Monitoring strawberry production to get grip on strawberry quality

GreenCHAINge Fruit & Vegetables WP3, BO-29.03-001-010

Fátima I. Pereira da Silva, Sabine K. Schnabel, Bastiaan Brouwer and Manon G. Mensin

WAGENINGEN
UNIVERSITY & RESEARCH

Monitoring strawberry production to get grip on strawberry quality

GreenCHAINge Fruit & Vegetables WP3, BO-29.03-001-010

Authors: Fátima I. Pereira da Silva, Sabine K. Schnabel, Bastiaan Brouwer and Manon G. Mensink

This research project has been carried out by Wageningen Food & Biobased Research commissioned by the Dutch Ministry of Agriculture, Nature and Food Quality and funded by TKI (project number 6239090303).

Wageningen Food & Biobased Research
Wageningen, December 2018

Public

Report 1940

WAGENINGEN
UNIVERSITY & RESEARCH

Version: final
Reviewer: Eelke Westra
Approved by: Nicole Koenderink
Client: the Dutch Ministry of Agriculture, Nature and Food Quality
Sponsor: TKI Foundation TKI Horticulture, Driscoll's and Bakker Barendrecht

This report can be downloaded for free at https://doi.org/10.18174/503786 or at www.wur.eu/wfbr (under publications).

PO box 17, 6700 AA Wageningen, The Netherlands, T + 31 (0)317 48 00 84, E info.wfbr@wur.nl, www.wur.eu/wfbr. Wageningen Food & Biobased Research is part of Wageningen University & Research.

Contents

Summary

The Greenchainge project is a large project financially supported by the industry and Foundation TKI Horticulture comprising different sub-projects focussing on different fruit and vegetable products. One of the sub-projects (work package 3) is dedicated to strawberry and is carried out with and by Driscoll's BV, Bakker Barendrecht BV and Wageningen Food and Biobased Research (WFBR). One of the main goals of the soft fruit project is to contribute to the understanding of strawberry quality and as such pave the way towards controlling quality to supply high and constant strawberry quality. Therefore, one of the key research question in this project is which chain parameters affect quality directly.

Hence, a large-scale quality monitoring research was set up together with the companies involved in the project. The main goal of this monitoring research was to get insight in the pre and post-harvest parameters that influence the quality of strawberries, meaning the quality at harvest and shelf life.

The scope of the monitoring was well defined to allow a sound data analysis. Hereto specific production fields at both companies were assigned for the monitoring. During some years data was gathered resulting in a large data set. The data consisted of recorded growing conditions as temperature, relative humidity (RH), radiation and amount of carbon dioxide, features related to the growing system, location, cultivar, week production, etc. and quality characterisation. In addition, some batches from the same fields were sent to WFBR and a set of complementary quality measurements was carried out.

The relations between pre/post-harvest parameters and quality variables were quantified in correlation coefficients. A statistical analysis was carried out to select only those correlation coefficients that were significant. In general the correlation coefficients found between pre/post-harvest parameters and quality variables were low. This may indicate that the quality of strawberries depends on several parameters. The data analysis was done separately for both companies.

Conclusions Driscoll's data set

- The quality variables were mainly related to different growth condition parameters:
    - For brix: temperature during the growth period is more relevant than the RH;
    - For Vscore day 8 (shelf life variable): the RH is more correlated to this quality parameter than the temperature.
- The calculated parameter Tempdiff (temperature difference between day and night) is rather valuable and more relevant to monitor than the temperature only.
- Combining the temperature and RH in the water vapour deficit (wdd) leads as well to a relevant parameter to monitor.
- Regarding the effect of the growing degree hours period (GDH) defined in this research it can be concluded that in general the correlation coefficient increased with increasing GDH period. This means that monitoring the pre-harvest parameters for a longer period before picking improves the average correlation coefficient. This effect is for temperature and temperature difference larger than for the RH and wdd.
- Monitoring the temperature and temperature difference through the whole production period, from planting to picking moment gives a better correlation with quality than monitoring only the period close to picking.
- Monitoring the RH from 7500 – 10000 GDH until the harvest correlates the best with the product quality.
- The 5000 GDH is a particular relevant moment for both the temperature difference and for the water vapour deficit monitoring.

Conclusions Bakker Barendrecht data set

A small amount of statistical significant correlation coefficients was obtained from the data set of Bakker Barendrecht. However the correlation coefficients found support the conclusions based on the Driscoll's data set.

Conclusions batches measured at WFBR

Regarding the Driscoll's batches measured at WFBR it can be concluded that the firmness seems to be negative correlated with both the cumulative week production and the cumulative week production corrected. This means that the more strawberries are produced in a production field the softer the strawberries seem to be. Based on these results it can also be concluded that the cumulative week production is a more valuable parameter to predict strawberry quality than the week production only.

The shelf life (both corrected and uncorrected) and the decay score at the beginning of the storage period at WFBR seem to be related to all four pre-harvest parameters (RH, Tempdiff, wdd and temperature). Hence the RH is in general the most relevant pre-harvest parameter for the decay. This is in agreement with the trends found in the analysis of the Driscoll's data.

It should be stressed that the conclusion are based on the current research scope, focussed on one cultivar, 4 regions in Europe, several growers and throughout the whole year. This scope led to a large number of records which also increased the noise (due to larger mistakes chance) and variance in the data (several cultivers, growers, etc). This may be another reason why the correlation coefficients found were low. The conclusions are not based on an analysis of variance (ANOVA), therefore the conclusions give indications and reflect trends. Moreover, only the statistical significant correlation coefficients were considered in the data analysis. The correlations indicate a possible causal relationship. If the link is indeed causal will have to be proven in practice.

1 Introduction

1.1 Background

The Greenchainge project is a large project financially supported by the industry and Foundation TKI Horticulture comprising different sub-projects focussing on different fruit and vegetable products. One of the sub-projects (work package 3) is dedicated to strawberry and is carried out with and by Driscoll's BV, Bakker Barendrecht BV and Wageningen Food and Biobased Research (WFBR).

One of the main goals of the soft fruit project is to contribute to the understanding of strawberry quality and as such pave the way towards controlling quality to supply high and constant strawberry quality. Therefore, one of the key research question in this project is which chain parameters affect quality directly. Hence, a large-scale quality monitoring research was set up together with the companies involved in the project. This data set delivers the unique opportunity to search for links between a wide number of chain aspects and the actual product quality.
This document reports the approach and results of the monitoring research that has been conducted objectively and independently.

1.2 Objective

The main goal of this monitoring research to get insight in the pre and post-harvest aspects that do influence the quality of strawberries, meaning the quality at harvest and shelf life (figure 1). The objective is to find out how strong do these parameters affect quality i.e. what is the correlation between growing conditions and quality in a well-defined production and chain setting. Another important aspect of this work is related to the development of a quality prediction model for strawberries. The knowledge and information generated within the monitoring research forms the basis for building the prediction model.

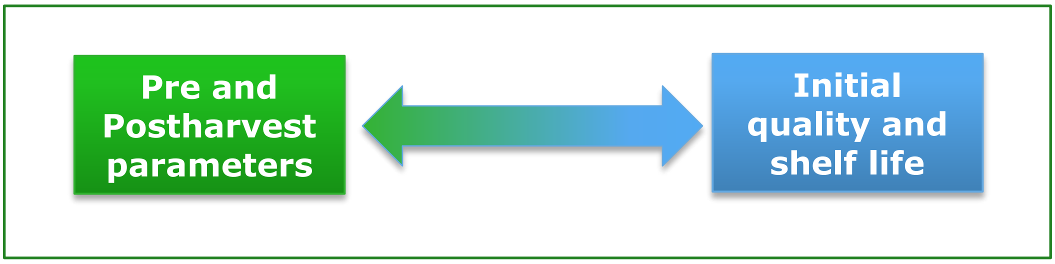


Figure 1 The main objective of the monitoring is to find relations between pre/post harvest parameters and strawberry quality.

1.3 Report

This document is intended to report the extensive work carried out in the monitoring research to all parties involved in the soft fruit project. The success of this study was very much depending on how to approach it and thus a large amount of resources was invested in the setup of the monitoring. Therefore, the following chapter describes the steps that were taken and choices that have been made in detail. Chapter 3 contains the results and discussion. Finally, the main conclusions are drawn in the last chapter.

2 Research approach

The following steps were followed to achieve a sound and accurate monitoring of quality that was used to identify correlations between pre/post-harvest parameters and quality

2.1 Step 1: extensive inventory of relevant pre and postharvest parameters

The experts involved in the project have extensive knowledge on the parameters that could affect quality. Therefore, the work was started by gathering the current knowledge. Hereto an excel sheet was set up to collect this information. The next chain information levels were defined:

- Basic grower information (geographical localisation, growth system, etc.)
- Growing conditions
- Postharvest handling
- Client

For each one of these levels monitoring parameters the next aspects were defined:

- Monitoring parameter
- Options
- Measuring method
- Priority

For instance, for the level "growing conditions" the parameter "temperature" was identified as a monitoring parameter that could affect quality. And for the level "Basic grower information" the parameter "growing media" was identified as monitoring parameter that could affect the quality. For each parameter the respective options were also identified. For the parameter "growing media" the options "soil and substract" were identified for instance. In addition to the options, the measuring method to record the monitoring was also established.

One of the most important elements of this excel sheet was the priority, which was scored between 1 and 3. The priority expresses how relevant the respective parameter is to the quality. A score of 1 means that the parameter has a high influence to the quality whereas a 3 means that the influence is low. Expert knowledge was used to establish the priority score of each monitoring parameter. The final excel sheet is presented in appendix 1.

2.2 Step 2: Choose parameters and establish methodology

The choice of monitoring parameters was done based on the priority level established in the previous step. However, in practical terms it was not possible to record all parameters with a level 1 priority. After discussion within the consortium a set of monitoring parameters was decided upon. These parameters are presented in table 1. Also, clear protocols were established to measure the chosen parameters. Unambiguous measuring methods / protocols were agreed upon to ensure that standard protocols were used overall and by both companies.

Table 1 Pre and Postharvest parameters recorded during the monitoring.

Type	Parameters
Dynamic growing parameters	
	Temperature and relative humidity
	Radiation, CO2 concentration
Static growing parameters	
	Growing system, growing medium, plant density
	Variety
	Grower
	Planting type
	Water source
	Latitude
Time related parameters	
	Duration of growth
	Planting date
	Day number, week number, year (all of production date)
	Average time between picking and cooling
Quantity parameters	
	(Corrected) week production
	Uncorrected week production
	Cumulative week production

The temperature and relative humidity (RH) were measured with sensor-data loggers placed on the production spot, protected from direct sun light and rain exposure (figure 2).

Figure 2 Sensor-data loggers placed in one of the monitoring fields (source: Driscoll's).

In the greenhouse production system the atmosphere CO_2 concentration and radiation were measured with the normal production sensors of the greenhouse. In the open field production the radiation was obtained through weather data and the normal CO_2 concentration in air was used (renders a constant value).

2.3 Step 3: Define monitoring scope

The scope of the monitoring had to be well defined to allow for a sound data analysis. In addition, it had to include enough variation in the pre and post-harvest parameters so that the correlations between these parameters and quality could be studied. The following options were chosen by the consortium:

- two strawberry varieties (Driscoll's: "Lusa" and "Scarlet"; Bakker Barendrecht: "Elsanta" and "Murano")
- two geographical locations: North (The Netherlands, Belgium) and South Europe (Spain and Marocco)

- open field and greenhouse production
- 2-3 different growing systems and mediums (greenhouse/open field; substract/soil)
- 2-3 growers for each geographical location
- Several seasons/production years (Driscoll's has carried out the monitoring during 3 seasons: 2016, 2017 and 2018. Bakker Barendrecht only in 2016.)

Unfortunately it was not possible to implement a full matrix, for instance there is no data available for the combination greenhouse production in the south of Europe, nor is data available for the combination Elsanta in South of Europe and not all growers produce both cultivars. Scarlet was not taken along in the monitoring of 2018.

In agreement with the growers a specific field/lot was chosen for the monitoring (the growers were selected as described in the scope). The sensors-data loggers were placed on those fields at the same time as the plants were planted (this is not the case for fields in the south of Europe in the first year; the sensors were placed later). All batches that were produced on those fields were included on the monitoring data set. At the end of the production period the data of each data logger was downloaded. Also the data of all other pre and post-harvest parameters was collected in the respective forms (see the chosen parameters in step 2). Finally, the respective quality and shelf life data of all monitored batches was also collected.

It was agreed that both companies would use their own quality assessment protocols. In doing so the 2 data sets of the companies could not be combined (the companies apply different quality parameters and use different quality assessment protocols). Therefore it was agreed to send samples of the monitoring batches to WFBR in Wageningen, The Netherlands; a set of monitoring batches was selected for the WFBR measurements. The samples of both companies were assessed according to the same protocols allowing the combination of both data sets (see section 2.5).

2.4 Step 4: Data acquisition and data analysis

A number of excel sheets and forms was used to register the data. In order to be able to relate different parameter's records to the same batch, a unique code per batch was used to unambiguously identify each batch. Twice per year the collected data was sent to WFBR for further analysis.

