Review

Advances in non-destructive early assessment of fruit ripeness towards defining optimal time of harvest and yield prediction – a review

Bo Li¹, Julien Lecourt¹, Gerard Bishop¹

¹ NIAB EMR, East malling, Kent, ME19 6BJ, UK
* Correspondence: Gerard.bishop@emr.ac.uk

Abstract: Global food security for the increasing world population not only requires increased sustainable production of food but a significant reduction in pre- and post-harvest waste. The timing of when a fruit is harvested is critical for reducing waste along the supply chain and increasing fruit quality for consumers. The early in field assessment of fruit ripeness and prediction of the harvest date and yield by non-destructive technologies have the potential to revolutionize farming practices and enable the consumer to eat the tastiest and freshest fruit possible. A variety of non-destructive techniques have been applied to estimate the ripeness or maturity but not all of them are applicable for in situ (field or glasshouse) assessment. This review focuses on the non-destructive methods which are promising, or have already been, applied to the pre-harvest in field measurement including colorimetry, visible imaging, spectroscopy and spectroscopic imaging. Machine learning and regression models used in assessing ripeness are also discussed.

Keywords: Pre-harvest; ripeness; image analysis; machine learning; fruit phenotyping,

1. Introduction

There are many ways in which the status of global food security can be improved for the world’s increasing population. Increased fruit production through adding to the area cropped is not sustainable and thus productivity per unit land area must be increased. Simultaneously there is a need to prevent waste and for fruit production the timing of harvest is crucial to ensure production meets the commercial ripeness specifications. Over- or under-ripe fruits have a lower or even no retail value and represent a significant income loss and a waste of resources. For the consumer too early harvest reduces the taste and quality of fruits whilst a late harvest can lead to reduced shelf life, poor appearance, and “off” flavours and odours. The early in-field assessment of fruit ripeness and the prediction of both harvest date and yield will therefore greatly reduce the waste in the supply chain and thus help towards improving food security.

The non-destructive on-plant assessment of fruit ripeness has received increasing interest as it provides several advantages compared with traditional destructive methods, such as high-throughput assessment, simultaneous multiple measurements and real time decision making. The phenotypic changes during fruit ripening are complex and in most cases a green hard immature fruit becomes more colourful, softer, sweeter and aromatic. Numerous physical and chemical attributes that can quantified during ripening include size, shape, texture, firmness, external colour, internal colour, concentration of chlorophyll, soluble solids content (SSC), starch, sugars, acids, oils, and internal ethylene concentration [1]. It is not realistic to assess simultaneously all the quality attributes in the field with non-destructive methods and destructive laboratory measurements are time-consuming due to the large number of samples required to take account of the within-field variability [2]. Simple representative non-destructive measurements are thus required to assess the ripeness of a fruit.
The measurement of fruit maturity in a non-destructive manner dates back more than half a century ago with the development of light transmittance techniques [3,4]. Since then a variety of non-destructive techniques have been introduced including colorimetry [5], visible imaging [6], visible and near infrared (VNIR) spectroscopy [7], hyperspectral imaging [8], multispectral imaging [9], fluorescence imaging [10], acoustic impulse technique [11], Computed Tomography (CT) scan [12], Magnetic resonance imaging (MRI) [13], acoustical vibration technique [14] and electronic nose technique [15]. The first six techniques listed above will be considered in this review as they are the most likely to be used in portable devices to enable pre-harvest in field assessment of fruit ripeness. These techniques have been applied to studies of a large number fruits and external quality attributes have been measured and correlated with internal characteristics, as shown in Table 1. The non-destructive methods can produce a large amount of data with multiple variables and thus multivariate analysis is utilized to identify key discriminatory variables that correlate with the ripening status of a fruit [16]. Such key discriminatory variables can be used in regression models enabling an assessment of fruit quality and thus ripeness.

Currently, the harvest time is mainly estimated by counting of days after flowering, subjective tasting or visual assessment of fruit colour, texture or plant canopy structure [17,18]. All of these methods on their own or in combination are time consuming and not necessarily accurate. The quality attributes or the maturity indices derived from non-destructive techniques can however be modelled to predict the optimal time of harvest. Such predictions need to account for changes in environmental conditions and being a primary factor affecting the rate of plant development [19], air temperature is a key parameter utilized in models predicting optimal harvest time [20]. An example of a typical approach in developing a workflow for predicting optimal harvest date is given in Figure 1.

