively of the experimental trial located at Helfer, Ithaca for 2020, while the genetic correlations for 2021 were strongest for Red edge NDVI and biomass yield (0.76, 0.74, and 0.66) followed by Green NDVI and biomass yield (0.75, 0.76 and 0.60) for the first, second and third cuttings. The potential of random regression models was investigated using Legendre polynomial functions. Random regression model converged for most of the time points and showed potential for modeling genetic parameters associated with growth and development.

Keywords: High-throughput phenotyping (HTP), multi-spectral imagery, random regression model

### Introduction

The genetic gain in alfalfa has approached stagnation in the past few decades limiting the benefits to alfalfa farmers. Additionally, the complexity of the genomes and a high phenotypic burden has limited the adoption of new breeding technologies. Evaluation of breeding material requires multiple harvests per year for multiple years, limiting the size and number of field trials. The low heritability of forage yield also demands extensive replications, further limiting the breeding materials to be evaluated whereas the ability to screen more materials will lead to selection efficiency.

The alfalfa phenotype (P) is determined by its genotype (G), environment (E) and the interaction between genotype and environment (G×E) i.e.  $P = G + E + G \times E$ . Therefore, variety trials in an alfalfa breeding program are usually conducted in multiple environments to minimize the risk of discarding the genotypes that potentially perform poor in some but not in all environments due to the significant  $G \times E$ . Evaluation of the genotypes in multiple environments is important in breeding program to get the insight of  $G \times E$  and this is of great interest for crop breeders to assess how much selection process achieved in one environment could be carried to other environments. Although the impact of environment on different genotypes has been known and considered in alfalfa breeding for a long time, it is still considered to be a challenging issue. In such scenario, advanced technologies such as high-throughput genotyping and phenotyping can help to reduce phenotypic burden. High throughput phenotyping (HTP) could drastically reduce the phenotypic burden in alfalfa by replacing the plot harvester with an unoccupied aerial vehicle (UAV). Images taken throughout the production years of a stand will enable the evaluation of large number of breeding materials and provide insights into genotype-by-environment interactions (G×E), in which

varieties have differential growth responses under different conditions. Quantitative genetic models can be built to accurately predict forage yields from spectral imaging. Advanced statistical models such as random regression models with Legendre polynomials could be implemented to fit such longitudinal phenotypes like VIs collected over time to estimate growth curves, to access genetic variation in growth and the relations of VIs to end-use traits like forage yield and quality (Meyer 2005). Understanding the genetic signal in differential growth response will allow for identification of breeding targets and optimal population change for sets of predictable environmental conditions.

The main objectives of this project were to (1) incorporate aerial high-throughput phenotyping to predict performance and genetic merit of the breeding materials, and (2) fit longitudinal random regression models to estimate genotype-specific growth curves and estimate the heritability of key growth parameters.

## Materials and Methods

A total of 36 Cornell varieties and breeding populations were established in a replicated variety trial with 5 replicates in Helfer, NY in 2019. The planting of the experimental trial was done on June 12, 2019. Forage yield was measured using a plot flail harvester, and dry matter yield for each plot was calculated from fresh forage weight and dry matter content samples. Forage yield (FY) was collected for three cuts in 2020 and 2021.

## Aerial phenotyping

Aerial phenotyping commenced on April 6, 2020. A DJI Matrice 600 Pro unoccupied aerial vehicle (UAV) equipped with a Micasense Rededge-MX multi-spectral camera was used for all flights. A flight plan was designed to obtain an 80% overlap in images collected at a flight speed of 2 m/s and an altitude of 20 m. Flights were conducted within 2 hours of solar noon on clear days when possible. A total of 56 flights were conducted on average every 4.3 days across six harvests. Four ground control points positioned at the four corners of the trial were measured with a Trimble RTK-GPS, which was used to geo-locate plots. Orthomosaics were constructed using Pix4D mapping software, and were subsequently uploaded into Imagebreed ([www.imagebreed.org](http://www.imagebreed.org)), a plot image database developed by our lab (Morales et al. 2020), for image processing and storage and vegetative index (VI) calculation at the plot level. Normalized difference vegetation indices (NDVI) were calculated from mean pixel values of near infrared (NIR) and Red bands of plot level images as

$$NDVI = \frac{(R_{NIR} - R_R)}{(R_{NIR} + R_R)} \quad (1)$$

where  $R_{NIR}$  is the near infrared reflectance and  $R_R$  is the red reflectance. Green normalized difference vegetation indices (GNDVI) and Normalized difference red edge indices (NDRE) were calculated using green and red edge reflectance instead of the red reflectance in Eq. 1, respectively.