Three type of data was registered within the monitoring:

- Dynamic data: consists of data that is collected over time (amount of water, temperature, feeding plan, etc.); This data was recorded with different frequencies (some per day, per week, per month, per hour depending on the variable).
- Single data: this data consists of one value.
- Text data: this is the descriptive type of data or open field data (registration of exceptional conditions); this data has to be divided in levels/categories. Therefore the possible levels/categories have to be established first and presented in the registration form as such.

The dynamic data was particularly a challenge since the relation between pre and post-harvest parameters cannot be studied based on thousands of temperature or RH values over time. This type of variables had to be processed into a single value to be used in the further data analysis. In cooperation with all partners the best procedure to handle this kind of data was discussed and agreed upon. The following growing degree hour periods that are relevant for the fruit development were defined:

- 500 Growing Degree Hour (GDH)
- 2000 GDH
- 5000 GDH
- 7500 GDH
- 10000 GDH
- 12500 GDH
- 15000 GDH.

Temperature and relative humidity
For each GDH period, a matrix of values was calculated: the minimal, maximal and average temperature/RH for the night, the day or the whole 24 hours period (3x3= 9 variables per GDH). The night and day period were defined using the sunset and sunrise of the location and date of the respective batch.

Radiation and CO_2 concentration
For each GDH period the average day radiation and CO_2 concentration was calculated (radiation and CO_2 concentration values overnight were not used).

Combination of variables: Temperature difference and VPD calculation
Based on the temperature and RH records the vapour-pressure deficit (VPD) was calculated. In addition the temperature difference between the maximum day temperature and the minimum night temperature was calculated (thus one value per 24-hour period).

Similar to the temperature and RH, for each GDH period a matrix of values was calculated: the minimal, maximal and average VPD for the night, the day or the whole 24 hours period.

Regarding the temperature difference, only the average temperature difference for each GDH was calculated.

The data analysis consisted of the steps described in figure 3.

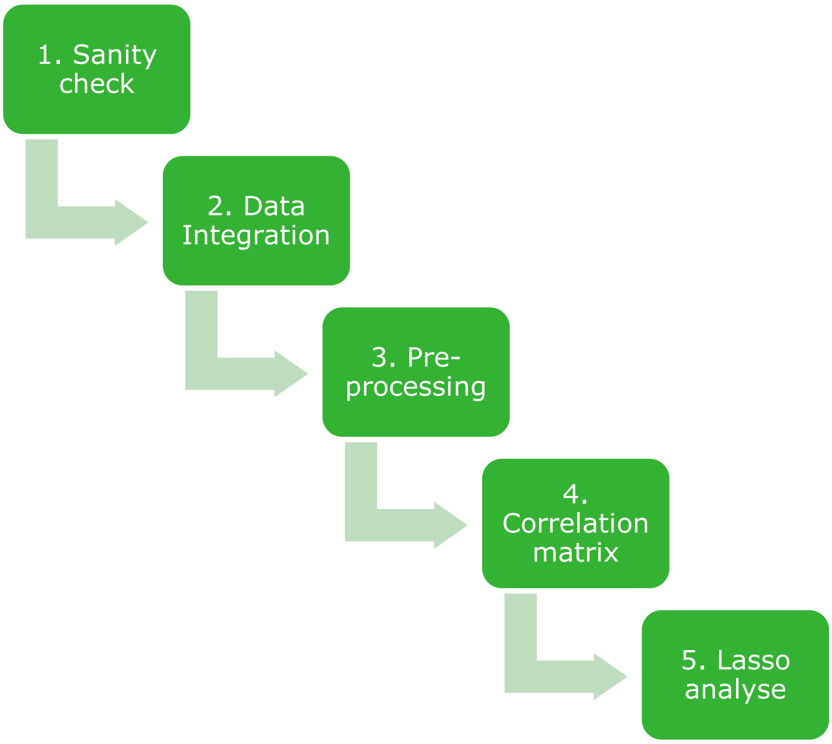


Figure 3 Steps in the data analysis of the monitoring data.

The first step in the data analysis was the so-called sanity check. The objective of this step is to:
- remove errors (wrong batch codes);
- assure that missing values are an empty cell;
- parameters with a zero value indeed show a zero in the respective cell;
- make sure that the data format is consistent through the whole data set (in some cases the download settings of the data loggers are not exactly the same resulting in different formats at the moment that the data is downloaded).

The second step is the data integration. All available/recorded data – both the pre/post-harvest parameters and the respective quality assessment data - of a specific batch has to be connected to this same batch code (alignment of the data). A macro was written in excel to support this step.

The final data preparation step is the pre-processing of data. This step includes the calculations on the dynamic data described above. In this step the GDH period and respective time period for each batch and defined GDH period were calculated and established. From the temperature records over time the daily amount of GDH and the cumulative amount of GDH were calculated starting from the harvesting moment of each batch. Afterwards the period of time corresponding to a specific GDH period for each batch was searched for in the temperature-time-GDH file (the harvest date of each batch is known and from this date back the GDH periods are established). Subsequently the calculation of averages, minimal and maximal values was done within the period of time corresponding to each GDH period.

After calculating all dynamic parameters and gathering all growth and production related characteristics per batch, the correlations for this data collection were calculated. The correlations were tested for significance (at level 0.05) and corrected for multiple testing as well. As the data matrix includes a substantial amount of missing values, the careful interpretation of the correlation values should take into account the actual sample size that went into a specific correlation between two variables.

While we studied the correlations between the (numeric) pre/post-harvest parameters, it could be assumed that not all pre/post-harvest parameters have (equal) influence on the quality parameters/variables. As we were also confronted with a large amount of pre/post-harvest parameters, we decided to include a variable selection step. This last step is also relevant for the quality prediction model (Bayesian network model) that was developed in the project based on the monitoring data. To this end we resorted to using the so-called Lasso technique (Least Absolute Shrinkage and Selection Operator) for variable selection for a set of quality variables that are included in our dataset. This is a regression technique which allows us to select the most important variables in our model in order to simplify the resulting model, not only in terms of complexity (less variables to include, less computing time etc.), but also in terms of interpretability. By this means the number of pre/post-harvest parameters that entered into the Bayesian network model were not only reduced but also the most important and relevant ones were selected.

2.5 Post-harvest cooling

In 2016, the cooling temperature of 25 batches of Driscoll’s (only North Europe) was also monitored with a sensor-data logger (in addition to all other pre and post-harvest parameters). The data loggers were placed in between packages short after harvesting.

Also in this case the logged temperature had to be processed to a single value to allow further data analysis. The following parameters were calculated from the temperature recordings and related to both the initial quality and the shelf life results of the respective batch:
- first measurement and last measurement
- minimum and maximum
- mean and median
- variance and standard deviation
- area under the curve (auc)
- number of points above 4.

Based on the results of 2016 it was decided that the cooling temperature should not be measured in the following monitoring years.

2.6 WFBR samples

As describes previously, a number of the monitoring batches samples were send to WFBR (from both companies). A fixed time schedule was followed for the samples that were transported to WFBR: harvesting- arrival at company-transportation WFBR-quality and shelf life assessment.

The schedule in 2016 was the following:

Wednesday: pick up strawberries
Thursday: transport to Driscoll's/BB and afterwards to WFBR (arrival in Wageningen Thursday afternoon)
Friday: start the quality evaluation measurements at the WFBR
Monday, Wednesday and Friday of week following week: quality evaluation at the WFBR

Parameters measured at the WFBR:
- Colour (light cabinet; whole strawberry including calyx)
- Volatiles (PTR)
- Visual inspection (according to the WFBR protocol); this includes the general quality assessment, calyx quality, fruit damage (including microbiological decay).

The schedule in 2017 was the following:

Tuesday: pick up strawberries
Wednesday: transport to Driscoll's and afterwards to the WFBR (arrival in Wageningen Wednesday afternoon)
Thursday: start the quality evaluation measurements at the WFBR
Tuesday and Friday of week following week: quality evaluation at the WFBR

Parameters measured at the WFBR:
- Colour (light cabinet; but first remove the calyx)
- Visual inspection (according to the WFBR protocol); this includes the general quality assessment, calyx quality, fruit damage (including microbiological decay).
- Firmness (Firmtech)

The samples/punnets were coded upon arrival and stored at 4°C and 80% RH over the assessment period. The day before the quality assessment the respective punnets were moved to another storage cell and were warmed up to 20°C in 6 hours. For each batch and for each assessment day 5 punnets were used.

Since it was not possible to measure at WFBR all batches that were included in the monitoring, a selection of samples was done at the start of the monitoring. This selection covered the whole scope of the monitoring. In 2016, 38 batches were transported to WFBR (in total 152 punnets were measured). In 2017, approximately 50 batches were measured of the in total almost 70 planned batches. Not all planned batches were transported to WFBR, thus in total less samples were assessed.

3 Results and discussion

The two companies use different quality parameters in their daily operation and these quality parameters were used during the project monitoring leading to data sets with different quality variables. Therefore the data analysis was done separately for each company. The results of the monitoring by Driscoll's are presented in section 3.2 and for Bakker Barendrecht in section 3.4. The variable selection carried out as input for the development of a quality prediction model is presented in section 3.3. The effect of cooling temperature (after harvesting) on quality was analysed separately on the Driscoll's data. These results are shown in the next section. Finally section 3.5 considers the results of the monitoring batches transported to WFBR.

3.1 Effect post-harvest cooling

The correlation coefficients between the calculated cooling parameters (described in chapter 2) and the quality parameters are presented in annex 7. The highest number of correlations between the cooling parameters and the quality parameters (0.6 - 0.7) were with standard deviation and variance. The results are very similar which makes sense because the 2 parameters are also similar. The minimal and maximal temperature also showed some relation with the quality parameters. The mean, median, area under de curve (auc), number of points above 4°C, first and last measurement correlate less with the quality parameters.

Table 2 gives an overview of the most significant correlations between the cooling temperature parameters and quality in the 25 batches analysed. The number of batches is limited and therefore relatively higher correlation coefficents were found.

Table 2 Overview of the most significant correlation coefficients between the cooling temperature parameters and quality (n=25 batches)

Quality	Temperature parameter	Correlation coefficient	Remarks
PQFscore	Variance	-0.67	The temperature variance is negatively correlated with the PQF score. This means that if the variance in the temperature is higher, the quality of the product is in general lower. The same holds true for the standard deviation of the temperature and the maximum temperature measured.
PQFscore	Standard deviation	-0.69	
PQFscore	Maximum temperature	-0.67	
T2 condition	Variance	-0.64	The temperature variance is negatively correlated with the T2 condition. This means that if the variance in the temperature is higher, the quality of the product is in general lower. The same holds true for the standard deviation.
T2 condition	Standard deviation	-0.62	
Dry bruises	Mean	0.56	The higher the temperature average (mean) the higher the amount of bruises. The same holds true for the the temperature median.
Dry bruises	Median	0.57	

The quality parameter PFQscore shows the most and highest correlations. Also it can be concluded that the temperature recording after picking is more related to the initial quality than to the shelf life results parameters (PFQ, T2 and dry bruises are all initial quality parameters).
The higher the PQF score, the better the quality. That is the reason why the relation between this quality parameter and temperature is negative.

3.2 Correlation matrix Driscoll's data

The correlation matrix for the Driscoll's data set consists of 238 input parameters (all pre and post-harvest parameters) and 17 quality parameters. In total 3140 batches were used to calculate the correlation coefficients (r) between the 238 pre/postharvest parameters and the 17 quality variables. The sample size per correlation coefficient (between 2 variables) is the number of complete records for the two variables in question (thus records with missing values for both or one of the 2 variables in question were not used in the correlation). In total 4046 correlation coefficients were calculated. However only **the correlations that are statistically significant** (α=0.05) were considered in the results analysis (see the data analysis approach in chapter 2).

The data set of Driscoll's contained batches of both the cultivars Lusa (in total 2073) and Scarlet (in total 1067). To avoid extra variance and noise in the data analysis (due to eventual differences in the relation between pre/postharvest and quality for the different cultivars), a separate correlation matrix was calculated for Lusa and used in the data analysis presented hereafter. The amount of Lusa batches is much higher than the Scarlet batches and in 2018 Scarlet was not included in the monitoring. For this reason the analysis is focussed on the cultivar Lusa.

Figure 4 shows a small part of the large Lusa correlation matrix (the whole matrix is presented in annex 1). The empty cells mean that the correlation coefficient was not statistically significant and therefore is removed from the correlation matrix.

Correlation coefficient (r)	PFQSCORE	T2.Appeara	T2.Condition	BRIX	SevScQC1	VScore8QC1
weekproduction	0.101		0.131	-0.221	0.243	
cumweekproduction	0.077		0.105	-0.081	0.153	0.127
temp.avg.night.500		0.108	-0.051	-0.108		
temp.avg.day.500		0.172	-0.133	-0.256	0.154	
temp.avg.all.500	0.051	0.161	-0.067	-0.164	0.101	
temp.min.all.7500	0.135	0.088	0.129		-0.081	-0.263
temp.avg.night.10000	0.062	0.107		-0.163	0.099	
temp.avg.day.10000	0.056	0.205	-0.099	-0.311	0.301	
PlantingDate	0.123	-0.072	0.252	0.147	-0.147	-0.462
hum.avg.day.2000	0.088		0.168		-0.316	-0.328
hum.avg.all.2000	0.069	0.061	0.052		-0.262	-0.242
wdd.min.day.2000		-0.088	0.142	0.138	-0.104	-0.247
wdd.min.all.2000	0.064	-0.067	0.158	0.177	-0.130	-0.270
wdd.avg.night.5000	-0.053	-0.131			0.151	0.263
PlantType		0.121	-0.057	-0.094	0.259	0.197
PlantDensity	0.103	-0.077	0.227	0.125	-0.171	-0.414
RowDistance	-0.134		-0.180		0.163	0.303
PlantDistance		0.109	-0.103	-0.137	0.248	0.280
WaterSource	0.195	-0.064	0.356	0.094	0.095	-0.476

Figure 4 Small part of the large Lusa correlation matrix (the whole matrix is presented in annex 1).