Figure 1. A scheme of the overall workflow for the prediction of the optimal harvest date.

Here we review the non-destructive techniques for the fruit ripeness assessment and the modelling for the prediction of an optimal harvest time. These predictions are based on imaging and/or spectroscopic techniques that quantify the colour and/or spectral qualities of fruits that change due to their molecular composition during the ripening process.
Table 1. Overview of the non-destructive methods for the assessment of fruit ripening.
Abbreviation: SSC (Soluble solid content), DM (Dry matter), MC (Moisture content), TTA (Titratable acidity) and TSS (Total soluble solid)

<table>
<thead>
<tr>
<th>Fruit</th>
<th>Colorimetry</th>
<th>Visible imaging</th>
<th>Spectroscopy</th>
<th>Fluorescence</th>
<th>HSI</th>
<th>MSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pear</td>
<td></td>
<td></td>
<td>Firmness [34], SSC [35]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peach</td>
<td>Colour [37]</td>
<td></td>
<td>Firmness [38], Chlorophyll [39], Colour [37]</td>
<td>Firmness [40]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nectarine</td>
<td>Colour [44]</td>
<td></td>
<td>SSC [45], firmness [45]</td>
<td>Firmness [40]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mango</td>
<td>Colour [46]</td>
<td></td>
<td>DM [47], starch [48], SSC [47], colour [49], firmness [50]</td>
<td>Firmness [8], SSC [8], water content [8]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banana</td>
<td>Colour [51]</td>
<td>Colour [51]</td>
<td>TSS [52], Chlorophyll [53]</td>
<td>Firmness [54], TSS [54]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tomato</td>
<td>Colour [55], Firmness [56], TSS [57]</td>
<td>Colour [58], firmness [59]</td>
<td>Lycopene [60], SSC [61]</td>
<td>a°/b° [62]</td>
<td>Phenolic [63], lycopene [63]</td>
<td></td>
</tr>
<tr>
<td>Melon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mandarin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cherry</td>
<td>Colour [69]</td>
<td></td>
<td>Firmness [70], SSC [71]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strawberry</td>
<td>Colour [72], TSS [73], Firmness [72], TTA [73]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. Colour measurement

The colour and appearance of a fruit is the initial quality assessment consumers use to judge the acceptability of a fruit. These criteria are related to physical and chemical changes occurring during fruit ripening [16,94]. In many fruits, during ripening colour change occurs due to chlorophyll degradation and the increase in the concentration of pigments such as carotenoids or polyphenols [95]. Several fruits have been studied for the relationship between maturity and colour including tomato [96], orange [95], guava [97], peach [98], nectarine [99], mango [46,100], blueberry [101], cherry [101] and pineapple [89]. To measure the changes in fruit colour the use of colorimeters and image capture and analysis are the two major methods.

2.1. Colorimeter

Colorimeters are traditional non-destructive instruments used extensively in the fruit industry to measure fruit colour [102]. They are more precise than human visual assessment and standardization using CIELAB colour space, which was introduced by Commission Internationale de l’Eclairage (CIE) in 1976, provides unified measurements [103]. The three coordinates of CIELAB colour space, L*, a* and b*, represent the respective values of lightness and the green to red and blue to yellow ratios. CIELAB is close to human perception of colour due to the uniform distribution of colours and all the colours can be perceived by human eye can be located on the three coordinates [104]. Several colour indexes have been developed as the indicator of ripeness. In certain experiments only the a* parameter was correlated with colour change [97,103,105]. The b* value has only been reported to be positively correlated with ripeness in peach [37]. Using more than one of the colour components improves the assessment of ripeness. The ripeness of tomato has been assessed using the ratio between a* and b* and it showed high positive correlation with lycopene concentration [106–108] and identifies significant differences for the six USDA ripening classes [56,102]. Equations based on a* and b* include determining the hue angle and chroma. The hue angle was found to be one of the best parameters to discriminate different ripeness stages for tomato, peach and guava [37,97,102,106,108]. L* value was also incorporated to the colour models with a* and b* including citrus [109] and tomato [103]. Portable colourimeters are now available

commercially [46] and can be carried to the field. However, single fruit measurement limits its application to map the ripeness of fruits in the whole field.

