Use of multispectral airborne imagery to improve yield sampling in viticulture

E. Carrillo, A. Matese, J. Rousseau, B. Tisseyre

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Use of multi-spectral airborne imagery to improve yield sampling in viticulture

E. CARRILLO[1], A. MATESE[2], J. ROUSSEAU[3] and B. TISSEYRE[1*]

Abstract

The wine industry needs to know the yield of each vine field precisely to optimize quality management and limit the costs of harvest operations. Yield estimation is usually based on random vine sampling. The resulting estimations are often not precise enough because of the high variability within vineyard fields. The aim of the work was to study the relevance of using NDVI-based sampling strategies to improve estimation of mean field yield. The study was conducted in 9 non-irrigated vine fields located in southern France. For each field, NDVI was derived from multi-spectral airborne images. The variables which define the yield: (berry weight at harvest (BWh), bunch number per vine (BuN) and berry number per bunch (BN)) were measured on a regular grid. This data-base allowed for five different sampling schemes to be tested. These sampling methods were mainly based on a stratification of NDVI values, they differed in the way as to whether NDVI was used as ancillary information to design a sampling strategy for BuN, BN, BW or for all yield variables together.

Results showed a significant linear relationship between NDVI and BW, indicating the interest of using NDVI information to optimize sampling for this parameter. However this result is mitigated by the low incidence of BW in the yield variance (4 %) within the field. Other yield components, BuN and BN explain a higher percentage of yield variance (60% and 11 % respectively) but did not show any clear relationship with NDVI. A large difference was observed between fields, which justifies testing the optimized sampling methods on all of them and for all yield variables. On average, sampling methods based on NDVI systematically improved vine field yield estimates by at least 5-7 % compared to the random method. Depending on the fields, error improvement ranged from -2 % to 15 %. Based on these results, the practical recommendation is to consider a two-step sampling method where BuN is randomly sampled and BW is sampled according to the NDVI values.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Key words: grape yield sampling, two-step sampling, remote sensing, NDVI

[1]UMR ITAP, Montpellier SupAgro/IRSTEA, 2 Place Pierre Viala. 34060 Montpellier Cedex 2. France

[*]Corresponding author: bruno.tisseyre@supagro.fr

[2]Istituto di Biometeorologia (IBIMET – CNR), Consiglio Nazionale delle Ricerche, via Caproni 8, 50145 Firenze. Italy

[3]Groupe ICV, Institut Coopératif du Vin. La Jasse de Maurin. 34970 Lattes, France.

Introduction

In order to optimize harvest organization and quality management, the wine industry needs to know the yield of each vine field. Ideally, yield has to be estimated a few days before harvest with a relative error of less than 10%. Although models have been developed to forecast the yield at the regional level (Baldwin, 1964; Cristofolini and Gottardini, 2000), their results were not precise enough to manage logistic issues linked to harvest operations at the farm or at the winery level. Therefore, precise estimation of vine field yield always requires fruit sampling and counting (Clingeleffer et al. 2001).

Estimation of yield must be carried out quickly (a few minutes per field) at a time when the workload at harvest or for the preparation of the harvest is important. Practical constraints, like the time available to visit all the fields before harvest, limit the number of sampled sites per field. Therefore, yield estimation is based on a low number of sampling sites (~4-5 sites) where variables which define the yield (number of clusters, number of berries per cluster, mean berry weight) are measured by an operator. These variables will be called hereafter yield components.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Accurate estimation of the yield remains difficult because of different sources of errors and/or variability:

i) Errors due to the operator, mainly counting errors caused by the difficulty in visualizing properly all bunches within the canopy. To overcome this constraint, Wolpert and Vilas (1992) proposed a two-step sampling method. This considers the different yield components on which yield estimation is based, independently. Indeed, depending on the phenological stage of the vine, some of the yield components are much easier to visualize. Thus, Wolpert and Vilas (1992) proposed estimating the average number of clusters early in the season, at flowering, when they are easy to visualize, and the mean bunch weight at the end of the season (just before harvest). It is thus possible to improve yield estimation by minimizing errors on bunch counting without increasing the time required to perform these observations. This approach assumes that the number of flowers does not change over the season and corresponds to the number of bunches at harvest.

ii) the second type of error is caused by the variability at the vine level (within-plant variability); bunch weight presents a high within-plant variability (Clingeleffer et al. 2001). Measuring this yield component is costly and destructive. Classical methods provide an estimation of this parameter using a small number of representative clusters (Clingeleffer et al. 2001). To minimize errors of estimation due to the choice of the clusters, different systematic methods have been proposed in the literature (Wulfsohn et al., 2012, Meyers et al., 2011). Other studies propose the use of alternative sources of information to facilitate or automate BW estimates. This is the case of image analysis that has been proposed to detect, count and estimate the weight of clusters (Diago et al., 2012, Reis et al., 2012, Nuske et al., 2011, Serrano et al., 2005, Dunn and Martin, 2004,) or to estimate the number and the volume of berries (Grocholsky et al., 2011,

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Rabatel and Guizard, 2007). Other authors have proposed continuous weighing devices positioned on the trellising system (Blom and Tarara, 2009) at a specific location in the field. These approaches aim to facilitate the work of observers and possibly reduce the estimation error by increasing *n*, the number of samples measured while maintaining or reducing the time required to make field observations.

iii) errors can be due to the inter-plant or plant-to-plant variability. To take into account this scale of variability, yield sampling methods generally rely on the definition of sampling sites that include a variable number of vines (usually between 3 and 10). Yield components are measured or sampled across all vines corresponding to the sampling site.

iv) finally, errors can also occur due to within-field variability; Taylor et al. (2005) showed that the coefficient of variation (CV) of the yield is very high in viticulture (CV ~ 50 %) and, what is more important, yield variation is not randomly distributed but presents a strong spatial organization (spatial autocorrelation). Surprisingly, none of the sampling methods proposed in the literature take into account the spatial organization of the yield; indeed, most of the yield estimation methods are based on a random selection of the sampling sites (Clingeleffer, 2001, Wolpert and Vilas, 1992).