As it can be seen in figure 4 the correlation coefficients are low. However these values have been used to analyse the relations between pre/postharvest parameters and quality. In order to increase the relevance of the work, 4 of the 17 quality variables were selected for this data analysis.

The 4 selected parameters give a **broad characterisation of strawberries quality**:

- T2 condition: this parameter reflects the appearance and physiological condition of the fruit close after harvest. This is a general score given at punnets level but including several condition aspects and score on scale from 1-100.
- Brix: this parameter reflects (to some extend) the taste; it is the measurement of the brix degrees close after harvest.
- Vscore day 8: this parameter reflects the shelf life/ keepability of the product (score between 1 and 3).
- Severity score: this parameter reflects the amount of fruit damage (dry bruises, wet bruising, rot) and is scored between 0-4. The evaluation is done per strawberry.

The graphs in figure 5 presents the average correlation coefficient for the dynamic pre-harvest parameters: temperature, RH, radiation, amount CO_2 and the subsequent variables water vapour deficit (wdd) and temperature difference (Tempdiff). The average includes all GDH period's, both the night, day and the 24 hour values and all calculation options (minimal, maximal and average values).

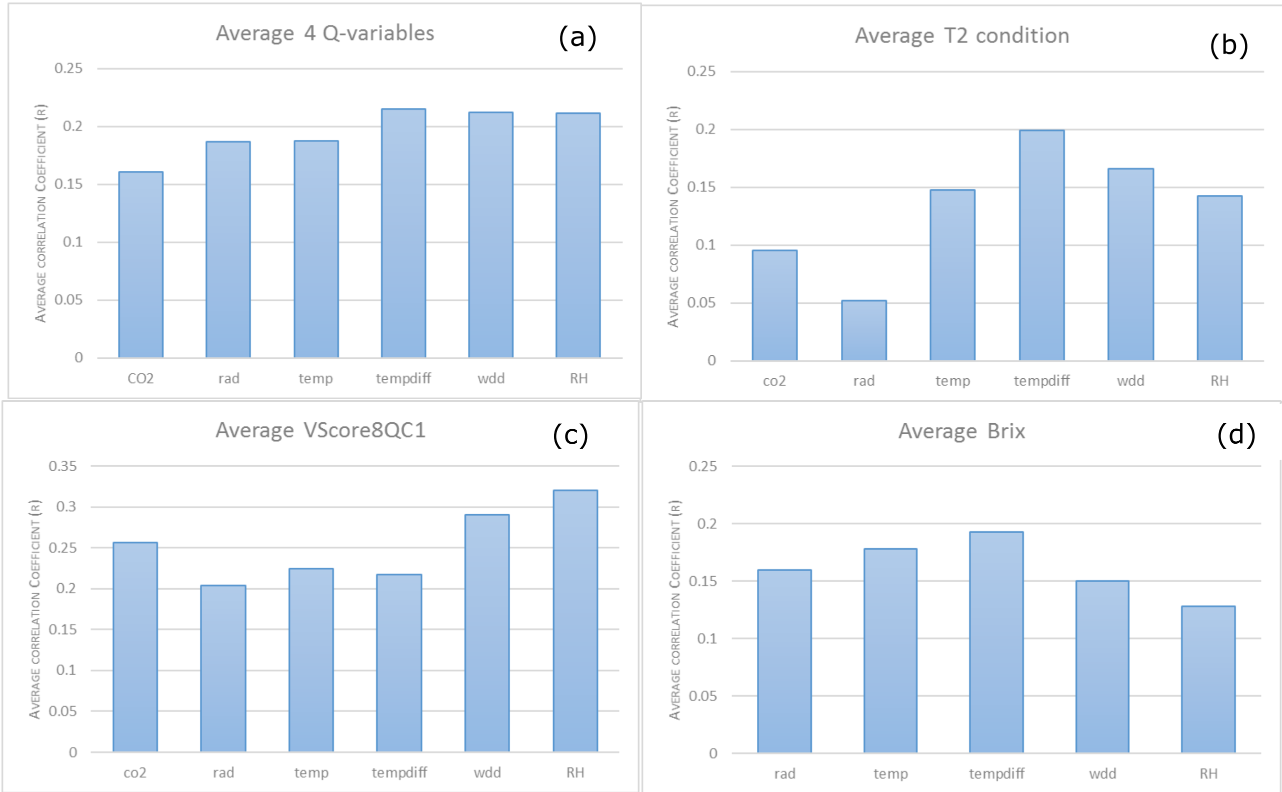


Figure 5 Average correlation coefficient for the dynamic pre-harvest parameters: temperature, RH, radiation, amount CO_2 and the subsequent variables water vapour deficit (wdd) and temperature difference (Tempdiff) on Average 4 Q-variables (a), Average T2 condition (b), Average VScore8QC1 (c) and Average Brix (d). The average includes all GDH period's, both the night, day and the 24 hour values and all calculation options (minimal, maximal and average values).

Despite the low correlations coefficients the average level of the pre-harvest parameters is different for each selected quality variable. There was a considerable amount of missing values by the parameters radiation and amount CO_2. Thus the number of statistical significant correlation coefficients for these parameters is clearly lower than for the temperature, RH, Tempdiff and WDD. Therefore the parameters radiation and amount CO_2 are less considered in the discussion of the results hereafter.

The most relevant pre-harvest parameters depend on the quality variable. For Brix (figure 5 d) the temperature seems to be more important than RH. On the other hand, for the Vscore day 8 (shelf life score parameter, figure 5 c) the RH seems to be more correlated to this quality parameter than the temperature. As such when considering the 4 selected quality variables, the role of all growth conditions parameters seems to be rather similar.

The results also indicate that the calculated parameter Tempdiff (temperature difference between day and night) is rather valuable and more relevant to monitor than the temperature only. In addition, combining the temperature and RH in the Water Vapour Deficit (wdd) leads as well to a relevant parameter to monitor. In the case of T2 condition (figure 5 b) the wdd is even slightly more related to the T2 than the temperature and the RH separately.

The graph in figure 6 represents the effect of the different GDH periods on the correlation coefficients. The average correlation for the same 4 quality variables is presented (average over the night, day and the 24 hour values and all calculation options (minimal, maximal and average values).

This graph gives insight in how the different moments during production influence quality in general. The 500 GDH period represents the period short before picking: half day to 1 day before harvest (depending on the moment in the season; if the weather conditions are warmer it means an half day only; if the weather is cooler it means approximately one day). On the other hand, the 15000 GDH period represents almost all production period from planting up to harvest.

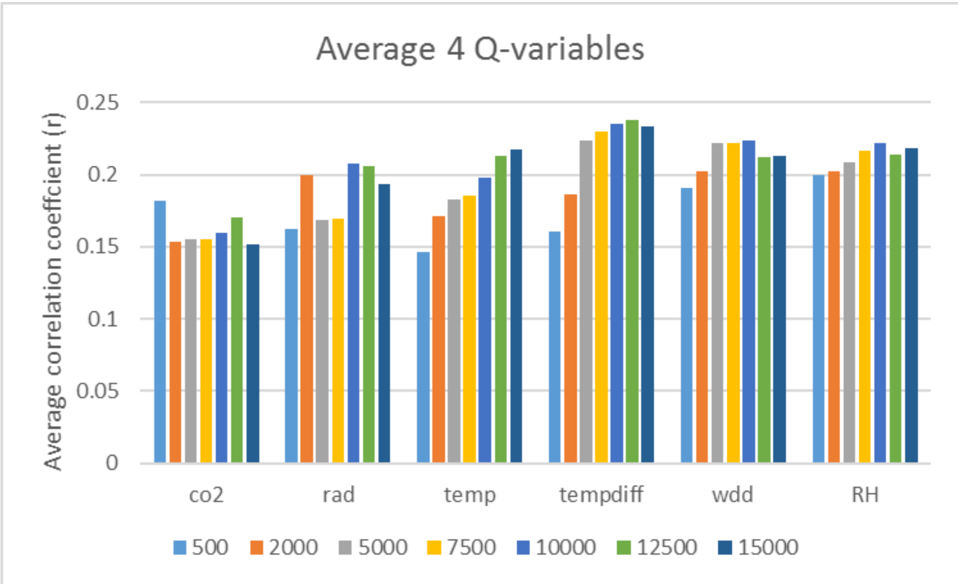


Figure 6 Effect of the different GDH periods on the average correlation coefficents of 4 selected quality variables (average over the night, day and the 24 hour values and all calculation options (minimal, maximal and average values)).

With the exception of CO_2, for all parameters the correlation coefficient increases with increasing GDH period (the radiation and amount of CO2 parameters have missing values thus the interpretation of the results requires some reservation). The parameter temperature difference (tempdiff), shows the largest increase whereas the RH the least difference. This means that monitoring the temperature and temperature difference through the whole production period, from planting to picking moment gives a better correlation with quality than monitoring only the period close to picking. On the other hand, according to the results, following the RH up to the planting moment seems to have less added value: measure the RH from 7500 – 10000 GDH until the harvest seems to correlate the best with the product quality. The 5000 GDH seems to be a particular relevant moment for both the temperature difference and for the water vapour deficit monitoring. This period corresponds to the period of time around 1 week before harvest and covers an important moment in fruit development.

The trends shown in figure 6 for the average of the 4 quality parameters is in some cases different for the separate quality variables. When analysing the effect of the GDH periods on the quality variables separately, it can be seen that:

- for Brix the correlation coefficient increases with increasing GDH period for the temperature and temperature difference but not for the RH or WDD.
- for the T2 condition and Vscore day 8 both the temperature difference and the RH increase with increasing GDH period (temperature difference increases more clear for T2 condition and RH increases more clear for the Vscore day 8).

These results are in agreement with the results shown in figure 5 here above.

Next step in the detailed data analysis is to investigate the effect of day-night and the whole 24-hours measurements on the correlation coefficients. In annex 2 the average correlation coefficient for the day-night-24 hour measurements (average over the minimal, maximal and average values) of the pre-harvest parameters for each GDH period is presented (per quality variable separately). The graphs presented in figure 7 show a summary of the detailed results in annex 2.

In agreement with the results discussed above, the effect of the day-night-24 hours measurements depends on the quality parameters. For Brix the night WDD measurements and the day temperature measurements leads to the highest average correlation coefficients. For the average Vscore day 8 and for T2 condition the differences between day, night and 24 hours period are limited. For the severity score the night temperature, night WDD and night RH measurements leads to the lowest average correlation coefficients. This explains the lower values for the night measurements in the graph of the 4 quality parameters together.

