

Biological Variance in Agricultural Products Theoretical Considerations

L.M.M. Tijskens¹ and P. Konopacki²

¹ ATO, P.O.Box 17, 6700 AA Wageningen, the Netherlands,

E-mail: L.M.M.Tijskens@ato.wag-ur.nl

² Research Institute of Pomology and Floriculture, Skierniewice, Poland

Abstract

The food that we eat is uniform neither in shape or appearance nor in internal composition or content. Since technology became increasingly important, the presence of biological variance in our food became more and more of a nuisance. Techniques and procedures (statistical, technical) were developed. The most widely used are sorting or grading on large-scale operations. Statistical rules were developed on the sampling size to obtain a reliable mean value. Experimental design strategies were developed for efficient research. All the techniques used are rather empirical and only loosely connected to the theoretical and fundamental knowledge we do have.

An effort is made to approach biological variance from a new perspective. The nature of biological variance is traced back to its origin in the processes of production and conversion, primarily driven by the "random" differences in climate and soil. By modelling the dynamics on the level of the individual units that constitutes a batch, rather than modelling the mean value for the batch itself, more fundamental models can be developed. These more fundamental models are generic in nature and describe changes in product properties for all kinds of circumstances like growing area, seasonal effects, harvest maturity and storage temperature.

INTRODUCTION

As long as man exists on this earth, he has eaten his daily food to stay alive and to survive. In the very early days, when gathering our food by hunting and searching, the variation in kind, amount and quality was tremendous. Mankind adapted to that situation and developed skills to distinguish between edible and inedible products.

After a while, mankind settled down in small settlements and started farming. The era of agriculture began. The variance in quality and properties of our daily food gradually decreased to a bearable and manageable level. In the recent past, the effects of biological variance were avoided as much as possible by sorting and grading the products solely based on external properties. Statistics provided sufficient and reliable means of experimental design to decrease the effect even further. So, we learned quite efficiently to live with the problem without too much trouble. But still we are faced with the hidden and not (and certainly not fully) understood rules of biological variance.

In more recent times, driven by the shift in importance from external properties (shape, colour, size) to internal properties (sugar content, taste, vitamins, potentially health promoting compounds), more problems have been encountered even in apparently similar batches of produce. Many studies have been devoted to the causes and effects of biological variance in agricultural products. Especially the growing of food has attained much effort, as that is the realm where technology fails most in controlling the onset of biological variance (see for some examples General: Tijskens et al. 2000).

Altogether, the issue of biological variance has so far mainly been dealt with by technological means, without too much understanding of the rules and changes that govern the processes that cause biological variance. In this lecture, a new approach is proposed, that is based on the understanding and modelling of the behaviour of individuals in a population, rather than starting from the behaviour of the mean values of batches. Some theoretical examples are deduced and shown, that have however, a strong implication for practice.

ORIGIN AND SOURCES OF BIOLOGICAL VARIANCE

To understand how biological variance works, how it affects the perception of quality in our food and the handling of our food commodities, we have to understand what it is and where it comes from. In a previous lecture (General: Tijsskens et al. 2000), an attempt was made to review these sources and the nature of biological variance. The results are briefly summarised as follows:

- Variance in any product is part of the **production process**, whether natural (growing) or technological (food processing). Without production of a commodity or conversion of one commodity into another, no variation will be present. Variance in commodities is a direct consequence of variations in the producing sequence.
- Usually, effects of **external factors**, different in space and/or time initiate this variation. Examples of the most frequently responsible external factors are sunlight, daily and seasonal temperature changes, rainfall, soil type and its spatial distribution, fertilisers and their spatial distribution. Superimposed on this type of variance, is the variance that originates from **internal factors** as e.g. genetic differences in the producing species.
- The search for the variations in the **primary production process** brings us to the full area of growing and cultivation. The effects of various circumstances, especially weather, climate, soil type and fertilisation have been studied for already a long time. It covers a list of over 1 million references. The literature references, provided in Tijsskens et al (General: 2000), constitute merely a short list of efforts. The biological variance occurring in agricultural products, caused by external factors, is generally subdivided into three categories, reflecting more the type of observation rather than the fundamental cause of the differences. For the time being, we will stick to this classification:
 - Regional variation: difference between growing areas caused by soil composition, fertiliser availability and different climatic conditions due to location (e.g. latitude, vicinity of sea, ocean, mountain, desert), different cultivation treatments by growers during the cultivation period, (plowing, fertilisation, and irrigation).
 - Seasonal variation: The most important driving force is here the difference in weather over the different years. Indices to optimise the harvest time belong to this category.
 - Batch variation: the variance between the individuals in a batch. The source of this type of variation is the result of the same driving forces as for the previous two origins of variation, but on a far smaller scale. At the moment of harvest, not all individuals in a batch are at the same stage of development and exhibit a distribution in maturity.
- **Technical variance** in machinery and processing equipment will lead to different holding and passing times (Residence Time Distributions) which on their turn will induce variance in the products under treatment.