Visible atmospherically resistant index (VARI) was calculated from mean pixel values of Green, Red and Blue reflectance.

$$VARI = \frac{Green - Red}{Green - Red + Blue} \quad (2)$$

Triangular greenness index (TGI) was calculated as,

$$TGI = Green - (0.39 * Red) - (0.61 * Blue) \quad (3)$$

## Random regression

Random regression models using third order of Legendre polynomials were used to fit a model for all time points using mean NDVI values. The general random regression model for a single trait can be formulated as (Schaeffer 2004):

$$\text{NDVI}_{tj} = \sum_k^{K1} \phi(t)_{jk} \beta_k + \sum_k^{K2} \phi(t)_{jk} u_{jk} + \sum_k^{K3} \phi(t)_{jk} p_{jk} + \epsilon_{tj},$$

where,  $\phi(t)_{jk}$  is a time covariate coefficient defined by a basis function evaluated at time point  $t$  ;  
 $\beta_k$  is a  $k$ th fixed regression coefficient for the population's mean growth trajectory;  
 $u_{jk}$  is a  $k$ th random regression coefficient associated with the additive genetic effects of the  $j$ th accession;  
 $K1$  is the number of random regression parameters for fixed effect time trajectories;  
 $K2$  and  $K3$  are the number of random regression parameters for random effects;  
and  $p_{jk}$  is a  $k$ th permanent environmental random regression coefficient for the accession  $j$ .  
Growing degree days (GDD) for each time points were calculated:

$$\text{GDD} = \frac{T_{\max} + T_{\min}}{2} - T_{\text{base}}$$

where  $T_{\max}$  is the maximum temperature,  $T_{\min}$  is the minimum temperature and  $T = 40^\circ\text{C}$  is the base temperature. Any temperature below  $T_{\text{base}}$  is set to  $T_{\text{base}}$  before calculating the average.

## Results

### *Heritability and Correlation of phenotypic indices close to the harvest date with biomass yield*

The heritability of different phenotypic indices ranged from 0 to 0.78 with the highest median value of 0.63 for GNDVI followed by 0.60 for TGI and 0.55 for NDRE.

The correlation was strongest for Green NDVI and biomass yield (0.4053, 0.7875, and 0.6779), followed by Red edge NDVI and biomass yield (0.417, 0.7898, and 0.6417) for the first, second and third cuttings respectively for 2020, while the genetic correlations for 2021 were strongest for Red edge NDVI and biomass yield (0.76, 0.74, and 0.66) followed by Green NDVI and biomass yield (0.75, 0.76 and 0.60) for the first, second and third cuttings (Figure 1). The difference in number of days between imaging and harvest were 9, 4 and 3 days respectively for first, second and third cutting of 2020 and 6, 3 and 3 days for first, second and third cutting of 2021 respectively.