Data obtained with colourimeters has been successfully correlated with fruit ripeness thanks to multivariate analyses. These statistical tests allow the simultaneous model with multiple variables for example multiple linear regression (MLR) has been used to predict accurately the maturity of mango using \(a^*b^*\) and their product as the variables [46]. Other regression methods were also tested in this study including Partial Least Square (PLS) regression and Principal Component Regression (PCR), but the prediction performance were slightly worse than MLR in the study by Jha et al. [46].

2.2. Visible imaging

Colorimeters are not able to obtain representative colour values due to the limited sampling area compared to the size of the fruit [110]. This limitation can be overcome by 2D colour imaging that converts photons reflected from fruit skin to electrical signals and received by camera with CCD (Charge-Coupled Device) or CMOS (Complementary Metal Oxide Semiconductor) sensors. Normally, the sensor receives the light and filters it to three channels, which are R (red), G (green) and B (blue) and the intensity values are always determined by fruit samples, illumination and the internal characteristics of the camera [111].

Similar to \(L^*a^*b^*\) colour space it is possible to analyze fruit ripeness in RGB colour space. Schouten et al. showed that the R component can be used to describe the progressive colour change of tomato at different stages of ripeness and correlates with changes in fruit firmness [59]. As ripening is a continuously changing process the exact colour boundaries between different ripeness stages are difficult to determine and if used arbitrary thresholds for each colour channel need to be provided. Fuzzy logic, a statistical analysis approach reviewed by Yuan et al. [112], can overcome the need for discrete thresholds and has been applied to the ripeness assessment of mango and apple [22,113]. Goel et al. used the difference between R and B values to enhance the classification of the different tomato ripeness stages reaching 94.3% accuracy [114].

Other statistical methods utilized includes unsupervised classification such as K-means and Gustafson-Kessel algorithms, as reviewed in the study of Hartigan et al. and Lesot et al. respectively [115,116]. These have successfully been applied to automatically separate banana of different ripeness stages based on their RGB values [22]. Rather than only using the average RGB values, the histogram of each channel was used to match with predefined reference histograms for each ripeness group [22].

RGB values from images can be transformed to \(L^*a^*b^*\) values and a comparable performance was obtained with a colorimeter for the internal quality assessment of tomato fruits including Brix and lycopene content [117]. Similar conclusion was drawn from studies in tomato [118], cherry [101] and banana [94]. The RGB values are, however, device dependent and not a perceptually uniform space, calibration is therefore crucial before the transformation of RGB values into \(L^*a^*b^*\) space and produce parameters comparable with a colorimeter [104]. Statistical modelling approaches including quadratic and neural network models, as described in [119,120], were the best models to convert RGB values into \(L^*a^*b^*\) space [111] and Taghadomi et al. adopted the neural network method and obtained a very strong correlation (\(R^2=0.99\) ) between actual \(L^*a^*b^*\) and RGB values in cherry [101].

Other colour space values such as HSI (for Hue, Saturation and Intensity) and HSV (Hue, Saturation and Value), can be derived from RGB values, can better represent human visual perception [89]. Hue is defined as the similarity to the defined colours as red, green, blue and yellow [121] and Saturation is used to describe how colourful a stimulus is relative to its own brightness [122]. Fuzzy logic was successfully applied to group pineapple into three ripeness stages using values derived from H, S and I [89]. Ukirade et al. used H, S and V values as the input of the neural network model to classify tomato into four ripeness groups [123]. EI-Bendary et al. proposed a more sophisticated method for tomato, which used the colour histogram in HSV space and the colour moments (mean, standard deviation and skewness) which measure the colour distribution in
an image as the colour features [124]. Principal Component Analysis (PCA) was applied to extract the features for both Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) models with more than 80% Correct Classification Rate (CCR) for five ripening stages [55]. Rather than only using one colour space, colour components from two colour spaces can be combined such as the study by Li et al, which used R, B and H from outdoor colour images of blueberry and it was found K-Nearest Neighbour (KNN), which classifies an object based on a majority vote among its neighbours [125] could achieve more than 85% CCR to separate four ripening stages [126].

Compared with a colourimeter, colour information can be obtained rapidly from larger area in a 2D image due to the high spatial resolution. The equipment can also be easily attached to moving platform such as tractor, robot or drone for rapid collection of multiple data collection measurements. There are however challenges that need to be overcome for 2D imaging including the difficulty of segmentation of fruits from background, device dependent RGB values and requirement of homogeneous illumination in the field.