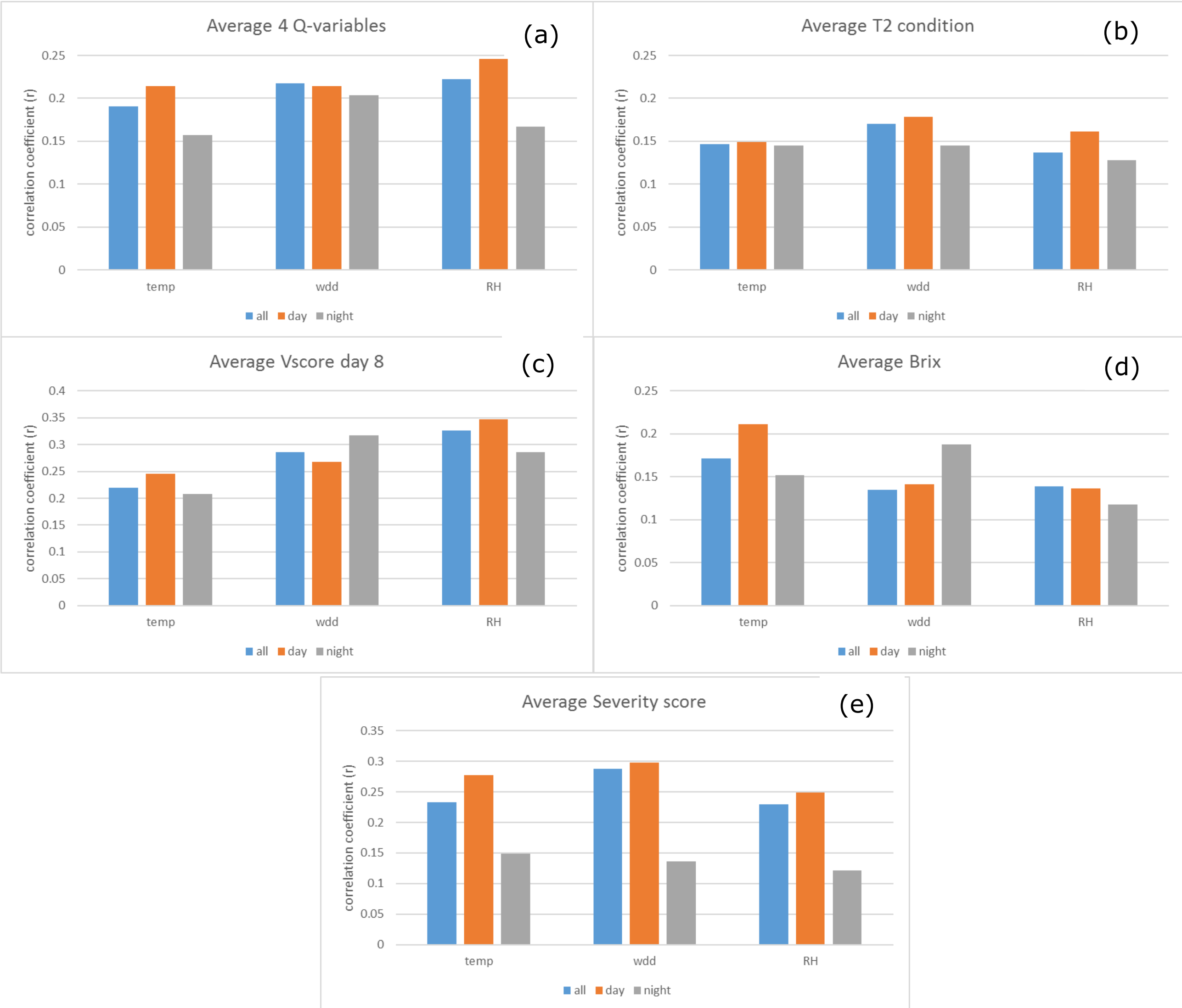


Figure 7 Effect of day-night and the whole 24-hours measurements on the correlation coefficients of some pre-harvest parameters (average of 4 selected quality variables (a), Average T2 condition (b), Average VScore day 8 (c) and Average Brix (d) and Average Severity score (e)).

The correlation coeficients of the non-dynamic parameters, as growing system, planting date, week production, etc are presented in annex 3. Several features considered in this correlation matrix are stongly interrelated. For instance the planting date is related to the type of ethylene cover, as the batches plated earlier in the season originate from a specific grower that uses a specif plant type, plant distance, type of soil, etheylene cover etc. Therefore these results have to be analysed carefully as the parameters are not fully independent.

Regarding the average correlation coefficients of the quality variables, Vscore day 8 has the highest average (0.36). The quality variables related to the initial quality PFQScore, T2.Appearance, T2.Condition and brix show lower averages. However the Vscore day 8 is measured in a smaller number of batches and that may also influence the correlation coefficient in a positive way (less samples may lead to lower variance).

It may be assumed that the parameters AvgTimePickingCooling (=average time between picking and cooling), week production (amount in kg produced strawberries in the monitoring field during the respective week), cumulative week production (sum of all week productions untill the respective week), week production corrected and cumulative week production corrected (week production and cumulative week production corrected for the amount of strawberries that the plant has produced but did not meet the commercial requirements) are more independente of the grower/growing system. The production related parameters show low correlation coefficients (average r= 0.11 or 0.15).

3.3 Results Lasso analysis (Data set Driscoll's)

The main goal of the monitoring research was to identify the pre/postharvest parameters that have the largest effect on strawberry quality. The number of pre/postharvest parameters is large and there is an high inter-relation between the variables (for instance several temperature parameters are strongly connected with each other). Therefore a Lasso analysis is carried out to select those pre/posthavets parameters that contribute the most to predict each quality variable. For each combination pre/postharvest parameter-quality variable a relevance coefficient is calculated. The higher this coefficient the more relevant the parameter. Table 3 shows the top 5 result of the Lasso analysis for the most importante quality variables. There is no Lasso analysis carried out for T2 appearance because this variable does not depend on the growing system or growing conditions. Togther with product experts it was decided to predict T2 appearance based on the cultvar, production week, grower, latitude and week production.

Table 3 Top 5 of the Lasso analysis for 5 quality variables (over all pre/postharvest parameters). The number on the pre-harvest parameter indicates the GDH period. Day/night/all refers to the measuring moment, respectively during the day, night or for the 24-hour period. Avg- average; min = minimal value and max = maximal value.

T2 condition	Severity score	Vscore day 8	Brix	Rot
wdd.avg.day. 7500	wdd.max.night. 5000	hum.min.day. 2000	hum.max.all. 2000	temp.max.night. 500
temp.max.day. 15000	wdd.max.night. 7500	hum.min.all. 2000	tempdiff.10000	hum.avg.night. 12500
wdd.min.nigh. 5000	Week production	hum.min.day. 10000	temp.min.day. 5000	Plant Type
temp.avg.all. 15000	wdd.max.day.500	hum.min.all. 10000	tempdiff.2000	wdd.avg.day. 2000
temp.avg.night. 15000	wdd.max.night. 10000	hum.avg.day. 7500	hum.min.night.15000	temp.min.day. 12500

The relevance coefficient of the top 5 parameters for each quality variable is rather similar, meaning that the parameter at the first position is not much more stronlgy related to the quality variable than the one in the second position. In addtion, not all pre/postharvest parameters have a relevance coefficient: several have a zero coefficient and thus do not contribute at all to predict the respective quality variable. The number of pre/postharvest parameters with a relevance coefficient, thus that do contribute to predict the quality, depends on the quality variable. For Rot there are only 12 parameters with a relevant coefficient wheras for Vscore day 8 there are 120 parameters with a relevant coefficient. Regarding the non-dynamic parameters as for instance the type of growing system, water source, planting date, etc the week production and corrected week production are the most relevant ones. In general these parameters appeared on a lower position in the lasso analysis indicating that the growth conditions as RH, temperatture, wdd and tempdiff have a stronger effect on the quality of strawberries than the growing system, water source, plant distance, etc (rot is however an exception; for this quality variable, the non-dynamic parameters showed high relevance coefficients).

The top 5 from the lasso results cannot fully be compared with the results of section 3.2 as both results were obtained following different approaches and with different objectives. The results and conclusions of both approaches are complementary and together contribute to the understanding and prediction of strawberry quality. The results are however in general in good agreement with each other.

3.4 Correlation matrix Bakker Barendrecht data and batches sent to WFBR

The monitoring data from Bakker Barendrecht was collected in 2016 at two growers. At one of the growers, no sensor-data logger was placed in the field and therefore the growth conditions were not recorded. The correlation matrix between the pre/post-harvest parameters of both growers and the quality variables – in total 44 batches - is presented in figure 8 (without the dynamic growing condition parameters as only one of the growers had recorded them). To facilitate the consultation of the correlation matrix, the same figure is also presented in annex 4.

The temperature after harvest was recorded and therefore the matrix includes the parameters related to the post-harvest cooling as described in session 2. The quality variables measured include:

1. quality variables measured at the grower: dry bruising old self, wet bruising new self, first impression self (=1stImprSelf), Colour self and Size self;
2. quality variables measured at Bakker Barendrecht: dry bruising old Bakker, wet bruising new Bakker, first impressionBakker (=1stImprBakker), Colour Bakker, Size Bakker, Taste evaluation, Brix, Acid average, Brix to acid ratio and End results shelf life;
3. quality variables measured at WFBR: score calyx after 7 and 10 days storage (= Calyx day7 and calyx day10), decay score after 7 and 10 days storage (=decay day7 and decay day10) and shelf life calyx and decay (this parameters represents the shelf life, ie the number of days in storage until the acceptance level was reached).

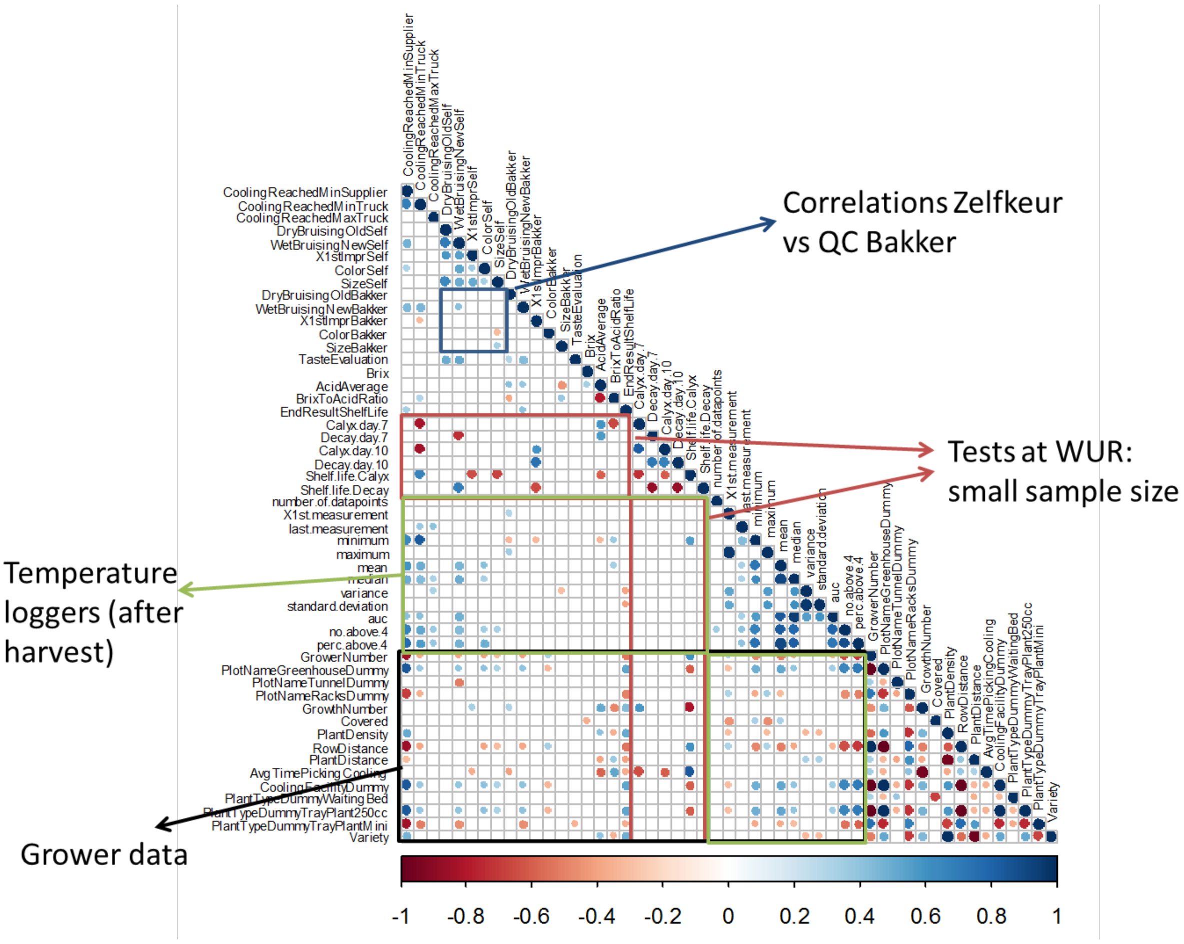


Figure 8 Correlation matrix between the pre/post-harvest parameters of both growers and the quality variables measured by the grower (Zelfkeur), Bakker Barendrecht (QC Bakker) and at WFBR; n=44 batches.

Only statistical significant correlations (α=0.05) are presented in the matrix (thus the absence of a colour dot means that the correlation that was found was not statistical significant). The stronger the correlation the stronger the colour blue for positive correlations or red for negative correlations. The colour bar under the matrix shows the relation between the colour and the correlation value. The relevant pieces of the matrix are highlighted with colours:

- In black the relation between growing system (ao greenhouse, tunnels, racks, etc) and quality.
- In green the relation between the temperature during cooling (data loggers placed after harvest) and quality.

- In red the relation between quality parameters measured at WFBR ("tests at WUR") and both growing system parameters and temperature during cooling parameters.
- In dark blue the relation between the quality measurements done by the grower self and the quality measurements done at Bakker Barendrecht.