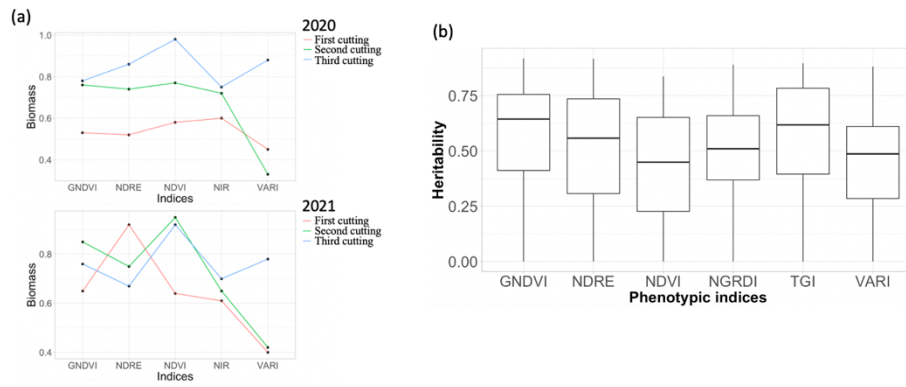


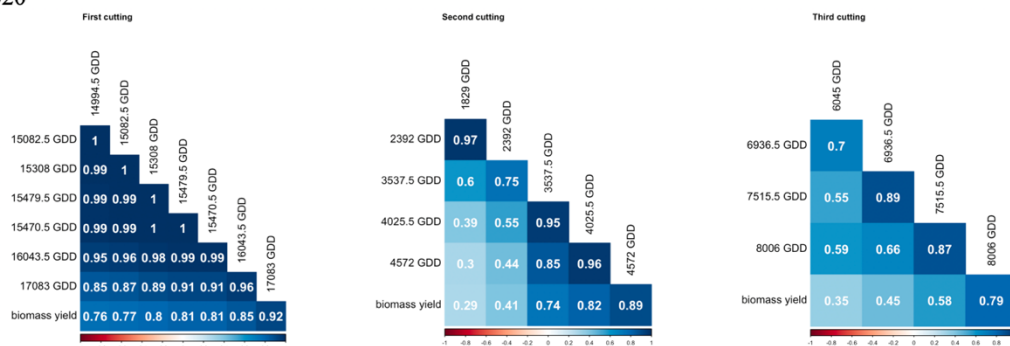
Figure 1(a). Box plot showing Heritability of GNDVI, NDRE, NDVI, NGRDI, TGI and VARI

Figure 1(b). Correlation of GNDVI, NDRE, NDVI, NIR and VARI with biomass yield for 2020 and 2021. Lines with red, green and blue color shows first, second and third cutting respectively.

### Genetic correlation

The genetic correlation between NDVI at first, second, third and last imaging timepoint were 0.76, 0.77, 0.8, 0.81 and 0.92, respectively, for the first cutting; 0.29, 0.41, 0.74 and 0.89 for the second cutting; 0.35, 0.45, 0.58 and 0.79 for the third cutting respectively for 2020 (Figure 2). The genetic correlation between NDVI at first, second, third and last imaging timepoint were -0.4, -0.37, -0.33, -0.10 and 0.86, respectively, for the first cutting; 0.59, 0.61, 0.69 and 0.98 for the second cutting; 0.4, 0.48, 0.72 and 0.99 for the third cutting respectively for 2021 (Figure 2).

2020



2021

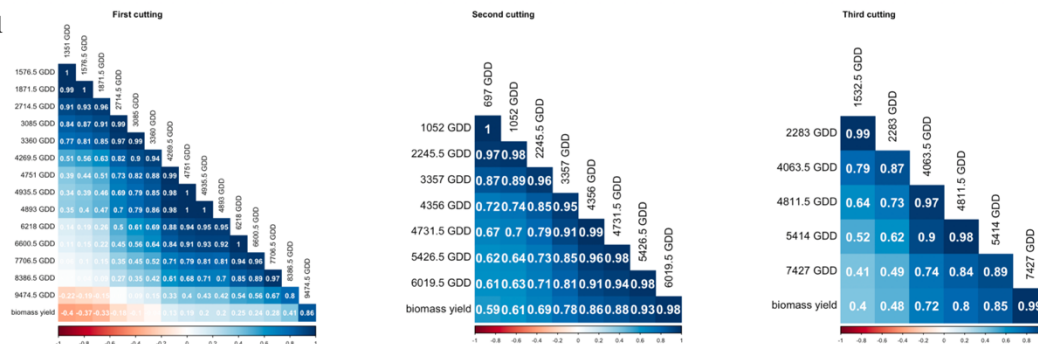


Figure 2. Heatmap showing genetic correlation of NDVI at different imaging timepoints (converted to Growing Degree Days (GDD)) with biomass yield.