3. Visible and infrared spectroscopy

When light hits the surface of a fruit it can be absorbed, scattered or re-emitted. The amount of each of these is determined by the physical properties and chemical constituents and thus ripeness of a fruit. Visible and Near InfraRed (VNIR) reflectance spectroscopy measures reflected light between 380 nm and 2500 nm, which is largely dependent on the light absorption by fruit sample and relates to almost all the major organic compounds. An example of the changes in spectra during ripening is given in Figure 2 that shows marked changes in the values between 400-700nm and also notable changes can be observed at longer wavelengths. VNIR spectroscopy has been widely applied as a non-destructive and fast measurement method for multiple quality attributes [127]. More importantly, a portable device has been developed and used in the field [128]. The recorded spectra can be analyzed and related to different ripeness stages by using spectral indices. The whole wavelength scan or values at key selected wavelengths are used in regression models to correlate with specific fruit qualities that is associated with fruit ripeness.

![Figure 2. Typical progressive change of reflectance spectra at different ripening stage of tomato and the legend showed corresponding colours](image)

3.1. Spectral indices

Spectral indices normally combine the surface reflectance at two or more wavelengths to indicate relative abundance of a feature of interest as it is difficult to use only one wavelength as
indices for in-field assessment because the values can be highly affected by the sensor, environmental illumination and particle size [129]. A number of spectral indices have been calculated to describe the progressive change of peel pigment concentration during ripening. Ruiz-Altisent et al. found that the reflectance at 450nm and 680nm were both associated with peach firmness, but when using the reflectance values from only two wavelengths the linear correlation factor was low ($R^2<0.6$) [130]. The index of absorption ($I_{AD}$) introduced by Ziosi et al., is a robust spectral index obtained in calculating the difference between the absorption at two wavelengths around the chlorophyll-a peak (670nm and 720nm) [131]. $I_{AD}$ range was found to be similar across different growing seasons and a good correlation with ripeness was found for peach [98,132], apricot [133], nectarine [134,135] and plum [92]. Merzlyak et al. identified several spectral indices for the quantification internal quality attributes of apple related to its ripeness with high correlation ($R^2>0.8$) [24] Zude et al. also developed three indices including the ratio between the transmission at 698nm and 760nm, NDVI and red-edge vegetation stress index (RVSI). Even though all these indices showed good correlations with colour change of apple peel, the correlation appears to be cultivar dependent as lower for the cultivar of ‘Jonagold’ than “Elstar” [23]. The same phenomenon was also noticed in the study of Shinya et al. that the correlation between firmness and $I_{AD}$ could be extremely different among three cultivars of peach [98].

3.2. Full or selected wavelengths

Spectral indices can describe the change of peel pigment concentration during ripening process and provide comparable values with the colorimetric method [37], but peel colour is not always the only criteria for ripening assessment. The correlation between internal quality attributes related to ripening such as firmness and SSC were investigated with full or selected wavelengths from VNIR spectra.

For the full wavelengths, PLS is the most used regression model to predict fruit quality. The prediction is achieved by extracting a set of orthogonal factors from the predictors called latent variables which have the best predictive power [136]. Another common regression model is the Principal Component Regression (PCR), which uses Multiple Linear Regression (MLR) to correlate with the principal components scores extracted from the predictors [137], has been applied in some studies for fruit quality assessment [138–140]. Compared with PLS, PCR showed the drawback that the principal components were obtained without considering the dependent variables.

The variability of physical sample properties and/or the performance of the hardware can result in undesired results including light scattering, path length variations and random noise generated in the extracted spectra. These factors will reduce the accuracy and robustness of the prediction models [141]. In order to improve the data analysis, a number of studies have applied different pre-processing techniques to the spectra obtained before modelling [142].