As it can be seen the correlations are low. Figure 9 shows an overview of the relations with a correlation coefficient above 0.55 (black, green and part of red area of the matrix). Only the parameters "shelf life calyx", "calyx score day7", "calyx score day10" measured at WFBR and "acid average" measured at Bakker Barendrecht seem to have a link to the growth system and cooling conditions. It is not clear why almost only the calyx parameters show some relation with the growth system and cooling temperature and most probably this is a data artefact. A higher "shelf life calyx" parameter means a longer average shelf life of the samples (regarding the freshness of the calyx), thus a better quality. The longer the average time between picking and cooling, the lower the calyx shelf life is expected. The correlation should therefore be negative, and is hence not in accordance with the results.

	shelf life calyx	**other Quality parameters**
CoolingReachedMinTruck	0.66	calyx day 7/10 (-0.82; -0.85)
X1st.measurement	_	
last.measurement	_	
minimum	0.55	
maximum	_	
mean	_	
median	_	
variance	_	
standard.deviation	_	
auc	_	
no.above.4	_	
perc.above.4	_	
Grower Number	0.61	calyx day7 (0.59)
PlotName Greenhouse	-0.61	
PlotNameTunnel	_	
PlotNameRacks	_	
Growth Number	-0.82	
Covered	_	
PlantDensity	_	
RowDistance	0.61	
PlantDistance	_	
AvgTimePickingCooling	**0.85**	acid average (-0.58) and calyx day 7/
Cooling Facility	-0.61	
Plant Type Waiting Bed	_	
PlantType TrayPlant250c	-0.61	
PlantType Tray Mini	_	
Variety	_	

Figure 9 Overview of the relations with a correlation coefficient above 0.55 (r > 0.55)

The dark blue area shows that the relation between the quality parameters measured at the grower and by Bakker Barendrecht is very poor. When comparing the relation between the quality parameters measured by the WFBR and the quality parameters from the grower and Bakker Barendrecht, it seems that there is a good relation between the wetbruising measured at the grower and the decay parameters measured at the WFBR (detailed results are presented in annex 5). The first impression measured at Bakker seems to be reasonably related to the WFBR quality parameters. Nevertheless more links were expected, particularly for the parameters dry bruising, wet bruising and decay.

An extra analysis is carried out for the grower where the growing conditions were monitored. The data set consisted of 31 batches. The CO_2 concentration and radiation were not monitored. The temperature and RH (RH/hum) records collected were pre-processed as described in chapter 2 (with the exception of the temperature difference and water vapour deficit). Finally a correlation matrix was also generated (see annex 4). Few relationships showed a statistically significant correlation.

The relations with a correlation coefficient above 0.55 are presented in table 4. None of the quality parameters measured by the grower or by the Bakker Barendrecht showed a correlation coefficient coefficient with the growing conditions higher than 0.55. The score for calyx and decay at day 10 of the storage period (measured by WFBR) seem to be related to the RH parameters. The data of 2016 contained only the 500, 2000 and 5000 Growing Degree Hour periods.

Table 4 Overview of the relations with a correlation coefficient above 0.55 (r > 0.55).

	Calyx.day.10	Decay.day.10
hum.avg.night.2000	_	-0.72
hum.avg.day.2000	_	_
hum.avg.all.2000	_	_
hum.min.night.2000	-0.75	_
hum.min.day.2000	_	_
hum.min.all.2000	_	_
hum.avg.night.5000	_	_
hum.avg.day.5000	_	-0.75

On average, the correlation RH and quality (average r=0.48; # correlation coefficients = 25) is slightly higher and showed more correlation coefficients than temperature and quality (average r=0.45; #correlation coefficients =13).

In total 6 different temperature parameters (temp.avg.day.500, temp.avg.all.500, temp.max.day.500, temp.max.all.500, temp.avg.day.2000 and temp.max.night.5000) correlated to brix, whereas only 2 RH parameters (hum.avg.day.500 and hum.avg.day.2000) correlated with Brix. This trend seems to be in agreement with the results found in the data set of Dricoll's. Moreover the decay score, a quality variable reflecting strawberrie's condition and shelf life, only showed correlation coefficients with RH parameters as shown in table 4. This result is also similar to what was found in the Driscoll's data analysis.

Due to the limited number of batches (44 or 31) combined with a large number of constant (and missing) values it can be concluded the data is not sound enough to carry out a lasso analysis for variable selection.

3.5 Results Driscoll's batches sent to WFBR

As mentioned previously, a number of the batches used to monitor the relation between shelf lfe/quality and pre/postharvest parameters were sent to the WFBR for extra shelf life/quality measurements carried out at WFBR. The results of the Bakker Barendrecht batches was discused in the previous section. The results of the Driscoll's batches will be presented in this sesction.

Despite the limited number of batches, the correlation coefficient (r) between the 238 pre/postharvest parameters and the shelf life/quality variables was calculated for both Lusa and Scarlet together (no separate analysis for each cultivar). This resulted in a large matrix which is analysed in a similar way as the rest of the Driscoll's data as presented here above. Likewise only **the correlations that are statistically significant** were considered in the results analysis.

The two major quality parameters measured at WFBR reflecting the strawberries condition are the calyx score (refects the calyx freshenss) and decay score (reflects the amount of dry/wet bruising and mould development). These parameters were measured during the storage period as described in chapter 2. An acceptance limit was set for both parameters and based on this limit the shelf life (in days) was calculated. Since the time between harvest and arrival at WFBR was not the samen for all batches the shelf life was corrected by adding the number of days between harvest and arrival of each respective batch. The correlation matrix is calculated for both the uncorrected as the corrected shelf life. The graph in figure 10 shows the average correlation coeficient for the pre-harvest (growing conditions) parameters. Since several radiaton and amount carbon dioxide records were missing and this analysis is done on a much smaller number of batches, these parameters were not considered.

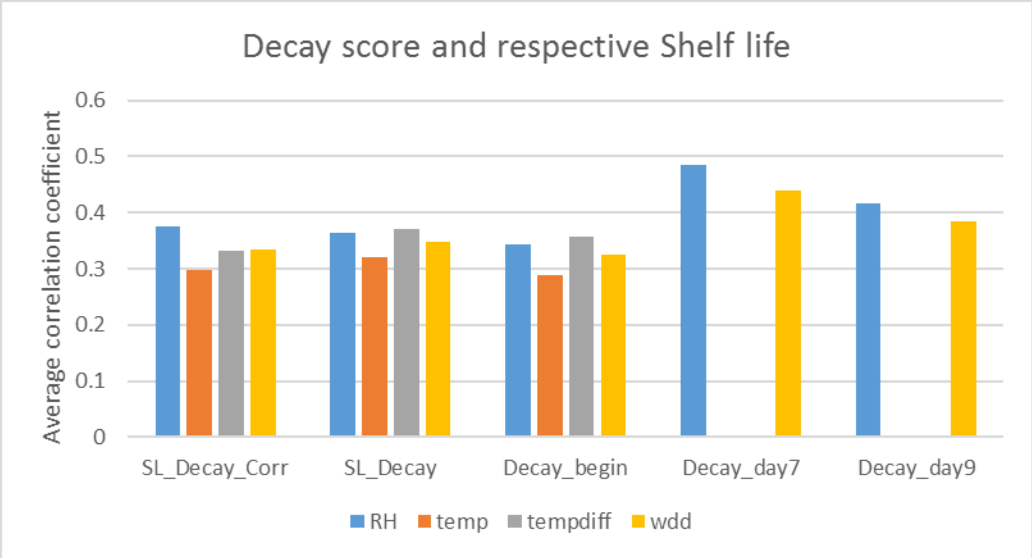


Figure 10 Average correlation coefficient for the quality variable decay score and respective shelf life. Decay = Score of decay; Begin = score the day after arrival at WFBR
Batches of 2016 were measure at day 7 and those of 2017 at day 9.
SL = Shelf life (amount of days until the acceptance limit for the decay score is reached)
SL_Corr = Shelf life corrected for the number of days between harvesting and arrival at WFBR

The shelf life (both corrected and uncorrected) and the decay score at the beginning of the storage period seem to be related to all four pre-harvest parameters, although the RH and tempdiff showed an higher average correlation coefficient than the other parameters. For the decay score after 7 and 9 days only the RH and wdd are relevant; the temperatue and tempdiff do not correlate with these quality variables. Hence the RH is in general the most relevant pre-harvest parameter for the decay. This is in agreement with the trends found in the analysis of the Driscoll's data presented in section 3.2. The Vscore day 8 parameter, also reflecing the condition of the fruit after storage showed higher correlations with RH and WDD than with temperature (see figure 5).

Next to the decay score, the colour and firmness of the WFBR batches was also measured. The firmness was only measured in 2017. There is one statistical signifficat correlation found between the firness and the growth conditions: this is the average correlation coefficient (r=0.42) between the temperature recods and the firmness measured on day 9 (end of the shelf life). This correlation coefficient is in the same range as the correlation coefficients between growth conditions and the decay score/decay shelf life values.

The colour is quantified in °Hue. The higher the Hue value the more red the strawberry. The colour was measured at the begining and during the storage period. The graph in figure 11 shows the average correlation coefecient between the colour and the growth condition parameters.

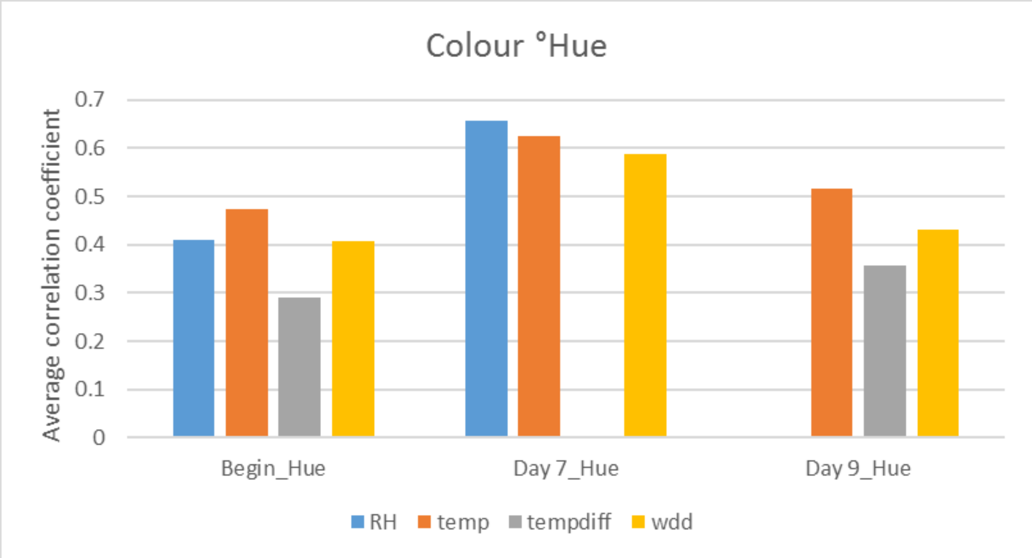


Figure 11 Average correlation coefecient between the colour (°Hue) and four pre-harvest parameters (growth conditions).

Likewise the decay results, the colour measurement in the begining of the storage period seem to be correlated to all 4 growth conditions (in less extend to the tempdiff). Also an increase of the average correlation coefficient relatively to the start measurement was observed. It should be noticed that the measurements of day 7 belong to the batches of 2016 and day 9 to the batches of 2017. The temperature and wdd is for all measurement moments relevant.

An overview of the correlation coefficients between the non-dynamic pre/post harvest parameters as growing system, water source, week production, etc is presented in annex 6. High correlation coefficients were found between the decay/calyx score and the week number (r=0.82 to 0.93). This indicates that the strawberry quality decreases over the time. Since this parameter is subjective (visual evaluation by product experts), this very strong relation may be (partly) influenced by a shift in the visual judgement during the season.

The grower variable is correlated to several quality variables. Since the type of growing system is directly connected to a latitude and a specif grower, these results should be considered carefully. It was not possible to test the same growing system at different latitues and growers, thus there is a direct dependence between these parameters. The parameters AvgTimePickingCooling (=average time between piking and cooling), week production (amount in kg produced strawberries in the monitoring field during the respective week), cumulative week production (sum of all week productions untill the respective week), week production corrected and cumulative week production corrected (week production and cumulative week production corrected for the amount of strawberries that the plant has produced but did not meet the commercial requirements) are independent of the grower/growing system.

The firmness is negative correlated with both the cumulative week production and the cumulative week production corrected. This means that the more strawberries are produced in a production field the softer the strawberries are. Based on these results it can also be concluded that the cumulative week production is a valuable parameter to predict strawberry quality and a good addition to the week production. The firmness is also correlated to the cultivar. This may be explained by specific cultivar featuress (some cultivars are firmer than others).

In general there are more statistical siginificant correlations between the firmness and the non-dynamic pre/postharvest paramerets than between the colour and the non-dynamic pre/postharvest paramerets.

Next to the correlation coefficients discusssed above, the direct relation between decay and firmness for these WFBR batches was investigated. Figure 12 shows this relation for the Lusa monitoring batches from 2017. As it can be seen there is a trend: the lower the firmness the higher the decay score however the correlation is low (and even lower for Scarlet). Regarding the relation between decay and colour (°Hue) the same low correlations were found. These high variance in the results may be explained by the large differences between harvesting and arrival time at WFBR of the different batches. These period ranged between 1 day and 5 days.