### Growth curves

The breeding values estimated from NDVI and cNDVI (Cumulative NDVI) using random regression model were used to plot growth curves of 36 genotypes across different imaging time points (Figure 4 and Figure 5). The imaging time points were converted to Growing Degree Days (GDD). The breeding values of high yielding genotypes (upper ten percentile) were high starting from the first imaging. The variance in breeding values was high during early growing season than later growing season i.e. close to harvest time.

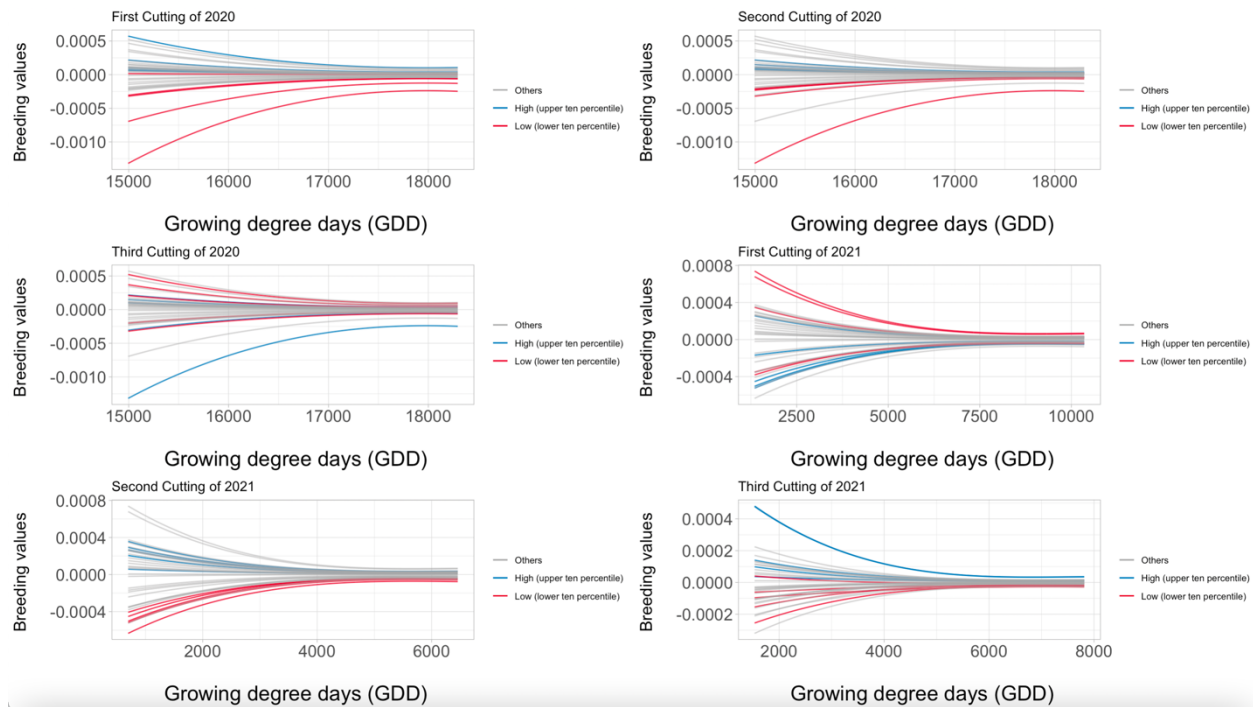


Figure 3. Growth curves of 36 genotypes across first, second and third cutting (2020) and fourth, fifth using breeding values of NDVI calculated using random regression model with third order of Legendre polynomials. X-axis shows Growing degree days (GDD) and Y-axis shows breeding values. The upper 10th percentile of high yielding genotypes and lower 10th percentile of low yielding genotypes are shown in blue and red color respectively.

