# Green mathematics: Benefits of including biological variation in your data analysis

TIJSKENS L.M.M.[1], SCHOUTEN R.E.[1], UNUK T.[2], SIMČIČ M.[2]

## ABSTRACT

Biological variation is omnipresent in nature. It contains useful information that is neglected by the usually applied statistical procedures. To extract this information special procedures have to be applied. Biological variation is seen in properties (e.g. size, colour, firmness), but the underlying issue is almost always to the variation in development or maturity in a batch of individuals generated by small scale environmental differences.

The principles of assessing biological variation in batches of individuals are explained without putting emphasis on mathematical details. Obtained explained parts increase from about 60 to 80 % for the usual approach to 95 when the biological variation is taken into account. When technical variation or measuring error is small even 99 % can be achieved. The benefit of the presented technology is highlighted based on a number of already published studies covering the colour of apples during growth and storage and the firmness of cut tomatoes during storage.

**Key words:** biological variation, biological shift factor, mixed effects nonlinear regression, indexed nonlinear regression, colour of apples, firmness of tomatoes

## IZVLEČEK

### ZELENA MATEMATIKA: KORISTI OD VKLJUČEVANJA BIOLOŠKE SPREMENLJIVOSTI V ANALIZO PODATKOV

Biološka spremenljivost je prisotna povsod v naravi. Vsebuje koristno informacijo, ki je navadno prezrta v navadno uporabljenih statističnih postopkih. Če hočemo izluščiti to informacijo je potrebno uporabiti posebne postopke. Biološko spremenljivost lahko opazujemo v lastnostih kot so velikost, barva, čvrstost itd., kjer je osnova zanjo skoraj vedno razlika v razvoju ali zrelosti vzorca analiziranih primerkov, ki jo povzročajo majhne notranje in okoljske razlike. Osnove ugotavljanja biološke spremenljivosti v naboru analiziranih primerkov so razložene brez poudarjanja matematičnih podrobnostih. Pojasnjen delež spremenljivosti je pri običajenm postopku med 60 in 80 %. Če upoštevamo biološko spremenljivost, se poveča na 95 %. V primeru, da odpravimo še napake meritve, lahko pojasnjen delež spremenljivosti povečamo na 99 %. Takšen način obdelave je dal zelo dobre rezultate v že objavljenih raziskavah razvoja barve jabolk med rastjo in hrambo ter pri meritvah čvrstosti paradižnika med shranjevanjem.

**Ključne besede:** biološka spremenljivost, biološki šift faktor, mešani učinek nelinearne regresije, indeksirana nelinearna regresija, barva jabolk, čvrstost paradižnika

## 1 INTRODUCTION

Variation is everywhere, in humans, in animals, in plants, in DNA, in climate, in weather and therefore also in measured experimental data. Nature is very generous in providing variation, but

[1]   Horticulture and Product Physiology, Wageningen University, NL; e-mail: Pol.Tijskens@wur.nl

[2]   Faculty of Agriculture and Life Science, University of Maribor, SI

[3]   Biotechnical Faculty, University Ljubljana, SI

amazingly, in a way nature is also very lazy. It uses the same process mechanisms over and over again but in endless combinations. The presence of all that variation from different sources poses a severe problem on the modelling of processes and on analysing experimental data properly.

In day-to-day horticultural practise, products are sorted and graded to remove the variation in batches. This sorting is very effective, but is only applied to external properties like colour, size and defects. Internal quality attributes like content of sugars, acids, dry matter, Brix, etc., constituting the eating quality, are hardly affected by that kind of grading.

All statistical procedures are built in such a way that the effect of variation is minimal. Mostly this is achieved by making the samples as uniform as possible (sorting) and by using mean values in one way or another.

Variation, however, contains useful information, not only with respect to differences between individuals but also with respect to the real mechanisms in action. Removal of variation, either by sorting or by statistical procedures, also removes the information contained in that variation, and prevents effectively the study and understanding of the dynamics of variation.

What we need is green mathematics, i.e. sustainable mathematics that takes this variation in account, to be used in plant biology and horticulture. But as Kermit the frog of Sesame Street put it: it is not that easy to be green (YouTube). To assess biological variation in experimental data, special lines of thinking have to be used, and data have to be analysed applying special statistical procedures like mixed effects and indexed non-linear regression. In this paper, the reasoning behind the technique (Tijskens et al. 2003, 2005), developed over the last couple of decades (see references) will be highlighted, however, without too much emphasis on equations. The technique will be illustrated based on some examples of skin colour of apples and tomatoes in storage and during growth.

