

# Modeling Plant Photosynthesis Heterogeneity under Dynamic Light Conditions

Oliver Tessmer, Jeff Cruz, David M Kramer, Jin Chen

May 16, 2014

## 1 Introduction

With the advent of models for biomedical image analysis, heterogeneity, a concept relating to the uniformity in a substance, has assumed new importance since last decade (Tiihonen 1997; Wieneke 1999, Wang 2000). In plant biology, photosynthetic heterogeneity refers to a plant comprising multiple regions, many of which have different photosynthesis properties, potentially because of vastly different  $CO_2$  intake capability, stomatal conductance and tolerance level to environmental changes.

Photosynthetic heterogeneity is a universal phenomenon among plants (Charles 2008). With the rapid development of lighting and imaging techniques, real-time non-invasive monitoring of photosynthesis became feasible, resulting in vast amount of phenotype data (Wituszynskaa 2013). A consensus view of the data is that the photosynthesis ability of a plant is not uniform across the whole area (Charles 2008, Meng 2007). By integrating plant morphological and physiological features, measuring plant photosynthetic heterogeneity aids interpretation of the sophisticated phenotyping data, particularly important for plant primary productivity estimation and modeling (Meng 2007).

The granularity of photosynthetic heterogeneity of plants ranges from cells to tissues, leaves, and even the whole plant level. While in-leaf variability in photosynthetic activity has been well studied for the understanding of the effects of stomatal conductance [2, 1], recent works show that photosynthetic capacity may decline with vertical gradient and leaf age, indicating that leaf-based photosynthetic heterogeneity is a key towards the understanding of plant total photosynthesis [5, 3].

To measure leaf-based photosynthetic variability in a nowadays popular large-scale phenotyping experiment, in which hundreds of plants are monitored simultaneously in real time, it is naturally required to segment each leaf from top-view photosynthesis images automatically. Photosynthesis images are false-color images, where the light intensity of every pixel is proportional to photosynthetic efficiency (Toet 1996). Differences between individual leaves with similar photosynthetic efficiency can be subtle, making the boundaries between them is difficult to define and creating a significant challenge for subsequent shape analysis. The difficulty even arises when individual leaves overlap and occlude one another in these false-color images.

We have developed a new leaf segmentation approach called leafPH to identify leaves from photosynthesis images...

In this paper, we present a new computational approach to compute leafPH. Comparing with the existing approaches, our method is novel in the following ways:

1. Our measure relies on a new leaf segmentation method, which considers plant-specific constraints including leaf shape similarity and symmetry. By taking the biological knowledge about the shapes of any possible leaf into account, we add flexibility to the model to fit to all the general leaves. Specifically, for leaf boundary identification, we develop a leaf symmetry based curve fitting model to define the initial shape of a leaf, and develop a plant layout based approach to define leaf tips and orientation, therefore defining the leaf initial position. Note that since leaf is assumed to be symmetric, curve fitting could converge without large amount of training.

2. Mathematically, leafPH refers to the differences between individual leaf photosynthesis and the pooled photosynthesis across all the leaves, with the weights being those used in the pooling method. With leafPH, transient regional variation events which do not affect the whole plant photosynthesis, can be easily discovered.

## 2 Background of heterogeneity measurement

The classical measure of heterogeneity is Cochran's Q-test (Conover1999), which is calculated as the weighted sum of squared differences between individual study effects and the pooled effect across studies with the weights being those used in the pooling method. Q is distributed as a chi-square statistic with  $k - 1$  degrees of freedom, where  $k$  is the number of studies, and each study is a comparison between observation and reference.

Q has low power as a comprehensive test of heterogeneity (Gavaghan et al, 2000), especially when the number of studies is small, i.e. most meta-analyses. Conversely, Q has too much power as a test of heterogeneity if the number of studies is large (Higgins et al. 2003): Q is included in each meta-analysis function because it forms part of the DerSimonian-Laird random effects pooling method (DerSimonian and Laird 1985). An additional test, due to Breslow and Day (1980), is provided with the odds ratio meta-analysis. It is arguably not possible to examine the null hypothesis that all studies are evaluating the same effect, by considering the only the summary data from the studies: The heterogeneity test results should be considered alongside a qualitative assessment of the combinability of studies in a systematic review.

