The data analysis challenge: from raw data to phenotypes

Wed, 11:15 AM

Jeff White and Jesse Poland

I. Introduction
   A. Useful to consider a recap of where we stand in the workshop:
      1. Monday: learned how to obtain data on height, growth/greenness/nitrogen (reflectance), transpiration (IRT)
      2. Tuesday: learned about vehicles and data logging
      3. Wed AM: learned how to associate data with plots (georeferencing)
      4. So we should now be confident that we have a workflow that will allow us to obtain data at plot level with high throughput
         a) Simplest to think of an array
         b) May include data besides from proximal sensing (Fig. 1)
   B. The rest of workshop is mainly concerned with:
      1. How to convert the basic data to useful information on phenotypes (Fig. 2)
      2. "Maximize the biologically useful information"
         a) Close relation to fundamental processes
         b) Reduce error in relation to cost of acquisition
            (1) Measurements of similar or related traits can vary greatly in cost
            (2) Leaf thickness example from White & Montes (2005)
      3. Three closely-linked topics
         a) Analysis of time series over the season (this lecture)
         b) Analysis of time series over short periods (A French)
         c) Inverse modeling (K Thorp/S Welch)
            (1) Using process-based ecophysiological models
            (2) Relate model parameters such as for photoperiod sensitivity to genotypic traits

II. "Biologically useful information"
   A. Not just final yield
      1. Often hear "which measurement has the strongest correlation with yield?"
      2. A direct estimate of yield is seldom the target for phenomics

a) Searching for mechanistic understanding

(1) Physiological traits related to underlying processes

(a) Plant height at day 50

(b) Maximum elongation rate

(2) The closer one gets to a fundamental process, the less G x E x M there will be:

(a) Time of flowering vs. photoperiod response

(b) Canopy temperature vs. root geotropism

b) Yield

(1) Integrates effects of many component traits

(2) Large G x E x M

B. Reduce error

1. Serial sampling can compensate for random error over time (Figure)

a)

b) Foundation of growth curve analysis (next lecture)

2. Analyses can compensate for other sources of variation

a) Co-variate analysis

(1) NDVI from active sensor -- height

b) Geospatial analysis

(1) Soil variability

c) Physical or ecophysiological models

(1) Account for weather (e.g., canopy temperature)

(2) Inverse modeling

III. Overview of processing methods (Figure 2)

A. How to convert single sets of observations to more useful data types

B. Direct output

1. Height

2. Canopy cover from imagery

C. Simple regression models to convert

1. NDVI to LAI

2. Canopy cover to LAI

D. Multivariate regression

E. Image analysis

F. Multivariate analysis
    1. Principal components analysis

IV. Time series (next lectures)
A. Long periods
B. Diurnal or a few days
C. Inverse modeling

V. Specialized analyses
A. Within-plot variability
    1. Canopy temperature
    2. Interplant spacing – covariate for other analyses
B. Flower or spike number
    1. Image segmentation (Thorp & Dierig, 2011)
    2. Tiller number from NDVI and CV of height (Scotford & Miller, 2004)

VI. Conclusion
A. Gathering the data is only half the challenge in field phenomics
B. Many options for analysis
    1. Maximize useful information
    2. If possible, reduce:
        a) Sampling error
        b) Effects of environment (soils, weather, management)

References

Scotford, I.M., Miller, P.C.H., 2004b. Estimating tiller density and leaf area index of winter wheat using spectral reflectance and ultrasonic sensing techniques. Biosystems Engineering 89, 395-408.

Thorp K.R., Dierig D.A. 2011. Color image segmentation approach to monitor flowering in lesquerella. Industrial Crops and Products 34:1150-1159. DOI: http://dx.doi.org/10.1016/j.indcrop.2011.04.002.

White J.W., Montes C. 2005. Variation in parameters related to leaf thickness in common bean (*Phaseolus vulgaris* L.). Field Crops Res. 91:7-22.

White J.W., Andrade-Sanchez P., Gore M.A., Bronson K.F., Coffelt T.A., Conley M.M., Feldmann K.A., French A.N., Heun J.T., Hunsaker D.J., Jenks M.A., Kimball B.A., Roth R.L., Strand R.J., Thorp K.R., Wall G.W., Wang G. 2012. Field-based phenomics for plant genetics research. Field Crops Research 133:101-112.

Wu R., Ma C.-X., Yang M.C., Chang M., Littell R.C., Santra U., Wu S.S., Yin T., Huang M., Wang M. 2003. Quantitative trait loci for growth trajectories in Populus. Genetical research 81:51-64.