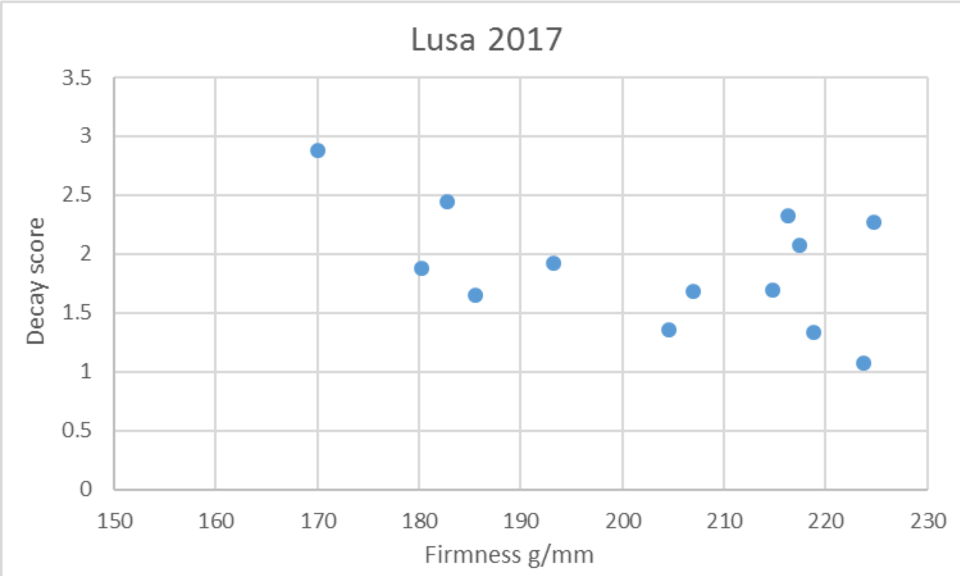


Figure 12 Scatter plot of the decay score versus the firmness for the WFBR batches (monitoring Lusa 2017).

4 Conclusions

In general the correlation coefficients found between individual pre/post-harvest parameters and quality variables were low. This indicates that the quality of strawberries depends on more than one parameter and hence the quality prediction requires a model approach where several parameters can be combined to predict the quality of strawberries.

It should be stressed that the conclusion are based on the current research scope, focussed on one cultivar, 4 regions in Europe, several growers and throughout the whole year. This scope led to a large number of records which also increased the noise (due to larger mistakes chance) and variance in the data (several cultivers, growers, etc). This may be another reason why the correlation coefficients found were low. The conclusions are not based on an analysis of variance (ANOVA), therefore the conclusions give indications and reflect trends. Moreover, only the statistical significant correlation coefficients were considered in the data analysis. The correlations indicate a possible causal relationship. Whether the link is indeed causal will have to be proven in practice.

Driscoll's data set
The correlation coefficients calculated out of the data gathered during the monitoring for the cultivar Lusa showed the following:

- The quality variables were mainly related to different growth condition parameters:
  - For brix: temperature during the growth period is more relevant than the RH;
  - For Vscore day 8 (shelf life variable): the RH is more correlated to this quality parameter than the temperature.
- The calculated parameter Tempdiff (temperature difference between day and night) is rather valuable and more relevant to monitor than the temperature only.
- Combining the temperature and RH in the water vapour deficit (wdd) is a relevant parameter to monitor. For some quality variables (T2 condition) the wdd is even slightly more related to the T2 condition than the temperature and the RH separately.
- Regarding the effect of the growing degree hours period (GDH) defined in this research it can be concluded that in general the correlation coefficient increased with increasing GDH period. This means that monitoring the pre-harvest parameters for a longer period before picking improves the average correlation coefficient. This effect is for temperature and temperature difference larger than for the RH and wdd.
- Monitoring the temperature and temperature difference through the whole production period, from planting to picking moment gives a better correlation with quality than monitoring only the period close to picking.
- Monitoring the RH from 7500 – 10000 GDH until the harvest correlates the best with the product quality.
- The 5000 GDH is a particular relevant moment for both the temperature difference and for the water vapour deficit monitoring.

Bakker Barendrecht data set
The data set of Bakker Barendrecht was limited in terms of number of records and thus a small amount of statitiscal signifficant correlation coeffcients was obtained for result analysis. However the correlation coefficents found support the conclusions based on the Driscoll's data set.

Batches measured at WFBR
Regarding the extra Driscoll's batches measured at WFBR it can be concluded that the firmness is negative correlated with both the cumulative week production and the cumulative week production corrected. This means that the more strawberries are produced in a production field the softer the strawberries are. Based on these results it can also be concluded that the cumulative week production is a more valuable parameter to predict strawberry quality than the week production only. In general there are more statistical siginificant correlations between the firmness and the non-dynamic pre/postharvest paramerets than between the colour and the non-dynamic pre/postharvest paramerets. This indicates that the firmness is more related to shelf life and quality than colour.

The shelf life (both corrected and uncorrected) and the decay score at the beginning of the storage period at WFBR are related to all four pre-harvest parameters (RH, tempdiff, wdd and temperature). Hence the RH is in general the most relevant pre-harvest parameter for the decay. This is in agreement with the trends found in the analysis of the Driscoll's data.

Further research
Due to several pratical limitations it was not possible to monitor all pre-harvest parameters that may play a role in quality. A next step in this work could be to monitor the amount of water and nutrients given to the plants during growth and fruit developement. Also the occurence of extreme weather conditions or plant diseases was not recorded and those may give extra information on quality.

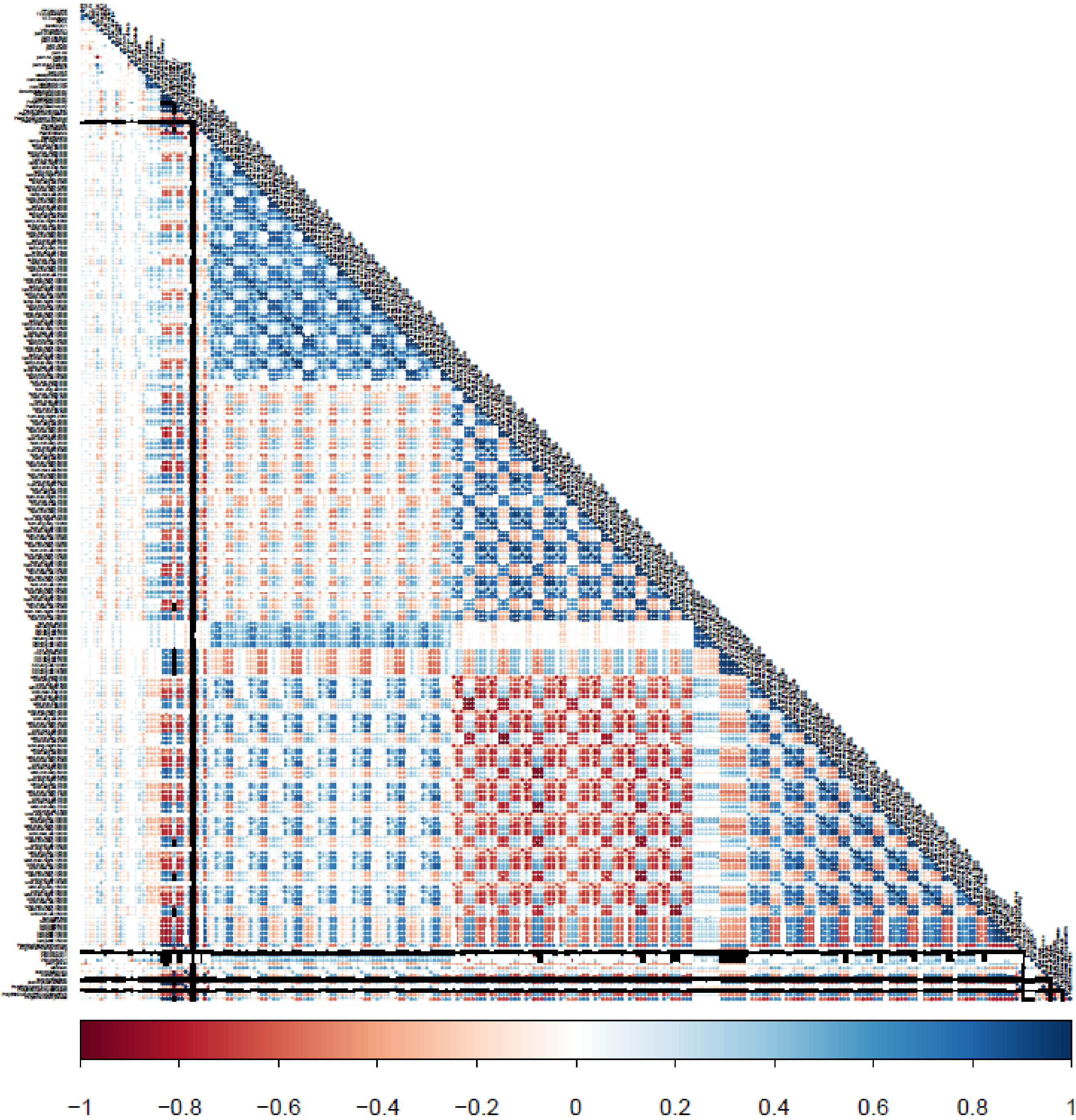
A large data set is important to obtain statistical signifficant correlation coefficients, but it may also introduce extra variance and noise in the data leading to lower correlation coeficients. Therefore increasing the amount of records would be relevant to confirm the relationships found but the quality of the data in terms of accuracy is of high importance. Also the lack of data must be avoided as much as possible. The parameters light intensity and carbon dioxode were less reliable in this work due to several missing values.

The research scope of the non-dynamic variables should be enlarged to obtain a balanced variables matrix, where the parameters are indepedent of each other. In this data there was a strong interdependency between grower, region and growing system which makes the conclusions from this specific parameters weak.

Acknowledgements

This work was funded by Foundation TKI Horticulture, Driscoll's Europe and Bakker Barendrecht. Furthermore we would like to thank Driscoll's Europe and Bakker Barendrecht for their collaboration in acquiring the material and growth information during this project and thank in particular Rob van Mullem and Erna van der Wal for their help in organizing the many shipments of strawberries and gathering the accompanying growth parameters.

Annex 1 Total correlation matrix for Lusa, data 2016, 2017 and 2018

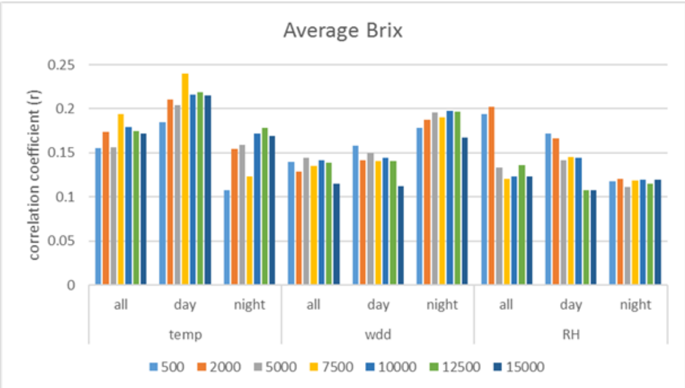


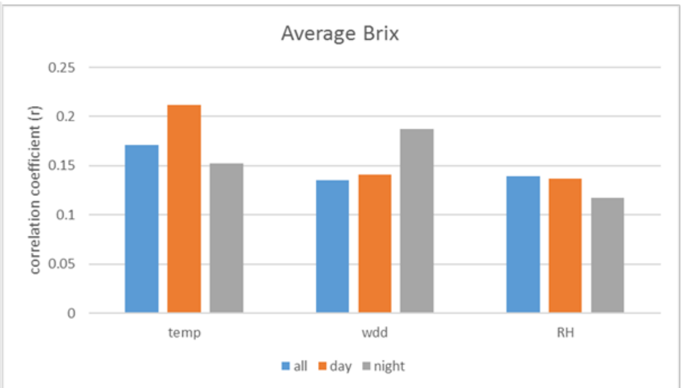
The colour intensity indicates how strong is the correlation coefficient and if the parameters are positively correlated (red) or negatively correlated (blue).