The  $I^2$  statistic (Higgins2003) describes the percentage of variation across studies that is due to heterogeneity rather than chance (Higgins and Thompson, 2002; Higgins et al., 2003).  $I^2$  is an intuitive and simple expression of the inconsistency of studies results. Unlike Q it does not inherently depend upon the number of studies considered. A confidence interval for  $I^2$  is constructed using either i) the iterative non-central chi-squared distribution method of Hedges and Piggott (2001); or ii) the test-based method of Higgins and Thompson (2002). The non-central chi-square method is currently the method of choice (Higgins, personal communication, 2006) it is computed if the 'exact' option is selected. The L'Abb plot can be used to explore the inconsistency of studies visually.

To measure leaf-based photosynthetic variability in a large-scale phenotyping experiment, in which hundreds of plants are monitored in real time, it is naturally required to identify each leaf from top-view photosynthesis images automatically. However, leaf identification is a difficult task, in that 1) all the leaves are very similar in appearance especially for model plant *Arabidopsis*; 2) leaf features are much fewer than the well-studied face recognition problem; and 3) leaves may overlapped with each other that makes the problem more complicated. In short, there are two challenging computational problems. First, how to separate overlapped leaves in photosynthesis images where the leaf boundaries are obscure. Second, how to recognize incomplete leaves due to

occlusion. Alternatively, approximation approaches can be used to estimate the leaf-based photosynthetic heterogeneity, including distribution based, region based and pixel based approaches. While these approximation approaches are easier to model, their performance may be reduced.

### 3 Data Acquisition and Preprocessing

In the photosynthesis phenotyping experiment, hundreds of *Arabidopsis thaliana* plants (wild type, genetic variations with gene knockout or over-expression, ecotypes, etc) were grown side-by-side under three different light conditions (constant, sinusoid, fluctuate), for in total three days. Top-view fluorescence images were collected every 15 minutes in order to observe the photosynthesis activity of all of the plants simultaneously. Each fluorescence image is a grey-scale image with a resolution of 1M pixels at 12-bit intensity.

To accurately capture the photosynthesis activities of plants from fluorescence images, a image segment method is applied to remove the background, identify every piece of leaf [10], measure the intensity of pixels on leaves, and finally convert the intensity values to the measure of four kinds of photosynthesis parameters. The extracted measurements of photosynthesis parameters are presented in the form of multi-dimensional time-series, one dimension for every photosynthesis parameter. Figure 1 shows that the measurements of a photosynthesis parameter  $\Phi_2$  at one time point are quite different for the leaves of the same plant (i.e. plant no. 3-7 and 5-5).

### 4 Computing Photosynthesis Heterogeneity

Heterogeneity is a concept relating to the uniformity in a substance (REF), which has assumed new importance since last decade with the advent of models for biomedical image analysis [7, 9, 8]. For example, heterogeneity in the aggressiveness of tumor cell populations has been adopted as an essential feature in predicting treatment success [6].

Photosynthetic heterogeneity refers to a plant comprising multiple regions, each of which has different photosynthesis efficiency, potentially because of vastly different tissues, sizes, developmental stages, and even transient variations over time. Photosynthetic heterogeneity is a universal phenomenon among plants (REF). With the rapid development of lighting and imaging techniques, real-time non-invasive monitoring of photosynthesis became feasible, resulting in vast amount of data (REF). A consensus view of the data is that the photosynthesis ability of a plant is not the same (or similar) across the whole area (REF). By connecting plant morphological and physiological features, measuring photosynthesis heterogeneity aids interpretation of the sophisticated plant phenotyping data, in particular as evidence of consistency across studies.

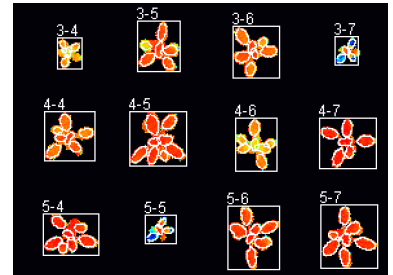


Figure 1: An example of leaf-level photosynthesis parameters in plants.

#### 4.1 Leaf-based photosynthetic heterogeneity

The traditional heterogeneity measure was designed to compute the differences between individual study effects, and for each individual study, people assume only two underlying populations

representing the experimental vs. control groups on a continuous outcome. However, to compute the leaf-based photosynthetic heterogeneity, we have to compare all the leaves where each leaf may have its unique underlying population. To solve the more complicated problem, we modified the current Q test, and describe it in this section.