## 2 VARIATION IN PROPERTIES AND DATA

### 2.1 Origin of variation

Just like product quality, biological variation is generated exclusively during growth. During the subsequent post-harvest storage, one can only try to minimise the further development of variation and to prevent some of its detrimental effects.

Once formed, plants, trees, organs like leaves and fruit are quite localised. They can't move from one place to another. That means that the small differences (not every day the same but always in the same order / direction) in e.g. micro climate, fertilisation, location in the canopy and soil type etc., accumulate over the entire growth period resulting in a sometimes considerable variation in properties of individuals.

What constitutes an individual depends on the focus of the study or application: fruit, plants, cells, organelles, harvest flights, fruit bins, pallets, containers etc. Scientific studies will (probably) focus on the smaller items (fruit, organs, cells), while commercial application will focus more on the larger items (harvest flights, fruit bins, pallets, containers). The principle however is always the same: determine the biological shift factor.

### 2.2 How to deal with variation

These sources of variation in growing conditions basically translate in variation in the stage of development of e.g. fruit. In Figure 1, an example is shown for a sigmoidal behaviour, frequently found and applied for colour development. The individual lines represent the individual fruit; the arrows indicate how much a line (individual fruit) has to be shifted in time to fall over the same (central) curve. That shift in time is called the biological shift factor (Tijskens et al. 2005). Irrespective the stage of development (how high or low on the y-axis) that shift factor is the same for each individual fruit: the red arrows are of equal length. The same is true for the green arrows for another individual. In short, the biological shift factor is a property of an individual not to be

attributed to sampling time. The biological shift factors of all the individuals in a batch will show a normal distribution (histogram) invariable in time, independent of the stage of maturity.

So, dealing with variation basically boils down to the estimation of the biological shift of all individuals in a batch. That can only be done, when data are collected repeatedly for the same individual (non-destructive testing). These are so-called longitudinal data: each individual is monitored in time for one or more properties like colour, firmness, size and number of cells in a leaf,

size and number of stomata in a leaf, photosynthetic activity in leafs etc.

Applying an appropriate model for the observed behaviour, and using fruit identification, the biological shift factor can be estimated by indexed or mixed effects nonlinear regression for each individual (random effects) along with the other model parameters are estimated in common (fixed effects) for all the individuals (Schouten et al. 2002, 2004, 2009, Hertog et al.2002, 2004, 2007, Tijskens et al. 2003, 2005, 2006, 2008, 2009, 2010, Unuk et al. 2012).



**Figure 1:** Behaviour of a normalised attribute of several individuals in a batch according a logistic mechanism (e.g. colour). Red and green arrows indicate the shift per individual to fall over the same generic line. The crossing line in the middle is obtained by analysing not the individuals but the mean values per sampling time when the biological variation is not included in the analysis

## 2.3 Benefits of including biological variation in data analysis

The overall benefit of including biological variation in data analysis is a better understanding of its behaviour, its dynamics and the rules that govern these. The major result of all these dedicated studies, crossing the borders of all disciplines, is that the occurrence, magnitude and behaviour of natural variation are as deterministic as all chemical reactions and reaction mechanisms. Understanding of the mechanisms and dynamics of variation will eventually result in a better prediction of quality and maturity and in ways to

deal with variation without or additional to the traditional sorting and grading. Using the statistical procedure of non-linear regression, without explicitly taking care of the variation, explained parts ($R^2_{adj}$) of 60 to 80 % can be obtained. The unexplained part is a mixture of the technical or measuring error and the biological variation not taken care of. When properly taking care of the variation by including the biological shift factor in the analysis, explained parts can reach as high as 95 to 99%. The unexplained part is now purely the technical or measuring error.

# 3 EXAMPLES FROM PRACTICE

## 3.1 Colour in the orchard

Details of this study can be found in Unuk et al. (2012). In 2009, three apple (*Malus domestica* Borkh.) cultivars ('Braeburn', 'Fuji', 'Gala') were grown in an orchard near Maribor (SI). Colour of individual apples was measured with a Minolta colour meter. The apples were selected at three locations in the canopy: shady, partially sunny and sunny. The raw (unprocessed) data for 'Braeburn' are shown in Figure 2. The effect of location (how much sunlight do these apples get) can clearly be observed in the stage of development. Moreover, the difference in and magnitude of the distribution in colour at different moments in time is indicated by the red ellipses.

Applying indexed nonlinear regression one arrives at a generic development curve with numerical information on time axis (biological time). All locations in canopy follow the same curve, with explained parts well above 95% (Figure 3).

