Longitudinal concordance correlation function based on variance components: an application in fruit color analysis

Thiago P. Oliveira, John Hinde, Silvio S. Zocchi

University of São Paulo, Piracicaba, Brazil and National University of Ireland Galway, Ireland

Email: thiago.paula.oliveira@usp.br

29th International Biometric Conference, Barcelona, Spain, 2018

Introduction

- Papaya (*Carica papaya* L.) is a tropical and climacteric fruit with antioxidant, anti-carcinogenic and anti-mutagenic properties, containing carotenoids and high nutritional value.
- Brazil is ranked second in the world for papaya production with 17.5% of global production.
- One of the most important criteria for determining the ripeness stage over time is the peel color.

Motivation

- Color is traditionally measured by a colorimeter — this procedure can lead to bias in the determination of the mean color depending on the number and location of sampled points on fruit's peel.
- A possible alternative to measure color is the use of image analysis.
- We extend the *Concordance Correlation Coefficient* ([1] and [2]) to assess agreement between colorimeter and scanner mesurements and explore if a sample on the equatorial region is representative of the whole peel surface in measuring the mean hue of papaya peel over time.

Case-study

- Study — evaluate the peel color of 20 papaya cv. Sunrise Solo with a flat-bed scanner (HP Scanjet G2410) and a tristimulus colorimeter Minolta CR-300

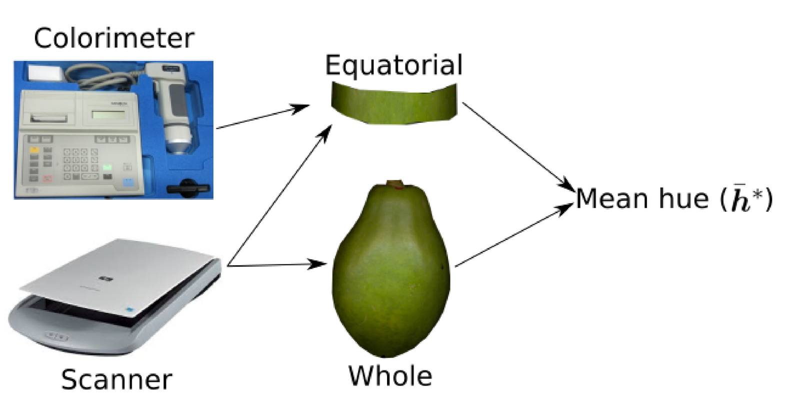

Figure 1: Vector of mean hue obtained from observed points on equatorial region with the colorimeter and scanner, and whole region with the scanner

- Colorimeter: observe four points on the equatorial region.
- Scanner: measure approximately 1,000 pixels on the equatorial region and 10,000 pixels over the whole region of the fruit's peel using scans of both sides.
- Problems with fungal diseases — some fruits did not have a complete set of responses.

The longitudinal concordance correlation (LCC)

- Let Y_{ijlk} denote the measurement on the ith fruit ($i = 1, 2, \ldots, N$) by the jth method ($j = 1, 2, \ldots, m$) on the lth region ($l = 1, 2, \ldots, r$) at time t_{ik} ($k = 1, 2, \ldots, n_i$), where n_i is the total number of observations taken on the ith fruit over time.
- Additional variability — interaction among method, region, and fruit $\Rightarrow$ new variable A with mr categories given by the combination of region and method levels.

$$d_{ick} = \begin{cases} 1, & \text{for category } c \text{ of variable } A \\ 0, & \text{otherwise} \end{cases}.$$

- Multiple mixed-effects regression model for longitudinal data

$$Y_{ijlk} = \sum_{h=0}^{p} \beta_{hjl} t_{ik}^{h} + \sum_{h=0}^{q} b_{hi} t_{ik}^{h} + \sum_{c=1}^{mr-1} \alpha_{ci} d_{ick} + \epsilon_{ijlk} \quad (1)$$

$$\boldsymbol{u}_i = \begin{bmatrix} \boldsymbol{b}_i \\ \boldsymbol{\alpha}_i \end{bmatrix} \sim MVN\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \boldsymbol{D} = \begin{bmatrix} \boldsymbol{G} & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & \boldsymbol{G}_{\alpha} \end{bmatrix}\right) \quad \text{and} \quad \boldsymbol{\epsilon}_i \sim MVN(\boldsymbol{0}, \boldsymbol{R}_i),$$