#### 4.1.1 Effect-size estimate for leaf pair-wise comparison

Given a plant with  $n$  leaves, we define a set  $T = \{T_{11}, T_{12}, \dots\}$ , where  $|T| = n \times (n - 1)/2$ , which is the total number of pair-wise leaf-to-leaf comparisons. In each tuple  $T_{ij}$ , we compare the photosynthesis values of leaves  $L_i$  and  $L_j$  using Equation 1.

$$T_{ij} = c(L_i, L_j) \frac{\text{mean}(L_i) - \text{mean}(L_j)}{S(L_i, L_j)} \quad (1)$$

where  $\text{mean}(L_i)$  is the averaged photosynthesis value of every pixel in leaf  $L_i$ , and  $c(L_i, L_j)$  is a correction factor for the positive bias suffered by the standardized mean difference with small sample sizes, and can be estimated by  $c(L_i, L_j) = 1 - 3/(4|L_i| + 4|L_j| - 9)$ , according to (Hedges and Olkin 1985) [4]. Such adjustment is necessary, because for small plants due to gene knockout, there may be only a few pixels for each leaf. According to the equation of  $c(L_i, L_j)$ ,  $c(L_i, L_j)$  is close to 1 for large leaves with many pixels, otherwise it is close to 0, downweighing the value of heterogeneity.  $S$  is a pooled estimate of the within-group standard deviation using Equation 2 [4].

$$S(L_i, L_j) = \sqrt{\frac{(|L_i| - 1)\text{std}^2(L_i) + (|L_j| - 1)\text{std}^2(L_j)}{|L_i| + |L_j| - 2}} \quad (2)$$

where  $\text{std}^2(L_i)$  is the standard variance of leaf  $L_i$  and  $|L_i|$  is the total number of pixels in leaf  $L_i$ .

The sample variance of  $T_{ij}$  is estimated using Equation 3 [4].

$$ST_{ij} = \frac{|L_i| + |L_j|}{|L_i||L_j|} + \frac{T_{ij}^2}{2(|L_i| + |L_j|)} \quad (3)$$

#### 4.1.2 Random effect-size estimate for all-leaf comparison

Once an effect-size estimate is obtained from each pair-wise comparison, we integrate them by calculating an average effect size, assessing the statistical heterogeneity around the average estimate, and searching for moderator variables when there is more heterogeneity than can be explained by chance. In general, the most realistic statistical model to integrate the effect estimates in a meta-analysis is the random-effects model, because it incorporates the two possible sources of heterogeneity among the studies, i.e., statistical variability caused by sampling error and substantive variability.

Given the set of all pair-wise comparisons  $T$ , we compute leaf-based photosynthetic heterogeneity LeafPH using Equation 4.

$$\text{LeafPH} = \frac{\sum_{i,j} w_{ij}(T_{ij} - \bar{T})}{|T|} \quad (4)$$

where  $\bar{T}$  is the weighted average of all  $T_{i,j}$ .

To compute the weight  $w_{ij}$  for each pair-wise comparison  $T_{ij}$ , we develop the following strategies. First, the value differences between large leaves should contribute more to the overall heterogeneity score. The leaf area based weight can be modeled using  $\min(|L_i|, |L_j|) / \sum L$ . Second, we also think the heterogeneity between different regions of a plant is more interesting than the heterogeneity within the same region. So the leaf region based weight can be modeled using a binary function  $[Pos(L_i) \neq Pos(L_j)]$ , where the function  $[x]$  return 1 if  $x$  is true, and return 0 otherwise. In summary, we define  $w_{ij}$  using Equation 7.

$$w_{ij} = [Pos(L_i) \neq Pos(L_j)] \times \frac{\min(|L_i|, |L_j|)}{(ST_{ij}^2 + \tau^2 + 1) \sum L} \quad (5)$$

where  $\tau^2$  is the between-study variability affecting the effect estimates. A commonly used estimator of the between-studies variance  $\tau^2$  is an estimator based on the method of moments proposed by DerSimonian and Laird (1986).