**Figure 2:** Behaviour of colour a* of 'Braeburn' apples at three locations in the canopy. Clearly the effect of location can be seen in the stage of development (where on the generic sigmoidal curve are the data located): left shady early, partial sunny (middle) and sunny (right) are already more developed.

**Figure 3:** Generic behaviour of colour a* of 'Braeburn', 'Fuji' and 'Gala' apples from the three locations in the canopy versus biological time (calendar time + individual biological shift factor).

## 3.2 Managing the orchard

Details of this study can be found in Tijskens et al. (2009). 'Golden Delicious' apples were grown near Maribor in 2001 and 2002. Colour of individual apples was measured with a Minolta colour meter. Different levels of crop load and fertilisation were applied. In Figure 4, the raw data are shown along with the estimated behaviour per individual. Again the effect of applied conditions can be seen in the stage of development: The less

crop load, the more developed. The effect of fertilisation is less clear (too small number of levels). When analysing the data including biological, all effects of crop load and fertilisation could be attributed to the biological variation. All individuals followed the same generic pattern (Figure 5). Striking is the now very clear difference in behaviour between the two season, both in range of change as in rate of change.



**Figure 4:** Behaviour of colour a* of individual Golden delicious apples at decreasing levels of crop load (CL: left to right, and increasing levels of fertilisation (NR: top to bottom). Full lines behaviour per individual, estimated using all data combined.

**Figure 5:** Standardised behaviour of colour a* of 'Golden delicious' apples in two seasons including the effects of crop load and fertilisation. The colour values are 'standardised' (col_stan) to get rid of the different values for the asymptotes at plus and minus infinite time.

### 3.3 'Granny Smith' in storage

Details of this study can be found in Tijskens et al. (2008). 'Granny Smith' apples were harvested at three orchards in south Slovenia in the season 1997 and stored at three temperatures. The colour of individual fruit was monitored with a Minolta colour meter. In Figure 6, the raw data are shown. The behaviour reflects the lower part of the normally observed sigmoidal behaviour (see e.g., Figure 3). Some clear 'outliers' in rate of change are indicated by the red arrows.

Analysing the data, taking care of the stage of maturity (biological shift factor) and the differences in lower colour values of the individual apples of all three orchards stored at all three temperatures, an explained part of 97 % was obtained. All 'outliers' complied with the model formulation. The generic pattern of development is shown in Figure 7 for the three temperatures separately. A small effect of a chilling injury process can be observed in the rate constants: at 1°C (Figure 7 left) the rate of colour increase is slightly larger than at 10°C (Figure 7 right).



**Figure 6:** The unprocessed data of apples from the orchard at Arnovo Selo as an example. Individuals that would normally be considered as 'outliers' are indicated by the red arrows.

**Figure 7:** The standardised data of 'Granny Smith' apples all orchards combined for the three temperatures. All 'outliers' comply with the same model (follow the same mechanism), but the value of their biological shift factor is outside the 'normal'.

### 3.4 Cut Tomatoes

Details of this study can be found in Lana et al. (2005). Tomatoes were harvested at 3 stage of colour development (breaker, pink and red). That actually reflects a grading of the fruit. Tomatoes were sliced and stored at 5 temperatures. Firmness was measured non-destructively by limited compression. The data of all slices were analysed using the same exponential model, and the biological shift factor for the three different stages of maturity was estimated. Clearly, firmness of cut tomatoes decreases according to the same exponential model with the same value of the reaction rate constant, irrespective of the stage of maturity at harvest (Figure 8).

That delivers a direct link between postharvest behaviour and growth of tomatoes.



**Figure 8:** Decrease in firmness of cut tomatoes (outer pericarp) as a function of storage time for the three stages of maturity at harvest. The symbols represent the mean of five replicates at the indicated temperatures. The time on the x-axis has been transformed to a standard temperature of 20 ∘C.

### 4 CONCLUSION

Biological variation is everywhere. In product properties as well as in experimental data on these properties. Applying mixed effects or indexed nonlinear regression based on fruit identification, the biological shift factor of individuals can be estimated, generating more reliable analyses and a

better understanding of the dynamics and behaviour of that variation.

Biological variation is not random at all, but complies with strict deterministic rules.

The technology is not restricted to fruit in horticulture, but can easily be expanded to cover all kind of areas in agriculture and botany on almost every conceivable property.