Savitzky-Golay (SG) is the most frequently used digital data smoothing filter [42,49,65,91,143,144], which applies linear least squares method to fit low-degree polynomial [145]. However, SG have contrasted effects on the performance of multivariate statistical models [146]. For example, Jha et al. compared different pre-processing techniques and found that smoothing did not show any improvement in comparison with other techniques for the assessment of the firmness in mango [49]. But Herrera et al. showed that SG filters using a second-order polynomial performed better than other scattering correction methods for the prediction of wine grape brix [143]. Standard Normal Variate (SNV) [72,78,147,148] and Multiple Scattering Correction (MSC) [42,67,140,143,144,149] are the two most frequently used techniques for scattering correction. MSC is used to eliminate the nonlinear scattering due to the non-uniform travel distance of light by linearizing each spectrum to a reference spectrum, which is the always the mean spectrum [138]. Previous research has shown the similarity between SNV and MSC, for example Ma et al. confirming the correlation coefficients were the same when assessing the sugar content of peach using PLS models with SNV and MSC [150]. SNV can however be applied to individual spectrum without requiring a reference [151]. In some studies, SNV was performed with de-trending, which was used to correct the baseline shift of spectra [72,72,148].
Generating derivatives of spectra are useful pre-processing techniques to enhance subtle differences and reduce the effect of specular reflection [79,143]. Guo et al. found that the PLS regression model performed better with first derivative of the spectra than SNV, MSC, and the second derivative for predicting the Soluble Solid Content (SSC) in strawberry [144]. A similar conclusion was drawn for the Total Soluble Solid (TSS) content prediction of strawberry [73]. Second derivative has also been used in the prediction of chlorophyll content of apple [23], the SSC of kiwifruit, strawberry, cherry, and peach [71,79,152,153], and the firmness of apricot [78]. Carlini et al. compared the second derivative, MSC and SNV methods and found the second derivative showed the best performance for the prediction of SSC of cherry [71]. Interestingly, pre-processing techniques are not always beneficial to the spectral analysis. Clement et al. applied all mentioned pre-processing techniques to the prediction of tomato ripeness, but it was found none of them showed improvement on a PLS model due to the low levels of noise [154]. Likewise, Jaiswal et al. also found that the best predictions of TSS and DM (Dry Matter) content of banana with PLS model was obtained with no pre-processing to the spectra [52].

In some studies, only a small number of selected wavelengths were used to reduce the multicollinearity among variables and modelled by Multiple Linear Regression (MLR) model. The key wavelengths can be identified manually or automatically. The manual selection of key wavelengths has been used for the prediction of brix of mango [155] and SSC for grape, lime and star fruit [84]. Guthrie et al. calculated the correlation coefficients between second derivative of spectra and brix values as the criteria of wavelength selection for MLR model. This method provided better predictions than when using the first derivative [47] and had previously been used in the prediction peach brix levels [156].

Automated wavelength selection method such as stepwise wavelength selection has also been used to aid the predictive power of models. The first wavelength is selected with the highest correlation to the dependent variable, and for each step the candidate wavelength was added to increase the correlation until none of the remaining wavelengths were significant. This method was used for the SSC prediction of peach [157], melon and pineapple [158]. Genetic Algorithm (GA), which uses natural selection and random mutations based on prediction accuracy is another efficient automated method in identifying key wavelengths and was successfully applied to SSC prediction of apple [159].

A comparison of the performance of PLS and MLR for the assessment of maturity of mango indicated that when using MLR, poorer correlation of data was observed and led to overfitting as seen by the large gap between the correlation coefficients of calibration and validation models [160]. Similar phenomenon was also observed for the prediction of TSS and DM for banana [52]. However, if the key wavelengths were selected with most of the variance of the whole spectra but low collinearity, MLR can show better performance than PLS such as the firmness prediction of mango [50]. These two methods of modelling were also compared for the prediction of brix value of mango and interestingly both correlation coefficients were high and comparable [155]. Consequently, it is unclear which model can provide better prediction as the performance of MLR is largely dependent on the wavelength selection.

The correlation was always higher for the prediction of SSC than firmness of fruit by using spectroscopic methods. Park et al. showed that the prediction of firmness was more complicated than SSC as it was not determined by a single analyte or a limited group of related chemicals [139]. For both SSC and firmness prediction, the performance of the model is always cultivar dependent, and the calibration model trained by using mixed cultivars produces lower correlation than when using results from an individual cultivar. Limited number of studies were focused on the assessment in field, but compared with indoor measurement, the prediction is less accurate [27,45,48].