- According to [3], under the model (1), we can define the LCC based on variance components for observations measured from different unique combinations of two factors at time t_{ik} as

$$\rho_{jl,j'l'}(t_k) = \frac{\boldsymbol{t}_k \boldsymbol{G} \boldsymbol{t}_k^T + \boldsymbol{d}_k \boldsymbol{G}_{\alpha} \boldsymbol{d}_k'^T}{\boldsymbol{t}_k \boldsymbol{G} \boldsymbol{t}_k^T + \frac{1}{2}\left\{\boldsymbol{d}_k \boldsymbol{G}_{\alpha} \boldsymbol{d}_k^T + \boldsymbol{d}_k' \boldsymbol{G}_{\alpha} \boldsymbol{d}_k'^T + \sigma_{\epsilon}^2 \left[g\left(t_k, \boldsymbol{\delta}_{jl}\right) + g\left(t_k, \boldsymbol{\delta}_{j'l'}\right)\right] + S_{jl,j'l'}^2(t_k)\right\}} = \rho_{jl,j'l'}^{(p)}(t_k)\, C_{jl,j'l'}(t_k) \quad (2)$$

- **Longitudinal Pearson Correlation** (LPC): $\rho_{jl,j'l'}^{(p)}(t_k)$ measures how far each observation deviated from the best-fit line at a fixed time $t_k = t$ (precision measure).

The longitudinal concordance correlation (LCC)

- **Longitudinal accuracy** (LA): $C_{jl,j'l'}(t_k)$ measures how far the best-fit line deviates from the 45° line at a fixed time $t_k = t$ (accuracy measure).
- $S_{jl,j'l'}(t_{ik}) = E\left(Y_{ijlk}\right) - E\left(Y_{ij'l'k}\right) = \boldsymbol{t}_{ik}\left(\boldsymbol{\beta}_{jl} - \boldsymbol{\beta}_{j'l'}\right)$, with $h = 1, 2, \ldots, p$ and $jl \neq j'l'$.
- $Var\left(\epsilon_{ijlk}\right) = \sigma_{\epsilon}^2 g\left(t_{ik}, \boldsymbol{\delta}_{jl}\right)$, where $g(.)$ is a variance function assumed continuous in $\boldsymbol{\delta}$, t_{ik} is the time covariate, and $\boldsymbol{\delta}_{jl}$ is a vector of variance parameters for observations measured by the jth method on the lth region.
- Non-parametric bootstrap confidence interval — uses simple case-resampling.

Simulation Study

- Whole versus equatorial region measured by colorimeter; 2 – whole versus equatorial region measured by scanner; 3 – colorimeter versus scanner on equatorial region, and 4 – colorimeter versus scanner on whole region.