## 5 REFERENCES

Hertog MLATM. 2002. The impact of biological variation on postharvest population dynamics. *Postharvest Biology and Technology* 26:253–63; DOI: 10.1016/S0925-5214(02)00044-3

Hertog MLATM, Lammertyn J, Desmet M, Scheerlinck N, Nicolaï BM. 2004. The impact of biological variation on postharvest behaviour of tomato fruit. Postharvest Biology and Technology 34:271–84; DOI: 10.1016/j.postharvbio.2004.05.014

Hertog, M.L.A.T.M., Lammertyn, J., De Ketelaere, B., Scheerlinck, N., Nicolaï, B.M. 2007. Managing quality variance in the postharvest food chain. *Trends in Food Science and Technology* 18, 320-332; DOI: 10.1016/j.tifs.2007.02.007

Lammertyn J, De Ketelaere B, Marquenie D, Molenberghs G, Nicolaï BM. 2003. Mixed models for multicategorical repeated response: modelling the time effect of physical treatments on strawberry sepal quality. *Postharvest Biology and Technology* 30:195-207; DOI: 10.1016/S0925-5214(03)00100-5

Lana M.M., Tijskens L.M.M., van Kooten O. (2005): Effects of storage temperature and fruit ripening on firmness of fresh cut tomatoes. Postharvest Biology and Technology 35, 87–95; DOI: 10.1016/j.postharvbio.2004.07.001

Schouten RE, Tijskens LMM, van Kooten O. 2002. Predicting keeping quality of batches of cucumber fruits based on a physiological mechanism. *Postharvest Biology and Technology* 26:209–20; DOI: 10.1016/S0925-5214(02)00017-0

Schouten RE, Jongbloed G, Tijskens LMM, van Kooten O. 2004. Batch variability and cultivar keeping quality of cucumber. *Postharvest Biology and Technology* 32:299-310; DOI: 10.1016/j.postharvbio.2003.12.005

Schouten RE, Zhang X, Verschoor JA, Otma EC, Tijskens LMM, van Kooten O. 2009: Development of colour of Broccoli heads as affected by controlled atmosphere storage and temperature. *Postharvest Biology and Technology* 51:27–35; DOI: 10.1016/j.postharvbio.2008.06.005

Tijskens LMM, Wilkinson EC. 1996. Behaviour of biological variability in batches during post-harvest storage. *AAB Modelling Conference: Modelling in Applied Biology - Spatial Aspects*. p. 267-8, 24-27 June 1996, Uxbridge, UK

Tijskens LMM, Konopacki P, Simčič M. 2003. Biological Variance, Burden or Benefit? *Postharvest Biology and Technology* 27:15-25; DOI: 10.1016/S0925-5214(02)00191-6

Tijskens LMM, Heuvelink E, Schouten RE, Lana MM, van Kooten O. 2005. The biological shift factor. Biological age as a tool for modelling in pre- and postharvest horticulture. *Acta Horticulturae* 687:39-46

Tijskens LMM, Herold B, Zude M, Schlüter O, Geyer M. 2006. Biological variation in the apple orchard. In: *Proceedings COST 924 workshop* (ed. J. Ovesná, V. Pouchová, K. Mitrová) *From plant to Human genes*, p27-34, November 2006, Prague, CZ. ISBN 80-86555-99-2.

Tijskens LMM, Konopacki PJ, Schouten RE, Hribar J, Simčič M. 2008. Biological variance in the colour of Granny Smith apples. Modelling the effect of senescence and chilling injury. *Postharvest Biology and Technology* 50:153–63; DOI: 10.1016/j.postharvbio.2008.05.008

Tijskens LMM, Dos-Santos N, Jowkar MM, Obando J, Moreno E, Schouten RE, Monforte AJ, Fernández Trujillo JP. 2009. Postharvest firmness behaviour of near-isogenic lines of melon. *Postharvest Biology and Technology* 51:320-6; DOI: 10.1016/j.postharvbio.2008.06.001

Tijskens L.M.M., Jacob S., Schouten R.E., Fernández-Trujillo J.P., Dos-Santos N., Vangdal E., Pagán E., Pérez Pastor A. 2010. Water loss in horticultural products. Modelling, data analysis and theoretical considerations. Acta Horticulturae 858, 465-471.

Unuk T., Tijskens L.M.M., Germšek B., Zadravec P., Vogrin A., Hribar J., Simčič M., Tojnko S. 2012, Effect of location in the canopy on the colour development of three apple cultivars during growth. *Journal of the Science of Food and Agriculture.* **92**:2450–2458; DOI: 10.1002/jsfa.5651

YouTube Kermit the frog: https://www.youtube.com/watch?v=hpiIWMWWVco &feature=player_detailpage&list=RDhpiIWMWWVco