Similarly with colourimeter, spectroscopy is not likely to be applied as high-throughput ripening assessment due to the low spatial resolution. The accuracy of the internal quality measurement is influenced by sample temperature, which needs to be compensated by an extra calibration model [48]. Spectroscopy has been used in assessing the quality of a large variety of
fruits and portable commercial spectrometers have been developed [48,78,128,161] but most of the studies focused on indoor post-harvest assessment of fruit maturity. Inconsistent performances were observed for the models developed by spectra taken indoor and on-tree. Predicting apple firmness and SSC both on the tree and during storage showed that ‘on tree’ PLS model had the best correlation coefficients for both firmness and SSC [27,162]. However, for nectarine, the ‘on tree’ model performed worse than post-harvest [45]. Consequently, for the on-tree ripeness assessment, it is necessary to build the prediction model with spectra taken in-field and understand the effect of environmental factors on the quality of spectra.

4. Fluorescence

Fruit degreening, i.e., the loss of chlorophyll, is an effective indicator of fruit ripening and thus measuring chlorophyll content using a fluorimetric sensor can be correlated to fruit ripeness traits [163].

One chlorophyll fluorimetric method measures the photochemical and non-photochemical processes with the illumination of actinic light [164]. A Pulse-Amplitude-Modulation (PAM) based fluorometer has been developed commercially, which uses visible light at blue region as the excitation and measures minimum ($F_0$) and maximum ($F_m$) emitted fluorescence. The maximum quantum yield ($F_m - F_0$)/$F_m$ was calculated and this parameter was found to be negatively correlated with the ripening stage of apple [165] and papaya fruit [166]. This chlorophyll fluorimetric method is popular in the laboratory, but difficult to be applied in field as the samples need to be dark-adapted. For example, the papaya samples in the study of Urbano et al. were dark-adapted for 30 mins with a dark towel [166]. Bodria et al. designed a fluorescence imaging system which measured the light emission at 690 - 740 nm with the excitation light in UV-blue and red regions, and good correlation was achieved between the quality parameters including firmness and SSC of fresh apple and detected fluorescence even though hue value of skin colour showed little change, but for peach and nectarine, the correlation was lower [163]. The fruit samples measured by this fluorescence imaging system were not dark adapted, but the equipment has only been designed for the laboratory use.

In order to reduce the influence of the environmental factors on the absolute fluorescence intensity at single band, which limits in-field application [167], more studies were focused on the understanding of the fluorescence ratios using various light sources of defined wavelengths, which has led to the development of a handheld multi-parametric fluorescence sensor – Multiplex® (Force-A, Orsay, France) with four LED light sources and three synchronized fluorescence detectors [10]. The most common informative indices utilized fluorescence from anthocyanins (ANTH), flavonols (FLAV), and chlorophyll (CHL) to indicate fruit ripeness. In the study of Betemps et al., CHL showed positive correlation with the firmness of apple and they also obtained good negative correlation between FR_RED and SSC [29]. In field assessment of CHL was also successfully applied to grape with a high correlation with TSS and the combination of CHL and ANTH can be used as a robust decision tool to predict harvest time [86]. For tomato, all the indices were found to correlate well with the time-shift in the tomato ripening process in the study of Hoffmann et al. [62]. The blue to red fluorescence ratio (B_UV/RF_UV) was measured as an effective parameter for the assessment of the ripeness of oil palm with rough skin, and with the Classification and Regression Tree (C&RT) method resulting in an overall correct classification rate of 90% for three different ripeness stages [168].

5. Spectral imaging

5.1. Hyperspectral imaging (HSI)

Hyperspectral imaging (HSI) has emerged as a powerful tool for the inspection of fruit quality. HSI generates a three-dimension imaging cube with images at a range of continuous wavelengths. A single spectrum can be extracted from each individual pixel representing the absorption properties and the textural information of fruit samples [169]. Similar with traditional visible
imaging and spectroscopic methods, HSI is non-destructive and requires little sample preparation, but it is advantageous that it can record both spatial and spectral information simultaneously [170]. For the assessment of fruit quality, two types of wavelength dispersion devices are normally used i.e. line scanning and area scanning coupled with an imaging sensor for the HSI image acquisition. A line scanning device has the imaging spectrograph dispersing the incident light into different wavelengths instantaneously between visible and near infrared wavelength range (380-1700 nm). Line scanning HSI cameras scan the samples continuously in one direction, so they can be attached to moving platforms such as tractors [171], robots [172] and unmanned aerial vehicles (UAV) [173].