In precision agriculture, sampling methods defined according to auxiliary data have been successfully used for the calibration of spatial models (Lesch, 2005). However, to the authors' knowledge, such approaches have never been used for estimating field yield and particularly vine field yield. On the other hand, NDVI may be relevant auxiliary information to optimize sampling for yield estimation. Meyers et al. (2011) proposed using NDVI information to optimize sampling to improve the estimation of vine canopy parameters. Many authors have shown that, for vines, the NDVI or a

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

similar vegetation index was correlated with the yield at the within-field level (Rousseau et al., 2008; Bramley et al., 2005; Martínez-Casanovas and Bordes, 2005). Moreover, it has been shown to be appropriate to characterize the spatial variability of vine fields at high resolution and sufficiently in advance (up to 15 days before veraison) to plan sampling before harvest (Kazmierski et al., 2011).

The aim of this work was to study the value of sampling based on NDVI (SBN) to improve estimation of the mean field yield. This study proposes to investigate the interest of optimizing the choice of within-field sampling sites. Considering the two-stage method of Wolpert and Vilas (1992), the study, i) investigated the possible relationship between each yield component and NDVI, and ii) tested, for each stage of the sampling method, the value of a sampling strategy designed according to the spatial distribution of within-field NDVI values.

Materials and Methods

Experimental Site

The experiment was conducted on 9 fields in the research vineyard owned by INRA Pech Rouge (Gruissan, Aude, France) (co-ordinates: E:709800, N:6226840, RGF93 datum, Lambert93). Table 1 presents information on the different fields including field size, training system, age of vines and grape variety. The selected fields are all very representative of Mediterranean vineyards in Southern France. Two different training systems were considered: vertical shoot positioning (VSP) and umbrella. These two training systems are the most common in this part of France. The nine fields were non-irrigated. The Pech Rouge vineyard has a Mediterranean climate with a hot dry summer. Precipitation occurs mainly in autumn and spring. A high evaporative demand usually leads to significant water restrictions in summer. The average water restriction over the

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

vineyard, estimated by pre-dawn leaf water potential measurements (Scholander et al. 1965) was between -0.75 MPa in August 2003 (dry year) and -0.60 MPa in August 2006 (wet year) (Taylor et al., 2010).

Previous work (Acevedo-Opazo et al., 2008; Kazmierski et al., 2011) showed that in this vineyard, water restriction is the main factor affecting the growth and the yield of the vines. The soil variability is the main factor explaining the within-field spatial variability of the vine water status, therefore vigour and estimated vigour through vegetation indices like the NDVI derived from airborne images, were relevant surrogate information to highlight within-field zones of water restriction (Acevedo-Opazo et al., 2008). As a result, NDVI presented a significant temporal stability of the spatial variability at both an intra-annual scale and inter-annual scale (Kazmierski et al., 2011).

As indicated in Table 1, the 9 fields were spread over three pedological units (PU1, PU2 and PU3). Coulouma et al., (2010) showed that each PU presents a significant soil variability which explains a significant variability in vine vigour, yield and vine water status at the within-field level (Taylor et al., 2010; Acevedo-Opazo et al., 2008). These previous works pointed out the opportunity of using NDVI information to design sampling strategies for yield estimation.

<Table 1>

Image Acquisition and Processing

Two multi-spectral airborne images were taken before veraison in 2008 (31st July) and 2009 (1st August). Images of 1m resolution were collected by Avion Jaune (Montpellier, Hérault, France). The spectral regions captured in the images were: i) blue (445-520nm), ii) green (510-600nm), iii) red (632-695nm) and iv) near-infrared (757-

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

853nm). The 1 m square image pixels were aggregated into 3 m square pixels using the methodology outlined in Acevedo-Opazo et al. (2008), which approximates the “mixed pixel” row spacing approach of Lamb et al. (2004). The calculation of NDVI (Rouse et al., 1973) was then made on the 3 m pixels (area of 9 m²). Note that mechanical or chemical weeding was performed over the inter-row spacing; therefore row cover crop did not affect NDVI values.

Sampling Sites

To compare NDVI values with ground measurements, a 15 m common sampling grid was defined. This common sampling grid was implemented field by field. Sampling grid nodes were taken as sample points, and were referred to as measurement sites (hereafter sites). To avoid border effects, on each side of each field, a buffer of 5 m was excluded from the sampling scheme. The resulting sampling rate was averaged over 40 measurement sites per hectare.

Depending on the shape and the area of the field, the number of sites per field was therefore different. The highest number of sites was obtained for the largest field (P22) with 45 sites and the lowest number of sites was obtained for the smallest field (P77) with 19 sites (Table 1). Given the precision level of image geo-referencing (+/- 1 m.), the smoothing introduced with image processing and the spatial footprint of field measurements (see next section), NDVI value was assigned to each site as the mean of 4 pixels corresponding to a square of 36 m².

Field Measurements

Yield components (berry weight at veraison (BWv), berry weight at harvest (BWh), bunch number per vine (BuN), bunch weight (BuW) and berry number per bunch (BN))

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

were measured in 2009. Each site was considered as 5 consecutive vines in the row. BW (at veraison and harvest) was estimated by the average weight of 100 berries randomly taken from the 5 consecutive vines. BuW was estimated at harvest by weighting 10 bunches (2 bunches per vine) also randomly taken from the same 5 consecutive vines. For each site, BN was then calculated as the ratio between BuW and BWh. Finally, BuN was determined by counting all bunches of the 5 consecutive vines of each sampling point. Note that BWv (berry weight at veraison) was measured to test the possibility of estimating mean field yield at an early stage of the growing season (6-7 weeks before harvest), although the study was focused on the yield estimation at harvest. This is why some yield components, such as BN and BuW, were measured or calculated only at harvest. The distance between vines along the row was 1 m. Data were associated with the spatial co-ordinates of the central vine.

The final data base was a set of 313 sites over the 9 different fields. Each site was then characterized by 5 field parameters (BWv, BWh, BN, BuN, and BuW) and 2 remote sensing parameters (NDVI08 and NDVI09), i.e. NDVI values measured in 2008 and 2009, respectively.

Data Analysis

A principal component analysis (PCA) was used to evaluate correlations between each yield component and NDVI values measured either in 2008 or 2009. Data were standardized on a per field basis before PCA was performed.