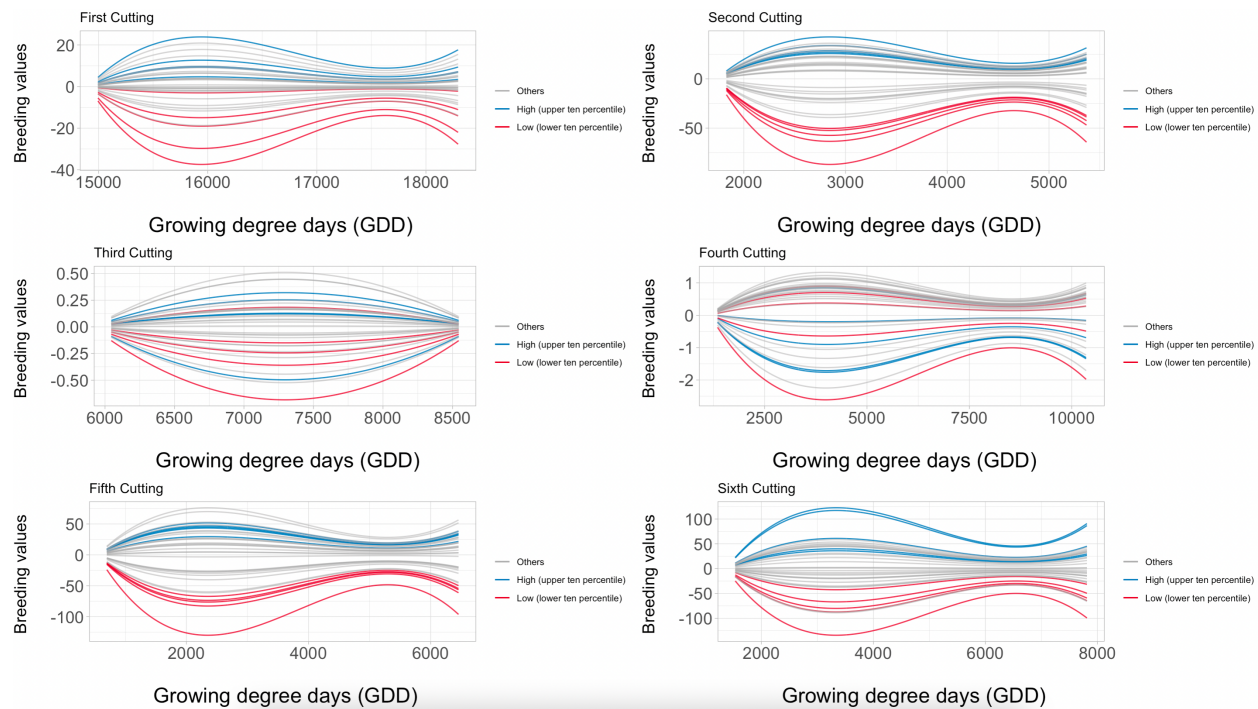


Figure 4. Growth curves of 36 genotypes across first, second and third cutting (2020) and fourth, fifth using breeding values of cNDVI (Cumulative NDVI) calculated using random regression model with third order of Legendre polynomials. X-axis shows Growing degree days (GDD) and Y-axis shows breeding values. The upper 10th percentile of high yielding genotypes and lower 10th percentile of low yielding genotypes are shown in blue and red color respectively.

## Conclusion

The correlation of phenotypic indices derived from multi-spectral imaging was strongest with yield when the difference between imaging day and harvest day days was less for both 2020 and 2021. The genetic correlation of NDVI with biomass yield was strongest for the imaging time point closest to the harvest date and weakest for the imaging time point earliest in the growing season for all the cuttings. The random regression model with third order of Legendre polynomials was converged and able to separate the high yielding (upper 10th percentile) and low yielding (lower 10th percentile) for most of the cuttings. Random regression models with a Legendre polynomials including other environmental trials are under evaluation to see the potentiality of these models to fit VIs from multiple time points.

## References:

Morales, Nicolas et al. (2020). “ImageBreed: Open-access plant breeding web–database for image- based phenotyping”. In: The Plant Phenome Journal 3.1, e20004.

Schaeffer LR (2004). Application of random regression models in ani- mal breeding. Livest Prod Sci 86(1–3),35–45