WHAT IS BIOLOGICAL VARIANCE?

Biological variance is difficult to describe in a generic way. The Webster dictionary defines variance as “the fact, quality, or state of being variable or variant”, and “an instance of variability: a degree of difference”. This definition does bring us not much further than what we already guessed. Some analogy with the definition of **quality** by Kramer and Twigg (General: 1970) is evident: “*the composite of those characteristics that differentiate individual units of a product, and have significance in determining the degree of acceptability of that unit by the buyer*”.

So, we could define **variance** as “*the composite of properties that differentiate individual units of a product*”.

This definition contains several interesting aspects and consequences:

- Variance can only occur in a **population** of individual units (entities). Without individual units in a greater set or part of the total population (e.g. batch, growing lot, season, partial population, total population etc) variance does not exist. This aspect brings in the vast knowledge on population dynamics and the manifold procedures developed by statistics to describe variance, population dynamics and statistical

distributions.

- What those **individual units** are and what these greater sets are depends more or less on the level of aggregation one wants to consider. One can take individual fruits in a batch, the cells in a fruit, the individuals in a grading category etc. The easiest and most practical way to define the level of aggregation is to adapt to the usual way consumers, trade and handling treat the products. For scientific purposes and for studying and understanding the occurring processes, the choice of units, however, is completely unlimited.
- The **composite of properties** contain all the differences between the individual units in a population. This implies that we cannot say that a sample of products is homogeneous, unless all properties are homogeneous. Therefore, even in apparently homogeneous samples, there always can be (or will be) some hidden variation. Of course, the importance of this variance for describing product behaviour and usability depends on the importance and level of variation of the varying property.
- Biological variance can only occur in **biological produce**. Variance can occur in man-made products, but this can be considered more technical or **technological variance** rather than biological variance. Technical variance in man-made products is most of the time far less and far less frequent than in nature-made products.

Normally, biological variance is limited to that variation in product properties that we cannot yet explain or do not yet understand fully. For example, if we consider apples, there exists some difference between apples Granny Smith and apples Jonagold. However, we know these are different cultivars, and we make good use of this “variance” in apples to add some choice in taste and appearance of the daily apples we eat. In other words, **as soon as we know and understand the reason of this variation in product properties, this variation ceases to be regarded as biological variance.** We can make good use of the existence of this variation if and only if we understand its origin and its behaviour. In our opinion, this is the most crucial reason to study on a fundamental level biological variance in agricultural products.

TECHNOLOGICAL WAYS OF DEALING WITH BIOLOGICAL VARIANCE

All approaches to deal with this ever-present and ubiquitous nuisance, apply more or less empirical means, either to avoid the effects of external circumstances during production (production control) or to decrease the effect of the variance present (sorting and grading). For describing the variance, the statistical distributions and the population dynamics have been used widely.

Production Control

To tackle the regional and seasonal type of biological variation, and to arrive at products as uniform as possible, production of agricultural products (growing) has been controlled as much as possible. In times of draught irrigation is used. Amount of fertiliser and frequency of fertilisation is adapted to the “needs” of the product, prescribed or at least influenced by regional and seasonal aspects that are beyond human control. Lately more and more modelling is being applied to understand the effects of external factors on variation and to deduce appropriate responses to that variation.

The ultimate form of this type of control is growing food in greenhouses where almost any aspect of the growing conditions can be controlled by man.

Sorting and Grading

In post-harvest handling of fruits and vegetables, sorting and grading is a long known and applied technique, almost standard nowadays, to reduce the effects of occurring variation in batches of produce. Of course, grading on one property can never eliminate the disturbing behaviour in another property. Until recently, the technique was almost exclusively applied based on external properties or quality attributes, like colour, appearance and visible defects. Lately sorting and grading based on internal quality attributes became feasible by virtue of the increasing possibilities of non-destructive

quality measurements. The driving force behind this development is the increasing importance of taste and quality for the modern consumer. Sorting on external attributes only can never lessen the variance in internal properties.