Variance-Based Sensitivity Analysis

A variance-based sensitivity analysis was carried out to assess the relative importance of the sampled parameters in explaining the variability of the mean field yield. More

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

specifically, the mean field yield (Y) was determined from the yield components (BuN, BN and BWh). This analysis involves the decomposition of variance following the method proposed by Sobol (1993). The variance of the mean field yield estimation (Var(Y)) was decomposed into terms attributable to each yield component as well as each interaction effect between them (Eq. 1).

$$Var(Y) = \sum_{i=1}^{3} V_i + \sum_{i<j}^{3} V_{ij} + \dots + V_{123} \qquad (1)$$

where:

Y: Mean grape yield of the field (kg ha^{-1}).

V_i: Variance of each yield component.

V_{ij}: Variance of the interaction effects between yield components.

The first-order sensitivity index S_i was then used to estimate the relative importance of the yield component "i" in the variability of the mean field yield (Eq. 2). Note that higher order interaction indices S_{ij}, S_{ik}… were also computed.

$$S_i = \frac{V_i}{Var(Y)} \text{ , with } \sum_{i=1}^{3} S_i + \sum_{i<j}^{3} S_{ij} + \dots + S_{123} = 1 \qquad (2)$$

where:

S_i: Main effect index for the i-th component of mean field yield.S_{ij}: Higher order sensitivity index, or effect of the ij-th interaction on mean field yield.

Proposed Sampling Methods

As proposed by Wolpert and Vilas (1992), Eq. 3 presents how the mean field yield (Y) is estimated using the mean field BuN and the mean field BuW. The distinction between these two yield components is justified by practical considerations related to optimal

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

periods of observation to provide relevant measurements. Indeed, BuN is more easily estimated at flowering by counting the number of inflorescences while BuW must be estimated as close as possible to harvest to provide the best yield estimation. Yield component BuW may be estimated directly by sampling bunches or by measuring two additional yield components: the mean field BN and the mean field BW.

$$Y = BuN * BuW \qquad (3)$$

$$\text{with} \quad BuW = BN * BW$$

The two-step approach proposed by Wolpert and Vilas (1992) allowed consideration of different sampling approaches. Depending on how the NDVI variable is used as a surrogate to design a target sampling strategy for BuN or BuW or both, 5 different sampling strategies were proposed and tested: i) random method (RM), ii) random-target method (RTM), iii) target method (TM), iv) random-model method (RMM) and v) model method (MM) (Table 2).

<Table 2>

Each sampling strategy is a combination of different sampling methods applied to yield components. All sampling methods are based on the selection of n sample sites. They differ in the way of selecting these sites. Tests were performed with a number of sites varying from n = 3 to n = 7. This interval was chosen to encompass current practices that correspond to a number of measurement sites equal to 5. Sampling methods are detailed hereafter.

- Random Sampling (RS)

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Random sampling was based on the selection of *n* sites randomly chosen among all the available sites. Yield components BuN and BuW are then computed from observations of the *n* sites to provide an estimate of the mean field yield, following Eq. 3.

- Target Sampling (TS)

Sites are chosen according to the distribution of NDVI values. For the field under consideration, NDVI values are divided into *100/n* % quantiles, with *n* corresponding to the desired number of sites. Among sites corresponding to each quantile, one site is randomly selected. Therefore, TS is a way to stratify the site selection according to NDVI values. Figure 1 illustrates the TS method with *n*=5 sites. This example led to the consideration of 5 intervals on the NDVI values corresponding to quantiles 0-20 %, 20-40%, 40-60%, 60-80%, and 80-100%. For each interval, one site is randomly chosen. Then, an estimation of the mean field yield is computed following Eq. 3.

<Figure 1>

- Model Sampling (MS)

The model sampling was only used to estimate BW (Eq. 3). This approach was defined to take advantage of having NDVI values with a high spatial resolution. The overall idea is to use a regression model (Hengl et al., 2003; Lesch et al., 1995) that provides an estimate of BW at each location where a NDVI value is available. A regression model was then considered between BWh and NDVI08 where BWh is the explanatory variable and NDVI08 is the dependent variable (Eq. 4). Note that both NDVI08 and NDVI09 were taken into account as dependent variables. However, NDVI08 and NDVI09 presented a significant correlation. Therefore, only NDVI08 was considered to present detailed results obtained with MS.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

This procedure was considered as a possibility to better take into account the spatial variability observed over the fields. The regression model was only used to provide an estimate of BWh according to NDVI values (Eq. 4).

$$\widehat{BWh}(s) = a * NDVI(s) + b \qquad (4)$$

The regression model provides estimates of berry weight at harvest ($\widehat{BWh}(s)$) for each site (s) for which a NDVI value (NDVI(s)) is available. a and b are the coefficients to be calibrated for each field.

The model method involved 4 steps: i) selection of the sites to calibrate the model; ii) model calibration; iii) estimation of BWh on each available site; iv) calculation of BW from $\widehat{BWh}(s)$. Each step is detailed hereafter.

i. To select the sites, the target method (TS) was used. As two coefficients (a and b) have to be calibrated for each field, the method could apply to only two sites. For practical reasons, a minimum of three sites were considered in this study.
ii. Classical least squares method was used to identify both parameters (a and b) of the model.
iii. The calibrated model was used to estimate $\widehat{BWh}(s)$ on each within-field location where a NDVI value was available.
iv. BW was calculated as the mean of $\widehat{BWh}(s)$ measurements.

Combining these 3 sampling methods (RS, TS, and MS), 5 different sampling strategies to estimate mean field yield were tested (Table 2).

<u>Evaluation of Sampling Methods</u>

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

The uncertainty of the sampling methods was estimated by bootstrapping (Efron, 1979). The same methodology was applied to each sampling method. Generically, *n* sampling sites were drawn and the estimated mean grape yield ($\widehat{Y_b}$) corresponding to the bootstrap sample b was calculated. This process was repeated B times, which provides B bootstrap samples. Bootstrapping was implemented with B=1000.