The hyperspectral image can be handled in two different ways 1) light scattering analysis and 2) spectral analysis. Modified Lorentzian Distribution (MLD), that correlates the data obtained with a predefined distribution curve by using a distribution function, can be used to describe the scattering profile and the fitting parameters were used as the variables of stepwise MLR model [30]. The results of this study suggest that spectral scattering from all wavelengths or selected wavelengths can provide more accurate prediction of apple firmness than using the secondary properties such as spectral absorption [30]. Similar methods were also employed for the prediction of peach firmness, but with MLR models different results were obtained when using two different cultivars [41]. Mendoza et al. combined both the spectral and image analysis techniques on scattering images including discrete and continuous wavelet transformation decomposition, first order statistics, Fourier analysis, co-occurrence matrix, and Variogram analysis but little improvement on the prediction of firmness and SSC of apple was found, with the performance of PLS model being cultivar dependent [174]. Wang et al. used two different feature selections methods which were Uninformative Variable Elimination (UVE) [175] and Supervised Affinity Propagation (SAP) [176]. The output of two PLS models with two feature selection methods were combined as the input into an ANN model that gave a correlation coefficient of 0.83 [177]. The scattering profile was also used for the prediction of firmness of peach [41].

The average spectra of the region of interest (ROI) have also been modelled for the assessment of fruit quality. The average spectra from whole wavelength scan (400-1000 nm) were used with a PLS model to predict the SSC of grape and the correlation coefficients were similar for both white and red grape [88]. The same method has also been used to predict the TSS of strawberry [178] and the firmness and SSC of blueberry [179]. Key wavelengths have also been selected before modelling in order to reduce the redundancy of the whole spectral dataset with different feature selection techniques. One of widely used feature selection criteria is based on beta coefficients derived from PLS model, which measured how great effect an independent variable have on the dependent variable. A comparison of the performance between the MLR models with selected wavelengths based on beta coefficients and PLS model with full spectra as input, showed that the final outcomes were similar for the internal quality measurement of strawberry [178]. The same feature selection methods were also used by Rajkumar et al. to predict the firmness and TSS of banana by MLR model, and good correlation was achieved for both quality attributes with correlation coefficients of 0.91 and 0.85 [54]. Another key wavelength selection method to solve collinearity problem is the Successive Projection Algorithm (SPA), which iteratively adds one wavelength until a specific number of wavelength was achieved with minimum redundant information content [180], has been used to select the feature wavelengths as the input of PLS model and high correlation ($R^2 = 0.92$) was found for predicting persimmon firmness [83].

Overall HSI is a promising technique for fruit ripeness assessment. In-field application of this technique will need to overcome the challenges of handling the large data output and the calibration of variable light levels whilst in the crop.

5.2. Multispectral imaging (MSI)

Multispectral imaging (MSI) is a form of HSI that collects data at specific wavelengths instead of scanning the whole wavelength range. This can be accomplished using a frame scanning imaging system with Liquid Crystal Tunable Filter LCTF coupled with CCD (charge-coupled device) or CMOS (complementary metal oxide semiconductor) sensor. Lu et al. used five wavelengths based...
on previous studies to correlate their scattering profiles with the firmness and SSC of apple using an ANN model and obtained reasonably high correlation for both quality attributes, which were $r^2 =$ 0.87 and 0.77 respectively [181]. Another lower-cost MSI system uses a rotating filter wheel containing a few bandpass filters instead of LCTF, but the tuning speed is lower than LCTF. This device has been used to predict the firmness and SSC of peach with the best combination of four wavelengths and high correlation coefficients were achieved as 0.94 and 0.97 respectively [76]. The prediction of firmness was higher than the prediction by HSI [41]. Similarly, Liu et al. used MSI with 19 wavelengths to predict the firmness and TSS of strawberry, and PLS, SVM and ANN were compared with the best correlation coefficient of 0.94 and 0.83 respectively [74], which was comparable with HSI for prediction of TSS of strawberry [178].

Compared with HSI, MSI can be lower cost and easier to convert into portable device and the output imaging dataset will be smaller. A portable MSI device with four narrow-band light sources and four reflectance sensors of different wavelengths at 570, 670, 750 and 870nm has been developed [130]. This device was used to classify oil palm into different stages of ripeness using quadratic discriminant analysis and discriminant analysis with Mahalanobis distance classifiers with a >85% of correct classification rate being achieved [182].

6. Prediction of optimal harvest date

Non-destructive methods are very promising for the ripeness assessment in field, but the most critical question is to how to link such assessment to predict yield and the optimal harvest date and [18]. This is highly challenging and made complicated by ripeness variability within and between plants.