Statistical Procedures

Statistics have developed techniques to minimise the effect of biological variation on experimental results. The most commonly applied technique is simply taking the mean over all the individuals in a sample. The size of a sample depends on the observed standard deviation in the sample. To make experiments more efficient and intelligent, adequate rules of experimental design have been developed.

A vast number of distribution functions and estimating procedures have been developed to describe the distribution of properties in a population or a sample. Those range from continuous normal or Gaussian distribution, to discrete Poisson and binomial distributions and include numerous empirical distributions as the Weibull and gamma distributions to account for skewness and kurtosis often encountered in biological variance. One has to be careful with taking the mean value in biological systems: a function of a mean is something quite different than the mean of a function. Allen developed a mathematical sound second order approximation procedure to account for this different behaviour (General: Allen 1988).

HOW SHOULD WE DEAL WITH BIOLOGICAL VARIANCE?

Without denying the value of research conducted so far in this field, time has come to try out a new approach to this old phenomenon. Based on the postulations so far, we have deduced that biological variance is a direct consequence of differences between the individual units of a batch. We have also deduced that these differences are a consequence of dynamic processes that occur during growth and conversion. From these deductions, the new approach is rather evident. The new approach for understanding and describing biological variance is mainly based on two important aspects and pathways. These two aspects are highly linked: one activity cannot exist without the other one.

Process Oriented Modelling

Repeatability of experiments is really the basis of modern science. In the same circumstances, the same process will always produce the same result. Aberrations to this rule will lead invariably to biological variance. The key-action to understand and describe biological variance, its source and its development, is therefore to find out and determine what the important processes are that occur in our food. Up to a certain level, this fundamental reasoning can be applied on mean data values. Examples can be found in the many more fundamental models (see references Modelling). From the results of all this modelling and analysis of data, it becomes clear that the reasons for the presence of biological variance are mostly hidden very deep in the growing mechanism. What also becomes clear is that for explaining and understanding the situation, fundamentally correct models are absolutely necessary. The problems at hand become very rapidly too complex to be studied and analysed in one effort. Consequently, sound and correct problem decomposition in terms of essential processes is necessary. To allow the problem to be split up in smaller and manageable sub-problems, it is essential that the knowledge obtained in one study can be used entirely in another study. This makes the transferability of values of model parameters essential. That is only possible with fundamentally correct models (General: Tijssens et al. 1998).

The model on the development of sugar content in potatoes during storage at various temperatures (cold induced sweetening) in relation to the maturity at harvest was of prime importance in the development of this new approach (Modelling: Hertog et al. 1997a, 1997b).

Individual Monitoring

Although biological variance only can occur in populations of individual units, it

is completely and solely determined by the properties of the individual units. The study on individual behaviour of units in a batch makes it necessary to apply non-destructive measuring techniques on all the individuals in a sample or batch. An individual in a population is the only entity that during its lifetime (e.g. during growth, storage, processing and consumption) always exhibits a behaviour based on exactly the same combination of initial conditions. Therefore, in repeating measurements on individuals, the process itself, instead of the observed phenomena, can be studied without interference of biological variance. Fortunately, several non-destructive measuring techniques have been developed in the past decade and are being developed for assessing both external and internal properties. The benefits of non-destructive measurements for data analysis were highlighted by Tijskens et al. (General: 1999).

A nice example of the effect of individual monitoring is that the relation between mean colour of a batch of French fried potatoes with mean sugar content has a statistical correlation of about 60%. The relation between colour and sugar content of the individual French fries was well over 95% (A. Braaksma, ATO, unpublished data).

FUNDAMENTALS OF MODELLING BIOLOGICAL VARIANCE

Example 1: Making Batches

Let us stick for the time being to colour as product property. We all can perceive it ourselves, without technical assistance, and we all have a fair amount of expertise. In due time the same line of reasoning will be applicable to internal product properties and quality attributes.

Suppose we have a large harvest batch of tomatoes, say 10,000. The maturity in such a batch will be not entirely homogeneous, however careful the picking will have been conducted. Suppose the distribution in maturity (expressed as days of development) can be approximated with a normal distribution. Out of this population, smaller consumer batches of 10 tomatoes are prepared. Every time such a small consumer batches is taken from the mother population, part of the variance in maturity of the population will be incorporated into the smaller batch. But every time a different part!