The estimated mean field yield was then computed as indicated in Eq. 5.

$$\hat{Y} = \frac{1}{B}\sum_{b=1}^{B}\widehat{Y_b} \qquad (5)$$

The estimated variance of the considered sampling method was defined as indicated in Eq. 6.

$$\widehat{V(Y)} = \frac{1}{B}\sum_{b=1}^{B}(\widehat{Y_b} - \hat{Y})^2 \qquad (6)$$

The error in % (Eq. 7) corresponding to the standard error of the mean was derived from the estimated variance $\widehat{V(Y)}$ and the estimated mean field yield.

$$Error\ (\%) = \frac{\widehat{\sigma(Y)}}{\hat{Y}} x100 \qquad \text{with} \quad \widehat{\sigma(Y)} = \sqrt{\widehat{V(Y)}} \qquad (7)$$

Assuming $\widehat{Y_b}$ is normally distributed, the interval corresponding to +/- Error(%). $\hat{Y}$/100 encompasses 68% of the samples. In order to verify the results obtained from a bootstrapping approach, THEO (%), the relative error computed from a theoretical field with a normal distribution of yield values corresponding to a coefficient of variation (CV) of 65 % was computed (Eq. 8). The value of 65 % was chosen as the mean CV of the yield for the overall fields of the database (Table 3).

$$THEO\ (\%) = \frac{1}{\sqrt{n}}.t.CV \quad (8)$$

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

where :

n : is the number of sites

t : is the value from the Student's table corresponding to the chosen *p-value*,

CV (65%): the coefficient of variation,

THEO (%): the percentage of error (%) defines the relative interval in which the true value may be found with a probability 1-*p*.

Note that THEO (%) was only used to verify the relevance of the results obtained by our bootstrapping approach with the Random Sampling (RS) strategy. Once verified, RS was considered as a reference to compare the different sampling strategies proposed in this paper.

Results and Discussion

Results are reported and discussed in two sub-sections. The first one aims at analyzing the variability of each yield component at the within-field level as well as the relationship between each yield component and NDVI. The second sub-section deals with the results of the sampling methods.

Yield Spatial Variability at the Within-Field Level

Table 3 summarises mean field yields and coefficients of variation (CV) observed for each field of the data-base. Mean field yields are low which is common in non-irrigated conditions on this type of soil under Mediterranean climate with high deficit in water balance.

<Table 3>

A significant heterogeneity in mean yields was observed between fields: the lowest yield was 2.76 t/ha (field p76) while the highest is 7.06 t/ha (field p22). The coefficient of variation (CV) showed a significant within-field variability in almost all the fields

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

(CV values are above 35%), confirming what Taylor et al. (2005) had already observed for a larger data-base obtained with grape harvesters and yield monitoring systems. In the current case, 4 fields (p65, p76, p77 and p80) presented very high CV values, above 70%. This result confirms the significance of the within-field variability for grape yield, and the potential interest of proposing sampling methods adapted to this yield variability.

Relationship Between Yield Components and NDVI

Figure 2 shows that, over the 9 fields of the experiment, the within-field variability may be summarized by two sets of parameters correlated to the first and second factor of the PCA. These factors represent 62 % of the variability of the 9 fields. One group was correlated to Factor 1, including NDVI at both dates (2008 and 2009) and BW at both stages (harvest and veraison). The second group was mainly correlated to Factor 2, which included BuN and BN.

<Figure 2>

Regarding the first set, as it includes both NDVI parameters, it was considered as representative of the vegetative expression. Therefore, as represented on Figure 2 with an arrow, an axis of vegetative expression can be defined. The position of the sites along this axis was in relation to their level of vegetative expression; sites located on the right present high vegetative expression and conversely for sites on the left. Note that this trend was temporally stable since both NDVI parameters measured either in 2008 or in 2009 were strongly correlated. This is consistent with the results obtained by Kazmiersky et al. (2011). This result also justifies the choice of considering only the NDVI measured in 2008 in the rest of the analysis. Hereafter, the NDVI terms will refer exclusively to NDVI08.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Figure 2 highlights a correlation between NDVI and BW either at veraison or at harvest. Therefore, sites with high vegetative expression correspond to sites with high BW and conversely for sites with low vegetative expression. Moreover, this correlation was temporally stable over the two years of NDVI acquisition. It was also temporally stable from veraison to harvest (at least over the two years of NDVI acquisition). However, no correlation was observed between NDVI and BuN or BN. Furthermore, trends highlighted by the PCA masked some disparity between the different fields as shown in Table 4 where the correlation coefficients between NDVI and yield as well as between NDVI and each yield component (BWh, BN and BuN) for the nine fields were calculated.

<Table 4>

Confirming the results of the PCA, a significant correlation between NDVI and BWh was observed for 5 fields (P22, P63, P65, P88, and P104). However, fields P76 and P80 show a low correlation, being practically non-existent for the other two fields (P77 and P82). Conversely, the results showed a low correlation between NDVI and BN except in the fields P65 and P80. Similarly, a low correlation between NDVI and BuN was observed except for field P82, P65 and P80. This resulted in a high variability of the observed correlations between yield and NDVI (Table 4). Considering all the fields of the database together, the observed correlation (r) between NDVI and yield was rather low ($r = 0.31$), although it was statistically significant ($p = 0.05$).

<u>Sensitivity Analysis</u>

The incidence of each yield component in the within-field yield variance was studied through a sensitivity analysis (Table 5). In addition, this analysis allowed calculation of the interactions between different components in the case that they were not

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

independent. According to Table 5, BuN explained over 60% of yield variance while BN and BWh explained 11% and 4%, respectively. An important interaction (second order sensitivity index) between BuN and BN was observed (20 % of the yield variance). This interaction means that BuN and BN were not independent. This observation is logical considering the correlation observed between these two parameters in the PCA (Figure 2). No other interaction was highlighted by the sensitivity analysis.

<Table 5>

As already observed (Rousseau et al., 2008; Santesteban et al., 2013), the results confirm the possibility of observing a relationship between the NDVI and yield at the within-field level. Considering each yield component, the results provided further information on a significant data-base which encompasses 9 different fields and two different varieties. Indeed, under study conditions, the results showed that among all the yield components, BW (BWv and BWh) was the most closely correlated with NDVI for most of the vine fields.