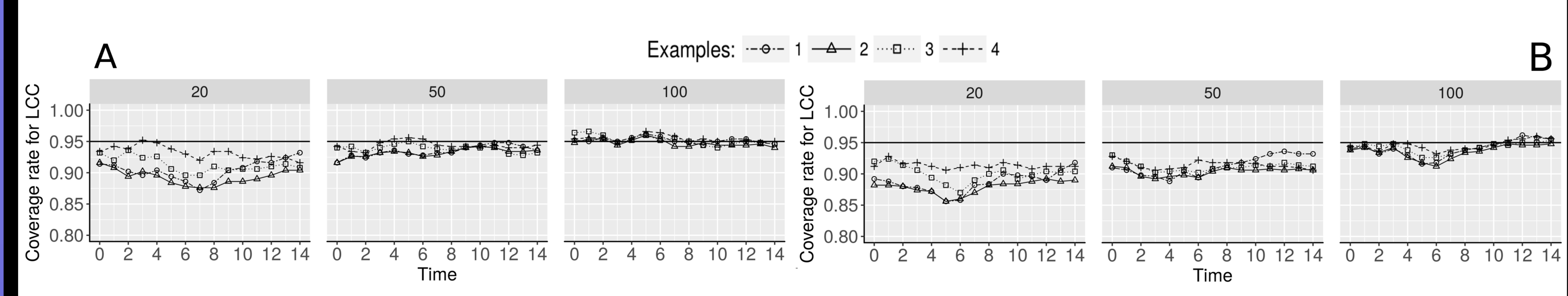

Figure 2: Estimated simultaneous coverage rate based on nominal 95% confidence intervals for $LCC(t_k)$, for $N \in \{20, 50, 100\}$ fruits under balanced (A) and unbalanced (B) designs.

Results and discussion

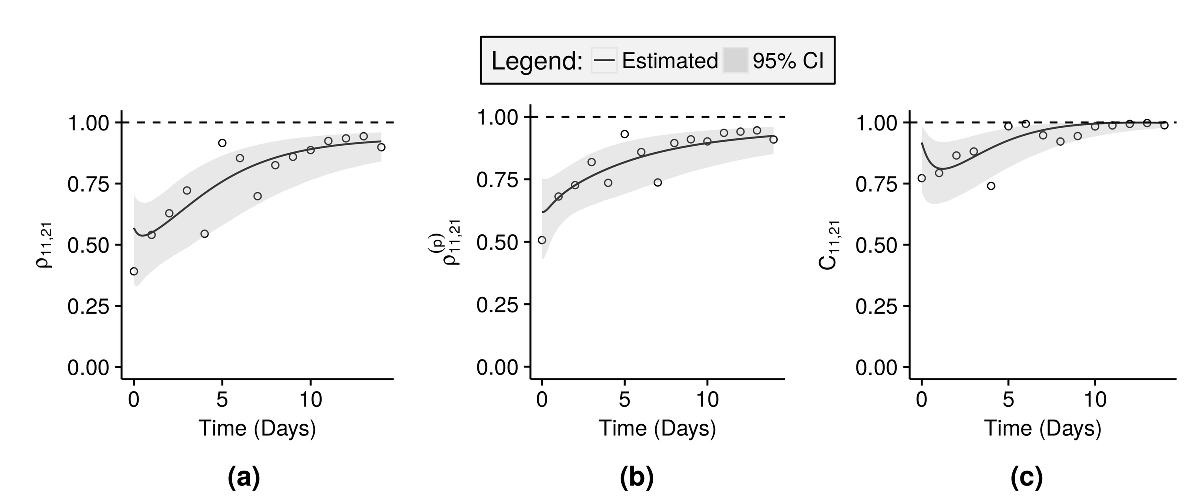

Figure 3: Estimate and 95% confidence interval (CI) for (a) LCC; (b) LPC;(c) LA between observations measured on the equatorial region by the scanner and colorimeter.