$$\tau^2 = \begin{cases} \frac{\tilde{Q} - |T| + 1}{\sum \tilde{w}_{ij} - \frac{\sum \tilde{w}_{ij}^2}{\sum \tilde{w}_{ij}}} & \tilde{Q} > |T| + 1 \\ 0 & else \end{cases} \quad (6)$$

where  $\tilde{Q}$  and  $\tilde{w}_{ij}$  are the results of Q test of fixed effect estimate and the corresponding weight.

$$\tilde{w}_{ij} = [Pos(L_i) \neq Pos(L_j)] \times \frac{\min(|L_i|, |L_j|)}{(ST_{ij}^2 + c) \sum L} \quad (7)$$

## 4.2 next step...

Plants have different shaped leaves as they have coped to live in different environments. Since leaves varying dramatically in shape and size and there is no model fit to all kinds of leaves, we focus on the model plant Arabidopsis in this study. A closed observation on Arabidopsis rosette leaves reveal that all the intact leaves are symmetry with their shapes close to oval, and all the leaves are at a similar height (2-3mm tall, does not have high canopy), but the photosynthesis values of different leaves can be statistically different under certain conditions.

## 5 Modeling Photosynthesis Heterogeneity

## 6 Conclusion

The photosynthetic properties of plants can vary dramatically across cells, tissues, and organs [], reflecting differences in development, stress responses, regulation of processes such as stomatal conductance [], photodamage [], and storage of photosynthate [] and contribute substantially to productivity [].

For example, we observed that the acclimation of photosynthesis in response to cold temperatures appears to be more rapid and robust in younger or emerging than older leaves, and ecotypes isolated from different latitudes show distinct heterogeneity patterns, implying that these responses are important for adaptation of photosynthesis to fluctuating temperatures. In other cases, exposure of plants to fluctuating light resulted in loss of photosynthetic capacity or increased photoinhibition in specific sets of leaves or leaf sectors. In many cases, older leaves are preferentially

affected, suggesting that resources for maintenance or acclimation responses are preferentially directed to younger leaves. However, we have also identified mutant lines where younger leaves are preferentially affected, which presumably affect the development of photosynthetic robustness.

In order to systematically study the leaf level photosynthesis phenotypes, especially in a high-throughput screen manner, we developed a novel computational tool to automatically conduct statistical analysis on leaf based photosynthesis.

## References

- [1] T.N. Buckley, G.D. Farquhar, and K.A. Mott. Qualitative effects of patchy stomatal conductance distribution features on gas-exchange calculations. *Plant Cell Environ*, 20:867880, 1997.
- [2] J.M. Cheeseman. Patchy - simulating and visualizing the effects of stomatal patchiness on photosynthetic co2 exchange studies. *Plant Cell Environ*, 14:593599, 1991.
- [3] C.P. Chen et al. The effect of leaf-level spatial variability in photosynthetic capacity on biochemical parameter estimates using the farquhar model: A theoretical analysis. *Plant Physiol*, 148(2):11391147, 2008.
- [4] Larry V Hedges and Jack L Vevea. Fixed-and random-effects models in meta-analysis. *Psychological methods*, 3(4):486, 1998.
- [5] K. Kitajima et al. Decline of photosynthetic capacity with leaf age and position in two tropical pioneer tree species. *Am. J. Bot.*, 89(12):1925–1932, 2002.
- [6] F. O’Sullivan et al. A statistical measure of tissue heterogeneity with application to 3d pet sarcoma data. *Biostat*, 4(3):433–448, 2003.
- [7] J. Tiihonen et al. Cerebral benzodiazepine receptor binding and distribution in generalized anxiety disorder: a fractal analysis. *Mol. Psychiatry*, 2:463–471, 1997.
- [8] N. Wang, C. Wilkin, A. Bocking, and B. Tribukait. Evaluation of tumor heterogeneity of prostate carcinoma by flow- and image dna cytometry and histopathological grading. *Anal Cell Pathology*, 20:49–62, 1999.
- [9] H. Wieneke, C. Zander, E. G. Eising, M. Haude, A. Bockisch, and R. Erbel. Non-invasive characterization of cardiac microvascular disease by nuclear medicine using single-photon emission tomography. *Herz*, 24:515521, 1999.
- [10] X. Yin, X. Liu, J. Chen, and D.M. Kramer. Multi-leaf alignment from fluorescence plant images. *IEEE Winter Conference on Applications of Computer Vision*, 2014.