The low correlations observed between NDVI and BuN or BN can certainly be explained by the impact of winter pruning which tends to control the BuN at the within-field level. Despite the significant impact that BuN has on yield variability, pruning is a manual operation adapted for each vine which may tend to smooth environmental effects and its potential correlation with vegetative expression (NDVI). In similar conditions, non-pruned vineyards show a decrease in BuN in low vegetative expression zones compared to high vegetative expression zones (Rousseau et al, 2013). It is not clear why BN is not affected by the vegetative expression in the current experiment, when, on the contrary, Champagnol (1984) reported that BN may be affected by vigour.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

However, BN is determined by many factors whose incidence is complex. Meteorological conditions and vigour during bunch initiation the previous year as well as meteorological conditions at flowering largely determine BN (Carbonneau and Ollat, 1993). Although meteorological conditions can be considered uniform at the within-field level, the complexity of the phenomena involved may explain the lack of clear correlation between NDVI and BN in this experiment.

Regarding the sensitivity analysis, note that a very similar analysis was carried out by Clingeleffer et al. (2001). These authors considered the impact of each yield component on the yield variability from one year to another. They showed that BuN explains 60 to 70% of the seasonal variation in vine yield. Yield fluctuation over the years was less sensitive to BN (~20%) and less sensitive again to berry weight (~10%). It is interesting to note that the relative importance of the yield components which affect yield variability is rather consistent both spatially and temporally.

Regarding the use of NDVI values to optimize the estimation of yield components, it is difficult to make a clear recommendation. BW presents the highest correlation with NDVI, therefore target sampling BW according to NDVI values could improve mean yield field estimation. However BW is the yield component with the lowest impact on yield variability (4%). Therefore, the expected improvement using BW in an optimized sampling strategy may have a limited effect on the quality of mean yield estimation. Conversely, BuN and BN present low correlations with NDVI for most of the fields, but these components have a high impact on yield variability. Therefore incidence of both BuN and BN optimized sampling may be significant on yield estimation (at least for some fields). This observation justifies testing all the sampling strategies (Table 2) with all the yield components.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Results of the Sampling Strategies

Figure 3 shows the mean error (%) (Eq. 7) observed over the nine fields with the different sampling strategies and for a number of sampling sites ranging from 3 to 7.

<Figure 3>

Three additional pieces of information were added to Figure 3 in order to analyse the results properly:

- THEO, the error computed from a theoretical field (Eq. 8),
- operative number of samples (5 samples) corresponding more or less to the current methods used by the wine industry to estimate the mean field yield at harvest
- and expected error (10 %) by the wine industry.

As expected, whatever the sampling strategy, the error (%) decreases with an increased number of sampling sites. The decrease is consistent for each method from 3 to 7 sampling sites; the error decreases by ~13%. THEO superimposes perfectly with error from RM. This result demonstrates the relevance of the bootstrapping method to approximate the distribution of mean yield estimations from a random sampling.

Although they are not statistically different, errors observed for each sampling method are ordered in the same way whatever the number of sites. Sampling strategies based on NDVI values (SBN) systematically improve the estimation by at least ~5-7 % compared to the random method (RM). A lower error is consistently observed when both yield components BuN and BuW (Eq. 2) are estimated from a NDVI distribution (TM or MM). Random selection of sites for estimating yield component BuN (RTM or RMM) therefore results systematically in a higher error. Note however that the difference between both sets of approaches (TM or MM vs RTM or RMM) is very low (~ 2 %) but consistent whatever the number of sites considered. Although BuN has a low correlation

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

with NDVI, the implementation of a target sampling for this yield component seems to be of interest to improve yield estimation.

The sampling approach currently used by the wine industry is similar to the RM method with 5 sites. This approach results in a mean error of ~29 %. This high value shows the uncertainty of the current methods caused by the large within-field variability of yield in viticulture. This also highlights the necessity to provide the wine industry with more efficient sampling methods. Figure 3 confirms the value of using sampling strategies based on NDVI values. However, regarding current operational constraints (5 sites), the best sampling method still leads to an error of ~ 23 %. Thus, none of the methods tested in this experiment achieve the error (10%) expected by the wine industry. As shown by error trends (Figure 3), to satisfy their expectations, the solution would be to increase the number of sampling sites in order to decrease the error. However, this solution is costly and would increase the working time.

Table 6 shows the large diversity of results observed for the nine fields of the database. It only focuses on RM, TM and MM with five sites, which may approximate the current sampling strategy used by the wine industry. For four fields (P65, P76, P77 and P80) a rather large decrease of the error was observed when implementing a sampling strategy based on NDVI values. These fields also present the highest CV (~77-80 %) (Table 3). For two fields (P82 and P88), a small decrease of the error was observed, and for the remaining fields (P22, P63 and P104) no decrease of the error was observed. This heterogeneity in the results could explain the lack of significance observed in Figure 3. As no clear relationship could not be demonstrated between field characteristics (soil unit, variety, mean yield) and the decrease in the error, it was assumed that the database may not be large enough to identify any clear relationship.

<Table 6>

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Sampling strategies based on NDVI are of value to improve grape yield estimation. Indeed, with the same number of samples and MM method, the grape yield estimation can be improved by 10 %, compared to conventional sampling (random). Depending on the fields, grape yield estimation can be improved from 20% to 0%. This shows that yield estimations may only be improved (and never damaged) when based on a sampling strategy based on NDVI values.

However, the proposed sampling strategies do not allow achieving the accuracy expected by the wine industry. Several issues deserve further consideration:

i) Significant improvements may be proposed in order to better take into account the distribution of NDVI values, but also the spatial structure of this information to optimize the location of the measurement sites as proposed by Stein and Ettema (2003),

ii) Sampling within sites was assumed to be optimal. Therefore, errors due to the operator in counting the number of clusters, incidence of berry selection as well as cluster selection to approximate BuW were assumed to be low. As investigated by other authors (Meyers et al. 2011; Wulfsohn et al. 2012), methods aiming at optimizing sampling within the plant or between plants at the within-site level may improve the yield estimation. These approaches deserve to be tested in addition to a sampling strategy based on NDVI.

iii) Incidence of the variety as well as training systems must be investigated. In particular, there is little work on the effect of these parameters on the relationship between BuN, BN and the NDVI. On non-pruned vineyards or mechanically pruned vineyards, a better correlation between NDVI and BuN is expected (Rousseau et al., 2013). Therefore, significant gain in yield estimation may be observed using sampling strategies based on NDVI in these training systems. BuN and BN are yield components which impact significantly the yield variance at the within-field level. If, for a given

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

variety or training system, the correlation between these components and the NDVI is higher than that observed in this experiment, then better results could be expected with a sampling strategy based on NDVI values.