Yin X., Goudriaan J., Lantinga E.A., Vos J., Spiertz H.J. 2003. A flexible sigmoid function of determinate growth. Annals of Botany 91:361-371.

Figure 1. Work flow for field phenomics hosing multiple types of data being recorded at different time scales (White et al., 2012).

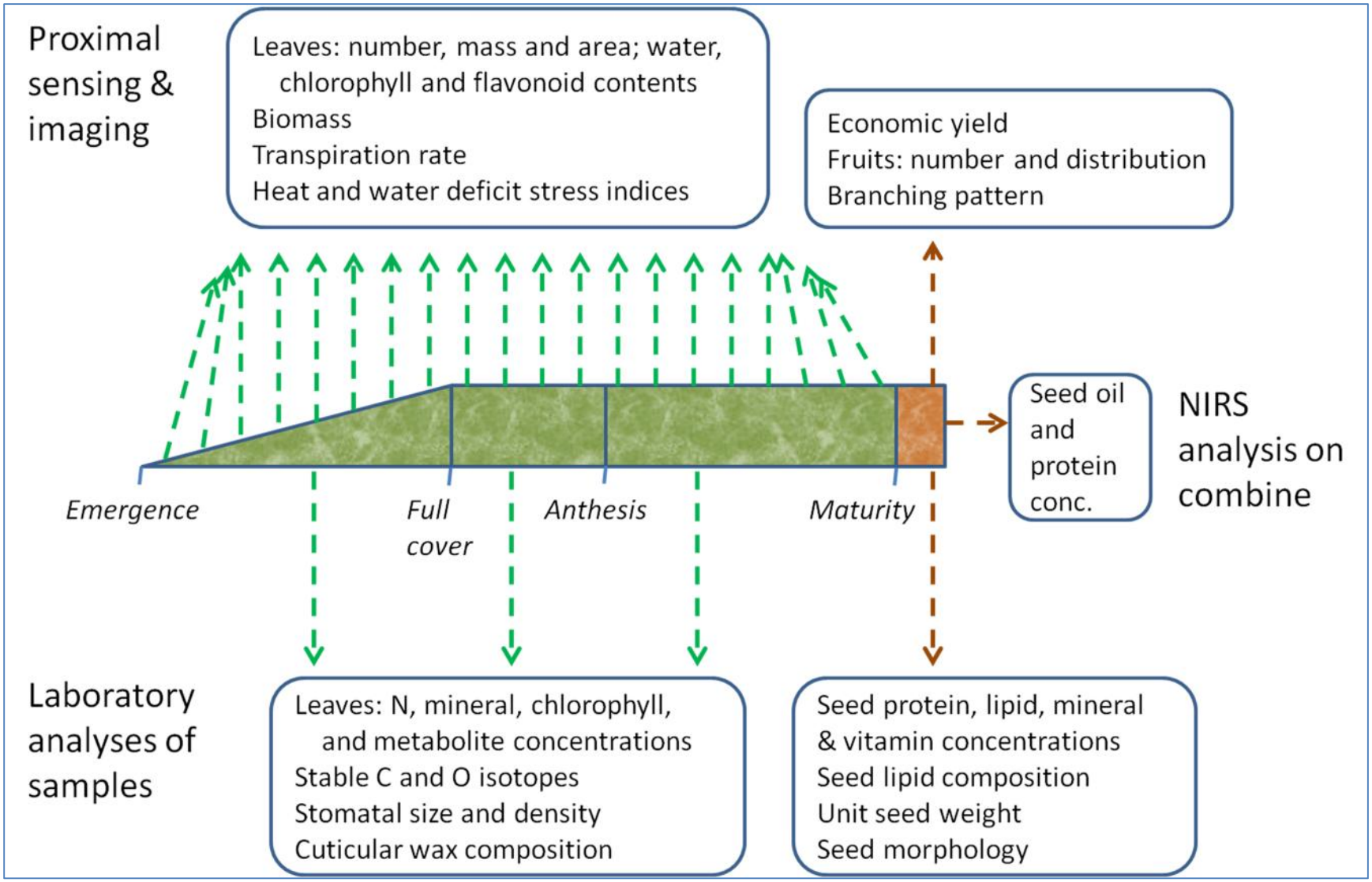

Figure 2. Examples of options for analyzing data from field phenomics (White et al., 2012).