For many products, colour develops according a sigmoidal function, often modelled as a logistic function. The standardised colour function, with all colour values standardised between 0 (green) and 1 (red or yellow) is:

$$c = \frac{1}{1 + e^{-k \cdot (t - t_m)}} \quad \text{eq. 1}$$

where c is the normalised colour, k the reaction rate constant, t is the time of storage, starting at some arbitrary point e.g. harvest, and t_m is the maturity at $t=0$, expressed in time, containing all information on initial conditions. As a consequence of the way the consumer batches are created, the distribution over t_m for all the individuals in a batch causes a skew distribution in the colour of all the individuals: tomatoes can not be greener then green, nor redder then red. This distribution will depend on the initial distribution of maturity t_m , and the rate with which the colour develops during growth and storage.

Usually, experimental set-ups work in their analysis with mean data over a number of batches. Because of the small number of individuals in consumer batches, the mean property (e.g. colour) in each of those consumer batches will develop in an apparently different and erratic way. In Fig. 1, an example is shown for the behaviour of colour in five different batches taken from the same mother population. The corresponding apparent standard deviation is shown in Fig. 2. From these graphs we can clearly see the "random" effect of taking small consumer batches out of a mother population, not only the apparent rate of development (tangent in Fig. 1) but also the estimated standard deviation over the colour changes sometimes drastically.

The smaller the standard deviation in the mother population is, the smaller the aberrations between theoretical and observed rate of development is (not shown). In other words, when the consumer batches are prepared with sufficient care with respect to

sorting and grading, the nuisance caused by biological variance is rather limited.

Example 2: Effect of Harvest Time and Storage

Accepting that there exists a distribution in the quality attribute under study (still colour as an example), it is interesting to know how this distribution looks like and how it behaves during storage.

Based on the assumption of a logistic change in colour during storage, and based on the fundamental rules of statistical distribution, one can deduce the equation that describes the distribution at all combinations of rate constant, initial standard deviation in initial maturity, and storage time and storage temperature (deduction not shown, publication in preparation). The distribution of a quality attribute that changes according a logistic behaviour looks like:

$$p = \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{1}{2} \left(\frac{kt_m + \ln\left(-\frac{c-1}{c}\right) - t}{\sigma} \right)^2}}{\sigma c(c-1)k} \quad \text{eq. 2}$$

where p is the density of a quality attribute, c the stage in quality attribute of an individual, k the reaction rate constant, t the time, σ the standard deviation in maturity of a distribution with mean assumed zero. With this function we can simulate and study the combinations important for applications in practice. An example is shown in Fig. 3. We clearly can observe that at the early stage of ripening, all individuals are predominantly green, exhibiting a distribution, which is skew (tailing) to the right, moving during storage towards a much flatter and wide but more normal distribution, and further on to a mirror image of the green stages when reaching the predominant red stage of coloration. Again the standard deviation of the initial distribution on maturity (σ) as well as the reaction rate constant (k) has a tremendous effect on the behaviour of biological variance.

This standard deviation on maturity (σ) is directly linked to the time between successive harvests of fruits from one production area: with a picking rate of once a week, σ is around 7 days, with a picking rate of twice a week, σ is around 3 days etc. This also confirms the long known rule: the more frequent fruits are picked, the more uniform the batches are, and the more acceptable the batch will be for the next actor in the food chain.

The magnitude of the reaction rate constant (k) is directly linked to the storage temperature. As for all reaction rate constants, the rate constant of the colouring process depends on temperature according the well-known Arrhenius law. In Fig. 4 an example is shown for the colour distribution of the same batch of fruits stored at four temperatures.

Practical Examples

Applying these views and findings to experimental data is even more complex and cumbersome than shown in these theoretical examples. In the first place, not too many data sets are available. In the second place, reality is more complex than theoretical analyses. One example using the described approach can be found at a poster presented at this conference (Tijsskens et al. 2001). A few other examples from practice can be found in Tijsskens et al. 2000.

CONCLUSIONS

To understand the occurrence and the behaviour of biological variance in agricultural produce, attention has to be devoted more to the processes involved and less to the observed phenomena themselves.

Combining good and reliable models to describe the behaviour of individual units in a batch with the fundamental knowledge available in statistics, one can arrive to an interesting new approach to this old nuisance.

The behaviour of biological variance and the corresponding distribution functions, theoretical deduced for very simplistic schemes and processes, are already directly connected to procedures used in practice.