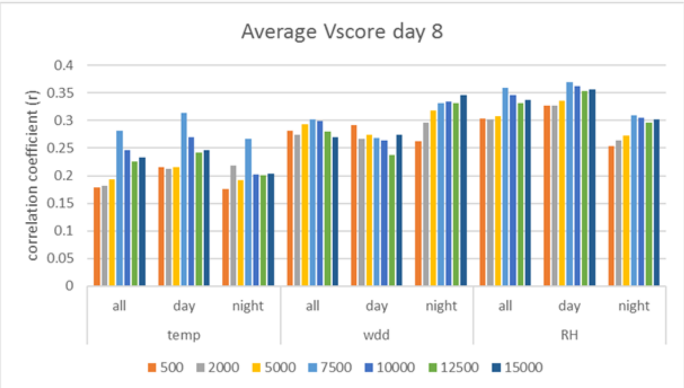
Overview of total amount Driscoll's batches per cultivar en per monitoring year:

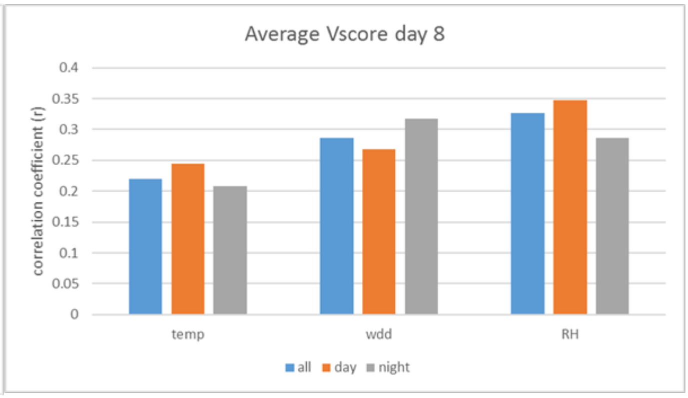
	2016	2017	2018
Lusa	827	882	364
Scarlet	629	438	0

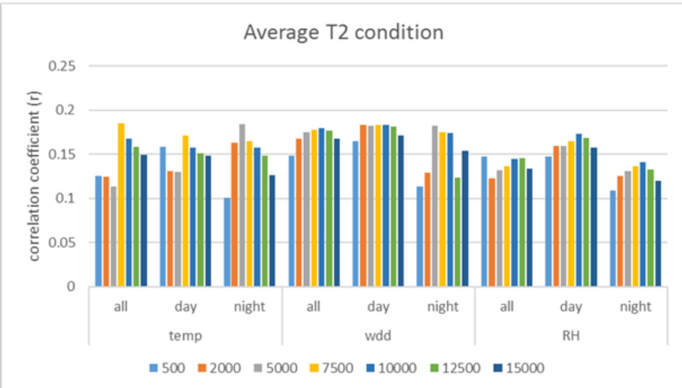
Annex 2 Average correlation coefficient between pre/postharvest parameters (minimal, maximal and average values) and quality variables for Lusa (calculated from the monitoring data)

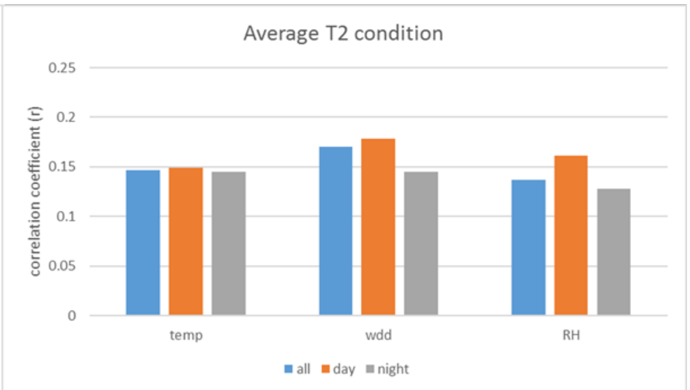










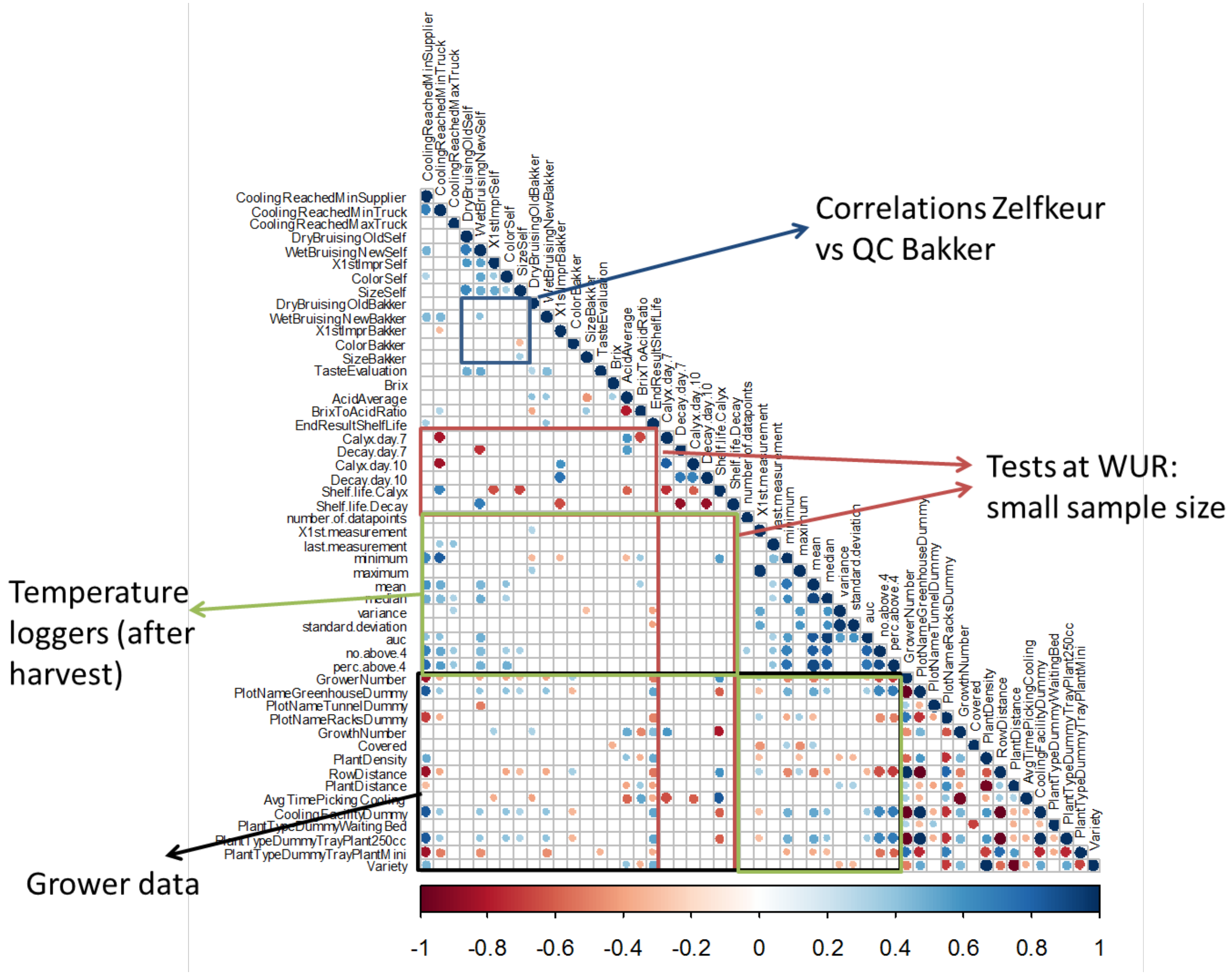


Annex 3 Correlation coefficients between non-dynamic parameters (as growing system, planting date, week production, etc.) and quality: Lusa data 2016, 2017 and 2018

Pre/postharvest parameter	PFQSCORE	T2.Appear	T2.Conditi	BRIX	SevScQC1	Vscore day 8	% not.fully.colore	%cracked	%overripe	%dry bruis	% wet bru	% rot	%no.defect	%no to light def.	% med.def	% wet.fr	%rot.fr	*Total number*
Growing Medium	0.12	-0.07	0.25	0.15	-0.15	-0.46	0.33		-0.14	0.41	-0.30	-0.21				-0.15	-0.38	*13*
Soil Type		-0.11	0.15	0.14	-0.09	-0.30	0.32		-0.13	0.44	-0.29	-0.17				-0.13	-0.32	*12*
Water Source	0.20	-0.06	0.36	0.09	0.10	-0.48	0.16	-0.08		0.05	-0.13	-0.24		0.15			-0.22	*13*
Drainage	0.12	-0.08	0.25	0.15	-0.16	-0.48	0.34		-0.14	0.41	-0.29	-0.21				-0.16	-0.37	*13*
PolyEth cover Type	0.22	0.13	0.23		0.26	-0.32	-0.18			-0.58	0.14		-0.25	0.52				*10*
PolyEth UV180micro	-0.21	-0.07	-0.27	-0.09	-0.10	0.44	-0.09		0.07	0.06		0.10	0.16	-0.29			0.28	*13*
PolyEth diffuse65																		*0*
PolyEth PECopolinerEVA		0.12	-0.06	-0.09	0.26	0.20	-0.27	0.05	0.09	-0.45	0.30	0.14		0.29		0.15	0.22	*14*
PolyEth Glass	0.12	-0.07	0.25	0.15	-0.15	-0.46	0.33		-0.14	0.41	-0.30	-0.21				-0.15	-0.38	*13*
PlantingDateDayNumber	0.13	-0.07	0.26	0.14	-0.13	-0.47	0.33		-0.14	0.39	-0.29	-0.22				-0.15	-0.38	*13*
Planting Date	0.12	-0.07	0.25	0.15	-0.15	-0.46	0.33		-0.14	0.41	-0.30	-0.21				-0.15	-0.38	*13*
Plant Type Tray	0.12	-0.07	0.25	0.15	-0.15	-0.46	0.33		-0.14	0.41	-0.30	-0.21				-0.15	-0.38	*13*
PlantType Misted	-0.21	-0.07	-0.27	-0.09	-0.10	0.44	-0.09		0.07	0.06		0.10	0.16	-0.29			0.28	*13*
PlantType Trimmed		0.12	-0.06	-0.09	0.26	0.20	-0.27	0.05	0.09	-0.45	0.30	0.14		0.29		0.15	0.22	*14*
PlantType Frigo																		*0*
Plant Density	0.10	-0.08	0.23	0.12	-0.17	-0.41	0.33		-0.11	0.37	-0.24	-0.17				-0.17	-0.21	*13*
Row Distance	-0.13		-0.18		0.16	0.30	-0.23	0.07		-0.13		0.05		0.14	0.19	0.16		*11*
Plant Distance		0.11	-0.10	-0.14	0.25	0.28	-0.30		0.11	-0.47	0.31	0.16		0.24		0.16	0.28	*13*
Morocco	-0.21	-0.07	-0.27	-0.09	-0.10	0.44	-0.09		0.07	0.06		0.10	0.16	-0.29			0.28	*13*
Spain		0.12	-0.06	-0.09	0.26	0.20	-0.27	0.05	0.09	-0.45	0.30	0.14		0.29		0.15	0.22	*14*
Belgium																		*0*
Netherlands	0.12	-0.07	0.25	0.15	-0.15	-0.46	0.33		-0.14	0.41	-0.30	-0.21				-0.15	-0.38	*13*
weekproduction	0.10		0.13	-0.22	0.24			-0.06			-0.08	-0.08	-0.19			0.13	0.31	*10*
cumweekproduction	0.08		0.11	-0.08	0.15	0.13		-0.05		0.10	-0.10	-0.10		-0.12		0.11	0.18	*12*
weekprod.corr	0.09		0.11	-0.25	0.28			-0.06			-0.06	-0.07	-0.18			0.13	0.32	*10*
cumweekprod.corr	0.07		0.10	-0.09	0.20	0.14		-0.05		0.10	-0.09	-0.10		-0.12		0.11	0.18	*12*
AvgTimePickingCooling	-0.13	-0.13	-0.08	-0.11		0.36	-0.06	-0.09	0.13				0.20	-0.27			0.36	*11*
week number picking	0.08		0.09	0.14		-0.29	0.16	-0.05		0.23	-0.07	-0.12			0.14		-0.21	*11*
Total number	*20*	*19*	*25*	*23*	*23*	*23*	*21*	*11*	*17*	*22*	*20*	*23*	*7*	*13*	*2*	*18*	*23*	
Maximal value	*0.22*	*0.13*	*0.36*	*0.15*	*0.28*	*0.44*	*0.34*	*0.07*	*0.13*	*0.44*	*0.31*	*0.16*	*0.20*	*0.52*	*0.19*	*0.16*	*0.36*	
Minimal value	*-0.21*	*-0.13*	*-0.27*	*-0.25*	*-0.17*	*-0.48*	*-0.30*	*-0.09*	*-0.14*	*-0.58*	*-0.30*	*-0.24*	*-0.25*	*-0.29*	*0.14*	*-0.17*	*-0.38*	
Average	*0.14*	*0.09*	*0.18*	*0.13*	*0.17*	*0.36*	*0.24*	*0.04*	*0.11*	*0.31*	*0.21*	*0.15*	*0.19*	*0.25*	*0.16*	*0.14*	*0.29*	

Annex 4 Overview correlations Bakker Barendrecht (all batches)

Correlation matrix two growers:



Correlation matrix including the dynamic parameters (data of one grower):

Temperature	DryBruisingOldSelf	WetBruisingNewSelf	X1stImprSelf	ColorSelf	SizeSelf	DryBruisingOldBB	WetBruisingNewBB	X1stImprBB	ColorBB	SizeBB	TasteEvaluation	Brix	Acid	BrixToAcidRatio	EndResult SL	Calyx.day.7	Decay.day.7	Calyx.day.10	Decay.day.10
temp.avg.night.500																			
temp.avg.day.500												0.54		0.41					
temp.avg.all.500												0.39							
temp.max.night.500																			
temp.max.day.500												0.40		0.48					
temp.max.all.500												0.40		0.48					
temp.min.night.500																			
temp.min.day.500																			
temp.min.all.500																			
temp.avg.night.2000																			
temp.avg.day.2000					0.39							0.45							
temp.avg.all.2000																			
temp.max.night.2000																			
temp.max.day.2000																			
temp.max.all.2000																			
temp.min.night.2000																			
temp.min.day.2000																			
temp.min.all.2000																			
temp.avg.night.5000																			
temp.avg.day.5000																			
temp.avg.all.5000																			
temp.max.night.5000												0.39							
temp.max.day.5000																			
temp.max.all.5000																			
temp.min.night.5000							0.52												
temp.min.day.5000							0.49												
temp.min.all.5000							0.52										average total:		0.45
RH																			
hum.avg.night.500	0.53																		
hum.avg.day.500					0.40							0.52		0.45					
hum.avg.all.500			0.40		0.51									0.49					
hum.max.day.500	0.52																		
hum.min.night.500																			
hum.min.day.500									0.38					0.45					
hum.min.all.500									0.38					0.45					
hum.avg.night.2000																			0.72
hum.avg.day.2000					0.41							0.45							
hum.avg.all.2000					0.51									0.39					
hum.min.night.2000																		0.75	
hum.min.day.2000																			
hum.min.all.2000																			
hum.avg.night.5000															0.50				
hum.avg.day.5000									0.41										0.75
hum.avg.all.5000		0.51				0.40													
hum.min.night.5000																			
hum.min.day.5000									0.39										
hum.min.all.5000									0.39								average total:		0.48

Annex 5 Correlation coefficient between the quality measurements carried out at Bakker Barendrecht and at WFBR

Quality	DryBruisingOldSelf	WetBruisingNewSelf	X1stImprSelf	ColorSelf	SizeSelf	DryBruisingOld_Bakker	WetBruisingNew_Bakker	1stImprB	Color_B	TasteEvaluation	Brix	AcidAverage	BrixToAcidRatio	EndResultShelfLife
Calyx.day.7	_	_	_	_	_	_	_	_	_	_	_	0.59	-0.67	_
Decay.day.7	_	-0.74	_	_	_	_	_	_	_	_	_	0.56	_	_
Calyx.day.10	_	_	_	_	_	_	_	0.61	_	_	_	_	_	_
Decay.day.10	_	_	_	_	_	_	_	0.74	_	_	_	_	_	_
Shelf.life.Calyx	_	_	-0.65	_	-0.67	_	_	_	_	_	_	-0.59	_	_
Shelf.life.Decay	_	0.70	_	_	_	_	_	-0.65	_	_	_	_	_	_

Annex 6 Overview correlation coefficients of the monitor batches measured at WFBR (non-dynamic parameters as growing system, etc.)