iv) The work assumed that NDVI information is available before flowering to design a sampling strategy. The spatial organization of NDVI values can vary between flowering and harvest (Kazmiersky et al., 2011, Hall et al., 2011). In this case, the NDVI image acquired before flowering may be more appropriate to design a sampling strategy for BuN. In the case where NDVI information would not be available for the first stage of the method (estimation of BuN, Eq. 3) at flowering, RTM or RMM (Table 3) should be recommended.

v) Practical aspects related to the measurement of some yield components were not considered in this work. Indeed, BuN is quick and easy to measure at flowering (< 1 min. per site), and it also presents the advantage of being non-destructive. Conversely, BuW (BN and BW) estimation is a destructive method and takes longer (> 5 min. per site). A simple recommendation would be to increase the number of sites randomly distributed for BuN at flowering while maintaining a limited number of sites defined with NDVI for BuW at harvest. This recommendation corresponds to RTM or RMM methods with a different number of sites at each stage. It has the advantage of limiting the sampling effort at harvest when technicians of wineries are usually very busy. Figure 4 shows the results obtained with this approach using the database. RMM and RTM were implemented with a number of sites between 5 to 15 for the first stage (flowering) and a limited number of sites between 5 to 7 for the second stage (harvest). Figure 4 shows that the combination of these two approaches improves yield estimation by 9 %. The lowest mean error observed (15 %) is close to the expected error.

< Figure 4 Near here>.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

vi) In this study, it is assumed that the mean of all sites within a field is the true mean. A more reliable reference would be to compare the estimations with the total amount of harvest weighed at the winery. In the current case, this was not possible because differentiated harvests (manual and mechanical) were carried out on several parts of each field. Differentiated harvests were due to the experiments undertaken by other researchers on the Pech-Rouge Vineyard. In manual harvesting, the whole bunch including the stalks is collected while in mechanical harvesting, only berries are collected. Stalks represent approximately 5 % of the bunch weight. These two types of harvest induced an inaccuracy in the total weight measured at the winery justifying the method used in this paper to estimate the mean grape yield.

vii) Finally, this study assumed that the number of missing or unproductive vines is correctly estimated. In some situations, the percentage of missing plants is an important source of imprecision in yield estimation. Furthermore, in areas with high levels of missing plants, a negative correlation was observed between NDVI and some yield components (low values of NDVI for high values of BW and BuN) because the few remaining vines had high vigor individually. Note that the use of remote sensing images with a suitable resolution can be used to count the missing plants (Robbez-Masson and Foltete, 2005) and may be helpful in detecting these specific situations.

Conclusions

This study, based on a database from nine different fields, showed the value of NDVI information to optimize yield sampling. NDVI presents the highest correlation with berry weight (BW) which, unfortunately, is the yield component with the lowest impact on yield variability. As a result, depending on the field considered, sampling based on NDVI provides marginal improvements on yield estimation. On average, yield estimation can be improved by 10%. Therefore target sampling based on NDVI can be recommended to the wine industry. Using NDVI as auxiliary information is particularly interesting to stratify sampling (target

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

sampling) for berry weight (and bunch weight) estimation but no significant value was demonstrated when trying to model the number of bunches. Note however that for some fields, yield estimation was significantly improved when target sampling was applied to the number of bunches, showing the potential interest of the approach in specific conditions.

Acknowledgements

We gratefully acknowledge the Experimental station of Pech-Rouge (INRA). This work was funded by the EC, French Government, OSEO, Région Languedoc-Roussillon, through Vinnotec project (Qualimed Pole of Languedoc-Roussillon region—France)

References

Acevedo-Opazo, C., Tisseyre, B., Guillaume, S., & Ojeda, H. (2008). The potential of high spatial resolution information to define within-vineyard zones related to vine water status. *Precision Agriculture*, *9*(5), 285-302.

Baldwin, J. (1964). The relation between weather and fruitfulness of the sultana vine. *Australian Journal of Agricultural Research*, 15, 920-928.

Blom, P.E. and Tarara, J.M. (2009). Trellis Tension Monitoring Improves Yield Estimation in Vineyards. *HortScience*, 44(3), 678-685.

Bramley R.G.V., Proffitt A.P.B., Hinze C.J., Pearse B. and Hamilton R.P., (2005). Generating benefits from precision viticulture through differential harvest. In: Stafford, J. V. (Ed.) *Proceedings of the 5th European Conference on Precision Agriculture*, Wageningen Academic Publishers, The Netherlands. pp. 891-898.

Carbonneau, A. and Ollat, N. (1993). Etude de la coulure et maîtrise de la production.(study on vines shattering and yield control) *Progrès. Agricole et Viticole*. 110 (15-16), 331-340.

Champagnol, F. (1984). *Elements of the physiology of the vine and of general viticulture*: Impr: Déhan, Montpellier, France.

Clingeleffer, P. R., Martin, S., Krstic, M., & Dunn, G. M. (2001). *Crop Development, Crop Estimation and Crop Control to Secure Quality and Production of Major Wine Grape Varieties: A National Approach: Final Report to Grape and Wine Research & Development Corporation*: Grape and Wine Research & Development Corporation [CSIRO and NRE: Victoria, Australia, 148].

Coulouma, G., Tisseyre, B., & Lagacherie, P. (2010). Is a systematic two dimensional EMI soil survey always relevant for vineyard production management? A test on two pedologically contrasting Mediterranean vineyards (Chap. 24). In: R. A. Viscarra-Rossel, A. B. McBratney, & B. Minasny (Eds.), Proximal soil sensing. Progress in soil science series . Heidelburg, Germany: Springer (in press). ISBN 978-90-481-8858-1.

Cristofolini F. & Gottardini E. (2000). Concentration of airborne pollen of Vitisvinifera L. and yield forecast: a case study at S.Michele all'Adige, Trento, Italy. *Aerobiologia*, 16, 171-216.

Diago M.P., Correa C., Millán B., Barreiro P., Valero C., Tardaguila J. (2012). Grapevine Yield and Leaf Area Estimation Using Supervised Classification Methodology on RGB Images Taken under Field Conditions. *Sensors*, *12*, 16988-17006.

Dunn, G.M.; Martin, S.R. (2004). Yield prediction from digital image analysis: A technique with potential for vineyard assessments prior to harvest. *Australian Journal of Grape and Wine Research*, 10, 196–198.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of Statistics*, *7(1)*, 1-26.