The number of individuals in a batch, the standard deviation in maturity of the populations and the temperature during storage have all three a major effect on the behaviour and development of biological variance.

Literature Cited

General

- Allen, J.C. (1988): Averaging functions in a variable environment: a second order approximation method. *Environmental Entomology* 17, 621-625.
- Kramer, A., Twigg, B.A. (1970): *Quality Control for the Food Industry*, AVI Pub. Co, Westport
- Tijkskens, L.M.M., Wilkinson, E.C. (1996): Behaviour of biological variability in batches during post-harvest storage. AAB Modelling "Confer"ence: Modelling in Applied Biology - Spatial Aspects. p. 267-268, 24-27 June, Brunel University, Uxbridge, UK.
- Tijkskens, L.M.M., Hertog, M.L.A.T.M., van Dijk, C. (1998): Generic modelling and practical applications. Proceedings COST 915 & COPERNICUS/CIPA workshop B.M. Nicolai, J. de Baerdemaeker (eds): *Food Quality Modelling*, p. 145-151, June, Leuven, B.
- Tijkskens, L.M.M., Hertog, M.L.A.T.M., van Kooten, O., Simčič, M. (1999): Advantages of non-destructive measurements for understanding biological variance and for modelling of the quality of perishable products. Proceedings 34. Vortragstagung der DGQ: Zerstörungsfreie Qualitätsanalyse, p.13 - 24, Freising Weihenstephan 22-23 March 1999
- Tijkskens, L.M.M., Konopacki, P., Simčič, M., Hribar, J. (2000): Biological variance in agricultural products. Colour of apples as an example. Proceedings IIR Conference: Improving Postharvest Technologies of Fruits, Vegetables and Ornamentals, 19-21 October 2000, Murcia, Es. (In press).
- Tijkskens, L.M.M., Konopacki, P., Simčič, M. (2001): Biological variance in Agricultural Products. Experimental examples. Proceedings 8th International Controlled Atmosphere Research Conference CA2001, 8 - 13 July 2001, Rotterdam, the Netherlands.

Modelling

- Heaton, J.W., Marangoni, A.G. (1996): Chlorophyll degradation in processed foods and senescent plant tissues. *Trends in Food Sci and Tech* 7, 8-15.
- Hertog, M.L.A.T.M., L.M.M. Tijkskens, P.S. Hak (1997a): The effects of temperature and senescence on the accumulation of reducing sugars during storage of potato (*Solanum tuberosum* L.) tubers: A mathematical model. *Postharvest Biology and Technology* 10, 67-79
- Hertog, M.L.A.T.M., B. Putz, L.M.M. Tijkskens (1997b): The effect of harvest time on the accumulation of reducing sugars during storage of potato (*Solanum tuberosum* L.) tubers: Experimental data described, using a physiological based, mathematical model. *Potato Research* 40, 69-78.
- Schouten, R.E., Otma, E.C., van Kooten, O., Tijkskens, L.M.M. (1997): Keeping Quality of cucumber fruits predicted by biological age. *Postharvest Biology & Technology* 12, 175-181
- Schouten, R.E., van Kooten, O. (1998): Keeping quality of cucumber batches: Is it predictable? Proceedings Model-IT (Tijkskens L.M.M. & Hertog M.L.A.T.M. eds.): *Acta Horticulturae* p. 349-355, Wageningen 1998
- Tijkskens, L.M.M., Polderdijk, J.J. (1996): A generic model for keeping quality of vegetable produce during storage and distribution. *Agric. Systems* 51, 431-452
- Tijkskens, L.M.M., Sloof, M., Wilkinson, E.C., van Doorn, W.G. (1996): A model of the

effects of temperature and time on the acceptability of potted plants stored in darkness. *Postharvest Biology & Technology* 8, 293-305

Tijksens, L.M.M., Rodis, P.S., Hertog, M.L.A.T.M., Proxenia, N., Van Dijk, C. (1999): Activity of pectin methyl esterase during blanching of peaches. *J. Food Eng.* 39, 167-177.

Van Boekel, M.A.J.S. (2000): Kinetic modelling in food science: a case study on chlorophyll degradation in olives. *J. Sci. Food Agric.* 80, 3-9.

Van Dijk C., Tijksens, L.M.M. (2000): Mathematical modeling of enzymatic reactions as related to the texture of fruits and vegetables after storage and mild preheat treatments. In S.M. Alzamora, S.M. Tapia, A. López-Malo (eds.): *Design of minimal processing technologies for fruit and vegetables*, Aspen Publ Co. USA.