Yang et al. recorded the HSI spectra of tomatoes at different growing stages and PLS model was applied to predict the growing stage with the best correlation coefficient of 0.89. It was also found that the key wavelengths were in the visible and infrared regions (400-2100 nm) [183]. A similar method was employed to predict the number of days before commercial harvest of apple. The calibration model was built with eight cultivars and good correlation ($R^2=0.93$) was found for the spectral range between 380 nm and 2000 nm [184]. Environmental factors such as temperature, light levels, humidity, etc. significantly influence the development of crop fruit, and it is essential to incorporate important environmental factors predictions for the determination of the optimal harvest date [185]. Several crop models have been developed since the 1960s by Loomis et al. [186] with the input of environmental factors both for in-field and greenhouse prediction. Such models are difficult to use due to the number of input variables [187–192]. Qiu et al. investigated the dominant environmental factors in greenhouse for tomato growth and it was found that temperature, humidity and photosynthesis active radiation (PAR) show positive or negative correlation to crop growth [192]. The influence of temperature has also been reported in a number of studies such as for tomato [185], grape [193], apple [194], mango [195], blueberry [18], apricot [196] and strawberry [197]. Shewfelt et al. studied the colour change at different constant temperatures, but the variation of the temperature within each day was not considered [96]. Munoz et al. developed a time series regression model for the prediction of harvest date of blueberry, and the minimum and maximum of daily temperature from weather forecast from two weeks ahead were used as the input to the model [18]. This method was closer to real application and potentially could be applied with non-destructive techniques, which could determine the current ripening stage. Environmental temperature affected the fruit growth, and it was also found the fruit temperature also influenced the near infrared reflectance spectrum in a non-linear way [146]. Kawano et al. compensated the surface temperature effect by developing combined MLR model which covered a variation of temperature ranged between 21 to 31 °C [198]. Peirs et al. compared global calibration model that covers a wide temperature range and calibration models for each temperature range and both methods performed well for the SSC prediction of apple, but for the practical purpose, global calibration model was preferred [26].
7. Conclusions

Although now becoming slighted outdated the key study by the Food and Agricultural Organisation (FAO) indicated that global food waste and losses in 2009 was estimated to be one-third (by weight) of all food produced in the world [199]. The prevention of such losses and waste is therefore a major driver to improve global food security. Here we have reviewed a range of non-destructive techniques and the data modelling methods for the assessment of fruit ripening and the prediction of optimal harvest date. Knowing when to pick will not only depend on the fruit to ensure the optimal taste, quality and postharvest performance but also on the local circumstances including, weather, the supply chain and markets. Having an affordable portable device to inform this decision is crucial and the non-destructive techniques discussed above have all been developed into such devices that will greatly help to reduce waste.

Such devices include colourimeters that can record accurate colour information but only for individual fruit. 2D imaging overcomes this limit as it images larger area and can be fitted on moving platform, but 2D imaging can only obtain colour information which is not adequate as the indicator of ripeness for all fruits. Similar with colourimeter and 2D imaging, fluorescence detect the colour change especially the change of content of chlorophyll. Spectroscopy in visible and NIR region correlates with both colour and internal quality attributes, which can provide better prediction of fruit ripeness. Hyperspectral imaging has the advantage of both spectroscopy and 2D imaging and can be integrated with moving platform, but due to the high expense and large dataset generation, it is still mainly used in the lab. Multispectral imaging has the potential to overcome the limitation of hyperspectral imaging but the reliability of the measurement in field needs to be further investigated. The miniaturization and computational capacity will be a major technical hurdle in ensuring that the devices can be used in the field and provide real time assessment of fruit quality and ripeness. Simultaneous to the development of the hardware the data modelling techniques need to be implemented that can utilize not only data from the controlled laboratory conditions but that from the more challenging field conditions, with the changeable environment.

Author Contributions: Dr. Bo Li is the key author, who reviewed previous literature and wrote the manuscript; Dr. Julien Lecourt supervised the work, provided ideas and contribute to writing; Prof. Gerard Bishop proposed the topic, supervised the work, contribute to writing and edit the manuscript before submission.

Conflicts of Interest: The authors declare no conflict of interest.

References

30. Peng, Y.; Lu, R. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit...


112. Klir, G.; Yuan, B. Fuzzy sets and fuzzy logic; Prentice hall New Jersey, 1995; Vol. 4;.