Grocholsky B., Nuske S., Aasted M., Achar S. and Bates T. (2011). A Camera and Laser System for Automatic Vine Balance Assessment. *ASABE Technical Library*, Paper no. 11-11651, ASABE, St Joseph, MI, USA.

Hall, A., Lamb, D.W., Holzapfel, B.P., & Louis, J.P. (2011). Within-season temporal variation in correlations between vineyard canopy and winegrape composition and yield. Precision Agriculture, 12, (1), 103-117.

Hengl, T., Gruber, S., & Shrestha, D. (2003). Digital terrain analysis in ILWIS. Lecture notes: *International Institute for Geo-Information Science and Earth Observation (ITC) Enschede, The Netherlands*, https://www.itc.nl/library/Papers_2003/misca/hengl_digital.pdf. Last access [22/04/2015].

Kazmierski, M., Glemas, P., Rousseau, J., & Tisseyre, B. (2011). Temporal stability of within-field patterns of NDVI in non irrigated Mediterranean vineyards. *Journal international des sciences de la vigne et du vin, 45*(2), 61-73.

Lamb, D. W., Weedon, M., & Bramley, R. (2004). Using remote sensing to predict grape phenolics and colour at harvest in a Cabernet Sauvignon vineyard: Timing observations against vine phenology and optimising image resolution. *Australian Journal of Grape and Wine Research, 10*(1), 46-54.

Lesch, S. (2005). Sensor-directed response surface sampling designs for characterizing spatial variation in soil properties. *Computers and Electronics in Agriculture*, 46(1-3), 153-179.

Lesch, S. M., Strauss, D. J., & Rhoades, J. D. (1995). Spatial prediction of soil salinity using electromagnetic induction techniques: 1. Statistical prediction models: A comparison of multiple linear regression and cokriging. *Water Resources Research, 31*, (2), 373-386.

Martinez-Casasnovas J.A. & Bordes X. (2005). Viticultura de precisión: predicción de cosecha a partir de variables del cultivo e índices de vegetación (Precision viticulture: yield prediction from crop variables and vegetation indices). *Revista de la Asociación Española de Teledetección*, 24, 67-71.

Meyers, J.M., Sacks, G.L., van Es, H.M. & Vanden Heuvel, J.E. (2011). Improving vineyard sampling efficiency via dynamic spatially-explicit optimisation. *Australian Journal of Grape and Wine Research*, 17, 306-315.

Nuske, S., Achar, S., Bates, T., Narasimhan, S., & Singh, S. (2011). Yield estimation in vineyards by visual grape detection. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011), [2352-2358] doi:10.1109/iros.2011.6048830.

Rabatel G. and Guizard C. (2007). Grape berry calibration by computer vision using elliptical model fitting. In: J.V.Stafford (Ed.) Precision Agriculture '07, Proceedings of the 6th European Conference on Precision Agriculture, Wageningen Academic Publishers, The Netherlands, pp 581-587

Reis M. J. C. S., Morais R., Peres E., Pereira C., Contente O., Soares S. (2012). Automatic detection of bunches of grapes in natural environment from color images. *Journal of Applied Logic*, 10, 285-290.

Robbez-Masson J. M., and Foltete J.C.(2005). Localising missing plants in squared grid patterns of discontinuous crops from remotely sensed imagery. *Computers & Geosciences* 31, 900-912.

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. In: S. C. Freden & M. A. Becker (Eds.), Third ERTS Symposium (pp. 309–317). Greenbelt, MD: NASA Goddard Space Flight Center

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Rousseau J., Dupin S., Acevedo-Opazo C., Tisseyre B., Ojeda H. (2008). L'imagerie aérienne : application à la caractérisation des potentiels viticoles et oenologiques [Airborne imagery: application to the characterization of viticultural and oenological potential], *Bulletin de l'organisation internationale de la vigne et du vin, 81*, 507-517.

Rousseau, J., Pic, L., Carbonneau, A. and Ojeda, H. 2013. Incidence of minimal pruning on wine quality. *Acta Horticolturae.* (ISHS) 978:309-316

Santesteban, L., Guillaume, S., Royo, J., & Tisseyre, B. (2013). Are precision agriculture tools and methods relevant at the whole-vineyard scale? *Precision Agriculture, 14*(1), 2-17.

Scholander, P. F., Bradstreet, E. D., Hemmingsen, E., & Hammel, H. (1965). Sap pressure in vascular plants negative hydrostatic pressure can be measured in plants. *Science, 148*(3668), 339-346.

Serrano, E.; Roussel S., Gontier L., Dufourcq T. (2005). Estimation précoce du rendement de la vigne: corrélation entre le volume de la grappe de vitis vinifera en cours de croissance et son poids à la récolte (Early grape yield estimation: correlation between the volume of the cluster of vitis vinifera during growth and harvest weight), In: Proceeding of the Groupe Européen d'Etude des Systémes de Conduite de la Vigne, Geisenheim/Schultz, Hans (ed), 311-318.

Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments, 1*(4), 407-414.

Stein, A., and Ettema, C. (2003). An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. *Agriculture, Ecosystems and Environment*, 94, 31–47.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Taylor, J., Acevedo-Opazo, C., Ojeda, H., & Tisseyre, B. (2010). Identification and significance of sources of spatial variation in grapevine water status. *Australian Journal of Grape and Wine Research, 16*(1), 218-226.

Taylor, J., Tisseyre, B., Bramley, R., Reid, A. (2005). A comparison of the spatial variability of vineyard yield in European and Australian production systems. In: J.V.Stafford (Ed.) Precision Agriculture '05, Proceedings of the 5th European Conference on Precision Agriculture, Wageningen Academic Publishers, The Netherlands, pp. 907-914.

Wolpert, J., and Vilas, E. (1992). Estimating vineyard yields: Introduction to a simple, two-step method. *American Journal of Enology and Viticulture, 43*(4), 384-388.

Wulfsohn D, Aravena-Zamora F, Potin-Téllez C, Zamora I, García-Fiñana M (2012) Multilevel systematic sampling to estimate total fruit number for yield forecasts. *Precision Agriculture*, *13(2)*, 256-275

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

List of Tables

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

List of Figures

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Table 1 Description of the characteristics and management practices of the 9 fields used in the study.