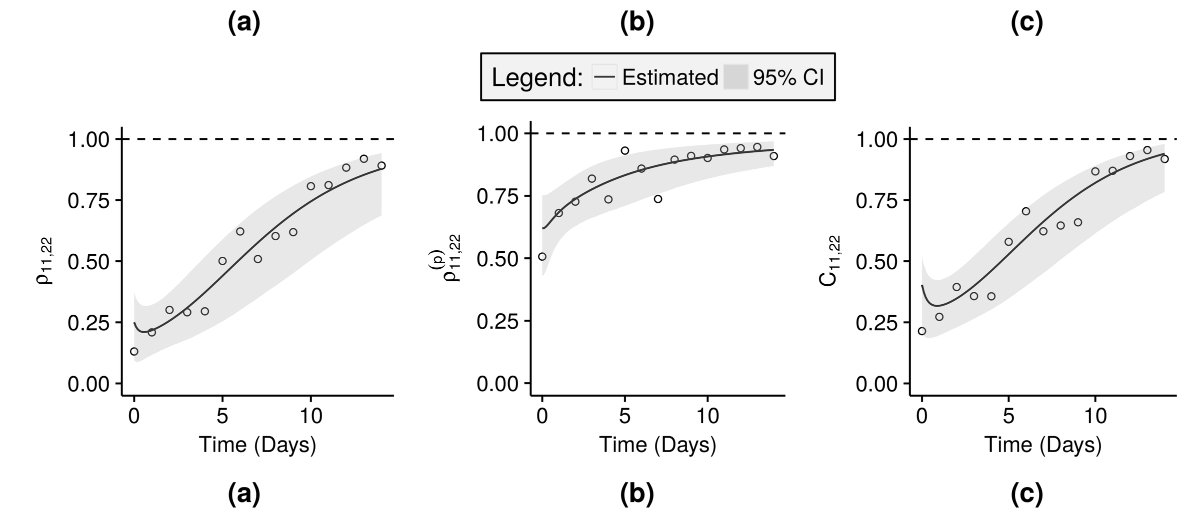

Figure 4: Estimate and 95% confidence interval (CI) for (a) LCC; (b) LPC;(c) LA between observations measured on the equatorial region by the colorimeter and whole region by the scanner.

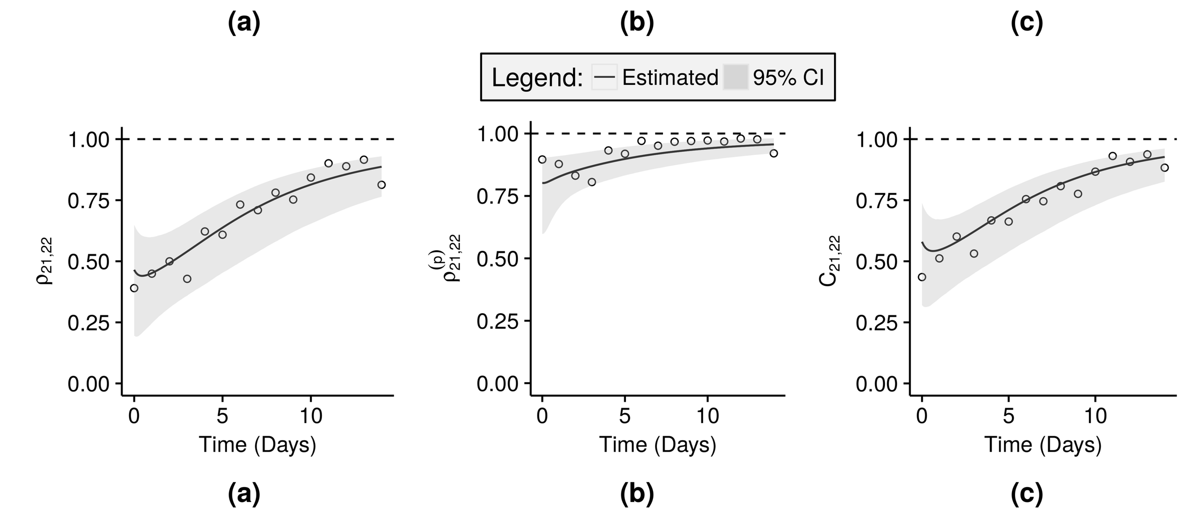

Figure 5: Estimate and 95% confidence interval (CI) for (a) LCC; (b) LPC;(c) LA between observations measured on the equatorial and whole regions by the scanner.

- Points only on the equatorial region are not representative of the whole peel region.
- Image analysis of the whole peel region should be used to compute the mean hue.
- Large LA between observations measured by the colorimeter and scanner on the equatorial region suggested that the topography and curved surface of papaya fruit did not affect the mean hue obtained by the scanner.

References

[1] **Lin, L. I.** (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, **45**, 255–268.

[2] **Carrasco, J. L., King, T. S., and Chinchilli, V. M.** (2009). The concordance correlation coefficient for repeated measures estimated by variance components. *Journal of Biopharmaceutical Statistics*, **19(1)**, 90–105.

[3] **Oliveira, T.P., Hinde, J., Zocchi, S.S.** (2018). Longitudinal Concordance Correlation Function Based on Variance Components: An Application in Fruit Color Analysis. *Journal of Agricultural, Biological, and Environmental Statistics*, **23(2)**, 233–254.