Pre-harvest parameters	Calyx_begin	Decay_begin	Calyx.day.7	Decay.day.7	SL_Calyx	SL_Decay	SL_Calyx_Corr	SL_Decay_Corr	Calyx_day5	Decay_day 5	Calyx_day9	Decay_d9	Rbegin_Hue_avg	Rday 7_Hue_avg	Rday 9_Hue_avg	Firmness_begin	Firmness_day 5	Firmness_day 9
PlantingDateDayNumber		0.28	-0.69	-0.49				0.27								-0.46	-0.46	-0.36
Planting Date	-0.37		-0.61		0.36		0.38	0.30		-0.48				0.48				
AvgTimePickingCooling				0.70		-0.27							0.27					
weekproduction																		
cumweekproduction							-0.27									-0.37	-0.46	
weekprod.corr		0.30																
cumweekprod.corr																-0.39	-0.48	
Growing Medium		-0.34				0.55		0.55		-0.59		-0.51						
Soil Type	-0.45	-0.67			0.35	0.82		0.70		-0.85								
Drainage		-0.36				0.54		0.54		-0.59		-0.51						
PolyEth cover Type																		
PlantType Tray				-0.63		0.48		0.40								-0.43	-0.46	
PlantType Misted				0.63														
PlantType Trimmed	0.31	0.49			-0.28	-0.59		-0.47	0.42	0.76								
PlantType Frigo																		
Plant Density										-0.38			-0.33	-0.48				
Row Distance				0.56														
Plant Distance		0.36				-0.42		-0.30		0.63								-0.35
Water Source				-0.62		0.35		0.37				-0.42			-0.36			
Latitude	-0.34	-0.49		-0.65	0.29	0.63		0.51	-0.43	-0.77								
Grower	-0.39	-0.41		-0.62	0.33	0.68		0.57	-0.43	-0.76								
cultivar		0.46	-0.78													-0.48	-0.49	-0.43
daynum		-0.54	0.76										-0.29	-0.63			0.36	0.36
weekno	0.93	0.86	0.82	0.86														

Correlation coefficients between de quality measurements carried out at WFBR and pre/post harvest parameters. Only the correlations that are statistically significant are presented in the table; therefore empty cells mean that the correlation coefficient was not statistical signifficant and thus not considered.

Calyx = Score of calyx
Decay = Score of decay
Begin = score the day after arrival at WFBR
Batches of 2016 were measure at day 7 and those of 2017 at day 5 and 9.
SL = Shelf life (amount of days until the acceptance limit for the calyx and decay score is reached)
SL_Corr = Shelf life corrected for the number of days between harvesting and arrival at WFBR (tis number of days was not constant for all batces therefore a correction has been applied)
Hue = °Hue represents the colour
AvgTimePickingCooling =average time between piking and cooling

Week production= amount in kg produced strawberries in the monitoring field during the respective week
Cumulative week production= sum of all week productions untill the respective week
Week production corrected and cumulative week production corrected = week production and cumulative week production corrected for the amount of strawberries that the plant has produced but did not meet the commercial requirements

Annex 7 Correlation matrix between cooling temperature parameters and the quality parameters of 25 Driscoll's batches

SMALL TABLE	PFQSCORE	T2.Appear	Green.uni	T2.Condit	Wet.bruis	Dry.bruise	BRIXLEVEL	Severity.s	Number.c	Vscore.da	perc.no.d	perc.ntl	perc.med	perc.wet.f	perc.rot.fr	X1st.meas	last.meas	minimum	maximum	mean	median	variance	standard.d	auc	no.abov.4
PFQSCORE	1.00	0.61	-0.37	0.83	-0.38	-0.04	0.32	-0.61	0.15	-0.55	0.36	0.43	-0.53	-0.20	-0.43	-0.57	-0.11	0.51	-0.67	0.22	0.13	-0.67	-0.69	-0.45	-0.02
T2.Appearance	0.61	1.00	-0.59	0.07	0.20	0.22	0.22	-0.44	0.13	-0.51	0.11	0.58	-0.46	0.35	-0.67	-0.40	-0.23	0.50	-0.41	0.12	0.09	-0.30	-0.36	-0.36	-0.19
Green.unripe	-0.37	-0.59	1.00	-0.05	-0.03	-0.04	0.08	0.26	-0.18	0.29	-0.15	-0.18	0.14	0.03	0.28	0.16	0.05	-0.52	0.18	-0.14	-0.15	0.11	0.16	0.09	0.31
T2.Condition	0.83	0.07	-0.05	1.00	-0.62	-0.20	0.29	-0.46	0.10	-0.34	0.38	0.13	-0.34	-0.48	-0.08	-0.44	0.02	0.30	-0.56	0.19	0.10	-0.64	-0.62	-0.31	0.10
Wet.bruising	-0.38	0.20	-0.03	-0.62	1.00	0.06	-0.26	0.04	-0.24	-0.05	-0.33	0.56	-0.14	0.28	-0.27	0.42	-0.45	-0.42	0.43	-0.26	-0.18	0.55	0.50	0.10	0.08
Dry.bruises	-0.04	0.22	-0.04	-0.20	0.06	1.00	0.37	-0.08	0.49	-0.42	-0.01	0.22	-0.22	0.34	-0.17	-0.05	0.56	0.41	-0.06	0.56	0.57	-0.02	0.01	0.33	0.05
BRIXLEVELRESULTS	0.32	0.22	0.08	0.29	-0.26	0.37	1.00	0.17	0.20	-0.26	-0.24	0.43	-0.09	0.58	0.15	-0.22	-0.02	0.30	-0.24	-0.05	-0.18	-0.33	-0.35	-0.37	-0.16
Severity.score.QC1	-0.61	-0.44	0.26	-0.46	0.04	-0.08	0.17	1.00	-0.21	0.51	-0.84	-0.31	0.86	0.11	0.57	0.13	-0.13	-0.46	0.30	-0.48	-0.45	0.48	0.49	0.01	-0.46
Number.of.fruits.QC1	0.15	0.13	-0.18	0.10	-0.24	0.49	0.20	-0.21	1.00	-0.20	0.12	0.04	-0.01	-0.28	-0.21	-0.33	0.34	0.61	-0.37	0.37	0.36	-0.34	-0.38	-0.03	-0.26
Vscore.day.8.QC1	-0.55	-0.51	0.29	-0.34	-0.05	-0.42	-0.26	0.51	-0.20	1.00	-0.25	-0.53	0.59	-0.12	0.45	0.39	-0.13	-0.45	0.50	-0.39	-0.37	0.43	0.48	0.16	0.08
perc.no.defects	0.36	0.11	-0.15	0.38	-0.33	-0.01	-0.24	-0.84	0.12	-0.25	1.00	-0.19	-0.70	-0.05	-0.17	-0.04	0.39	0.43	-0.19	0.48	0.43	-0.47	-0.43	0.16	0.50
perc.ntl	0.43	0.58	-0.18	0.13	0.56	0.22	0.43	-0.31	0.04	-0.53	-0.19	1.00	-0.51	0.00	-0.42	-0.17	-0.48	0.03	-0.25	-0.10	-0.09	-0.03	-0.09	-0.34	-0.07
perc.med.defects	-0.53	-0.46	0.14	-0.34	-0.14	-0.22	-0.09	0.86	-0.01	0.59	-0.70	-0.51	1.00	-0.18	0.38	0.21	-0.04	-0.33	0.35	-0.34	-0.29	0.41	0.42	0.13	-0.39
perc.wet.fr	-0.20	0.35	0.03	-0.48	0.28	0.34	0.58	0.11	-0.28	-0.12	-0.05	0.00	-0.18	1.00	-0.21	-0.26	0.14	0.06	-0.13	0.06	0.04	-0.04	-0.04	-0.12	-0.02
perc.rot.fr	-0.43	-0.67	0.28	-0.08	-0.27	-0.17	0.15	0.57	-0.21	0.45	-0.17	-0.42	0.38	-0.21	1.00	0.22	0.07	-0.33	0.25	-0.38	-0.36	0.25	0.30	0.13	-0.18
X1st.measurement	-0.57	-0.40	0.16	-0.44	0.42	-0.05	-0.22	0.13	-0.33	0.39	-0.04	-0.17	0.21	-0.26	0.22	1.00	0.17	-0.40	0.97	0.07	0.21	0.79	0.85	0.76	0.48
last.measurement	-0.11	-0.23	0.05	0.02	-0.45	0.56	-0.02	-0.13	0.34	-0.13	0.39	-0.48	-0.04	0.14	0.07	0.17	1.00	0.50	0.15	0.80	0.79	-0.09	0.01	0.63	0.36
minimum	0.51	0.50	-0.52	0.30	-0.42	0.41	0.30	-0.46	0.61	-0.45	0.43	0.03	-0.33	0.06	-0.33	-0.40	0.50	1.00	-0.43	0.67	0.61	-0.53	-0.54	0.00	-0.16
maximum	-0.67	-0.41	0.18	-0.56	0.43	-0.06	-0.24	0.30	-0.37	0.50	-0.19	-0.25	0.35	-0.13	0.25	0.97	0.15	-0.43	1.00	0.02	0.17	0.88	0.91	0.76	0.40
mean	0.22	0.12	-0.14	0.19	-0.26	0.56	-0.05	-0.48	0.37	-0.39	0.48	-0.10	-0.34	0.06	-0.38	0.07	0.80	0.67	0.02	1.00	0.98	-0.21	-0.15	0.58	0.43
median	0.13	0.09	-0.15	0.10	-0.18	0.57	-0.18	-0.45	0.36	-0.37	0.43	-0.09	-0.29	0.04	-0.36	0.21	0.79	0.61	0.17	0.98	1.00	-0.04	0.02	0.67	0.45
variance	-0.67	-0.30	0.11	-0.64	0.55	-0.02	-0.33	0.48	-0.34	0.43	-0.47	-0.03	0.41	-0.04	0.25	0.79	-0.09	-0.53	0.88	-0.21	-0.04	1.00	0.98	0.57	0.15
standard.deviation	-0.69	-0.36	0.16	-0.62	0.50	0.01	-0.35	0.49	-0.38	0.48	-0.43	-0.09	0.42	-0.04	0.30	0.85	0.01	-0.54	0.91	-0.15	0.02	0.98	1.00	0.62	0.23
auc	-0.45	-0.36	0.09	-0.31	0.10	0.33	-0.37	0.01	-0.03	0.16	0.16	-0.34	0.13	-0.12	0.13	0.76	0.63	0.00	0.76	0.58	0.67	0.57	0.62	1.00	0.55
no.abov.4	-0.02	-0.19	0.31	0.10	0.08	0.05	-0.16	-0.46	-0.26	0.08	0.50	-0.07	-0.39	-0.02	-0.18	0.48	0.36	-0.16	0.40	0.43	0.45	0.15	0.23	0.55	1.00

To explore the potential of nature to improve the quality of life

Wageningen Food & Biobased Research
Bornse Weilanden 9
6708 WG Wageningen
The Netherlands
www.wur.eu/wfbr
E info.wfbr@wur.nl

Report 1940

The mission of Wageningen University and Research is "To explore the potential of nature to improve the quality of life". Under the banner Wageningen University & Research, Wageningen University and the specialised research institutes of the Wageningen Research Foundation have joined forces in contributing to finding solutions to important questions in the domain of healthy food and living environment. With its roughly 30 branches, 5,000 employees and 10,000 students, Wageningen University & Research is one of the leading organisations in its domain. The unique Wageningen approach lies in its integrated approach to issues and the collaboration between different disciplines.