Van Kooten, O., Schouten, R.E., Tijksens, L.M.M. (1997): Predicting shelf-life of cucumbers (*Cucumis Sativus* L.) by measuring colour and photosynthesis. *Proceedings of the International Conference Sensors for Nondestructive Testing*, p. 45-55 (Northeast Regional Agricultural Engineering Service), February 18-21, Orlando, Florida, USA.

Figures

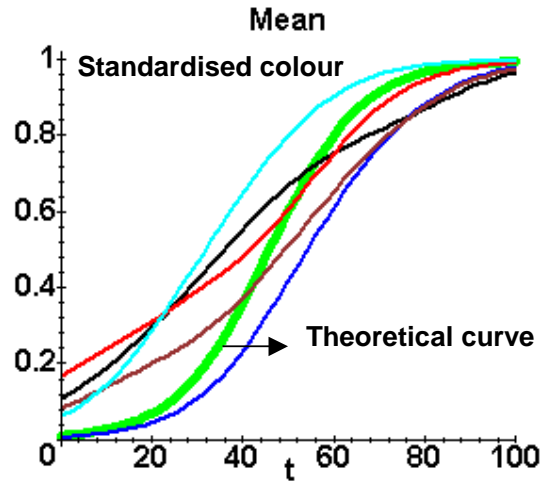


Fig. 1. Example for five batches of five individuals each, taken from a population with a standard deviation in maturity of 20 time units. The rate of colour development was set at 0.1/ time unit.

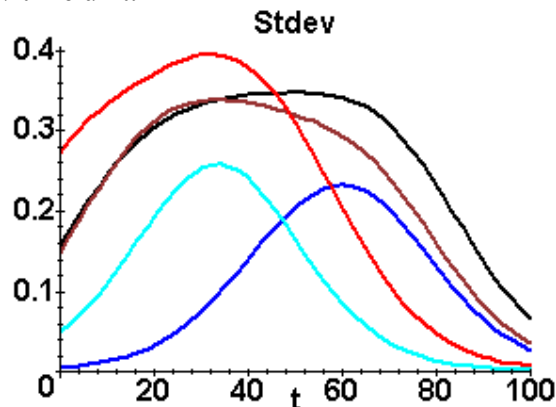


Fig. 2 Corresponding standard deviation in colour during development (storage).

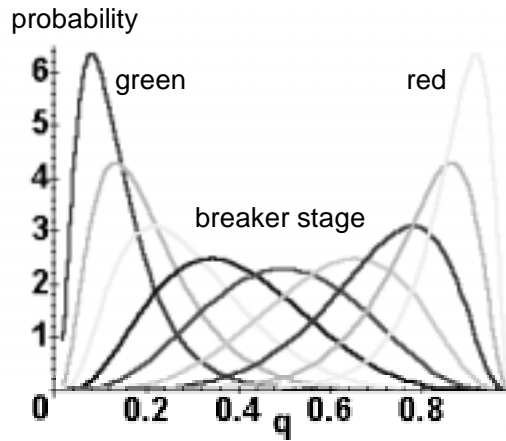


Fig. 3. Distribution in quality attribute q , assuming logistic behaviour in time, with a stdev of 7 time units, with a rate constant of 0.1 / time unit for t_m values of -20 (unripe) to $+20$ (ripe).

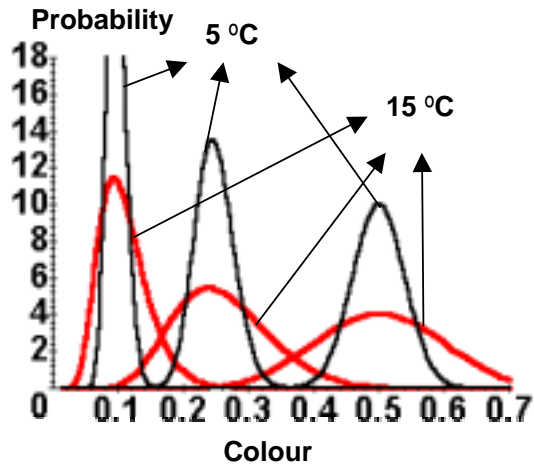


Fig. 4. Different shapes of the same distribution during storage at 5 and 15 °C, at three stages of storage: at start left), at $\frac{1}{4}$ (middle) and $\frac{1}{2}$ (right) of the product's expected life range.