Field (Id)	Area (ha)	Variety	Date of plantation	Training system	Row spacing (m)	Vine spacing (m)	Pedological Unit	Sampling sites
P22	1.72	Syrah	1995	VSP	2.5	1	PU3	45
P77	1.24	Syrah	1990	VSP	2.5	1	PU2	19
P76	1.14	Carignan	1990	VSP	2.25	1.5	PU2	37
P63	1.33	Syrah	1993	VSP	2.5	1	PU2	42
P80	0.54	Syrah	1978	VSP	2.5	1	PU2	40
P65	0.69	Syrah	1974	VSP	2.5	1	PU2	33
P82	1.15	Syrah	1977	Gobelet	2.5	1	PU2	53
P88	0.85	Syrah	2004	VSP	2.25	1.5	PU2	21
P104	0.81	Carignan	1961	Gobelet	2.25	1.5	PU1	23

VSP: Vertical Shoot Positioning. PU1: Calcisols/Regosols (clayic); PU2: Calcisols (skeletic); PU3: Endosalic Arenosols.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Table 2 Sampling strategies

Methods	**1st step (BuN)**	**2nd step (BuW)**	
		BN	**BW**
Random (RM)	Random sampling (RS)	Random sampling (RS)	Random sampling (RS)
Random-target (RTM)	Random sampling (RS)	Target sampling (TS)	Target sampling (TS)
Target (TM)	Target sampling (TS)	Target sampling (TS)	Target sampling (TS)
Random-model (RMM)	Random sampling (RS)	Target sampling (TS)	Model sampling (MS)
Model (MM)	Target sampling (TS)	Target sampling (TS)	Model sampling (MS)

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Table 3. Statistics over the 9 fields used in the study.

Field (Id)	Mean yield (t/ha)	Coefficient of variation (CV) %	Min. yield (t/ha)	Max. yield (t/ha)
p22	7.06	56.20	0.20	19.62
p63	4.42	61.18	0.36	13.83
p65	4.59	77.07	0.05	13.73
p76	2.76	80.24	0.27	9.48
p77	5.71	71.89	1.01	13.76
p80	3.31	76.60	0.07	10.40
p82	3.80	63.42	0.13	11.11
p88	6.88	35.81	3.21	14.65
p104	7.01	39.81	2.22	13.13

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Table 4. Correlation coefficients (r) between NDVI (2008), and each yield component, and the yield for the nine fields of the experiment. (BWh, Berry weight at harvest, BN; Berry number, BuN; Bunch number, Y; Yield)

Field (Id)	BWh vs. NDVI	BN vs. NDVI	BuN vs. NDVI	Y vs. NDVI
p22	0.49*	-0.12	0.07	0.04
P63	0.55*	0.05	0.22	0.25
P65	0.83*	0.82*	0.84*	0.81*
P76	0.28	0.11	0.20	0.25
P77	-0.02	0.33	0.43	0.51*
P80	0.16	0.32*	0.61*	0.47*
P82	0.03	0.08	0.35*	0.34*
P88	0.71*	0.27	-0.28	0.52*
P104	0.64*	-0.16	0.18	-0.04

***Significant at the 0.05 probability level**

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Table 5. Sensitivity analysis by Sobol's method.

Yield components	BWh %	BN %	BuN %	Interaction BuN and BN %
Sobol index	4	11	60	20

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

Table 6. Error on yield estimation for the random (RM), target (TM) and model (MM) strategies, for a 5-sites sampling.

Field (Id)	RM %	TM %	MM %
P22	24.90	25.15	25.97
P63	26.88	25.43	25.61
P65	35.03	15.42	20.48
P76	34.80	30.69	28.63
P77	33.51	26.68	25.61
P80	32.85	27.17	25.77
P82	27.37	25.12	22.72
P88	11.96	9.66	8.78
P104	17.84	19.46	19.50

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

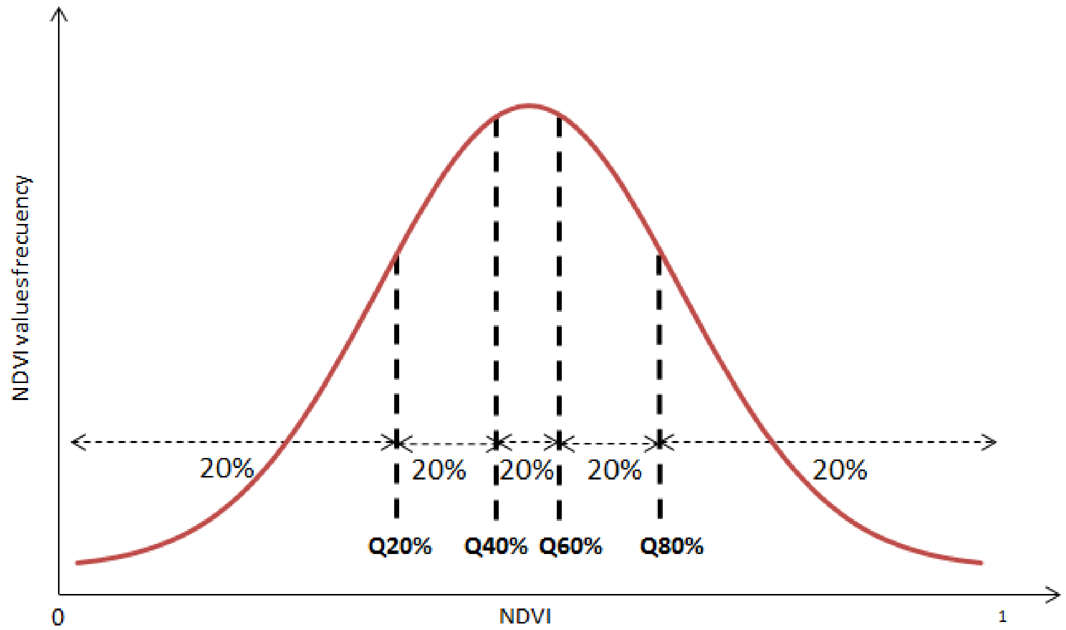

Fig. 1 Example of 20% percentile division of NDVI distribution to define intervals used to perform targeted sampling for 5 sites.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

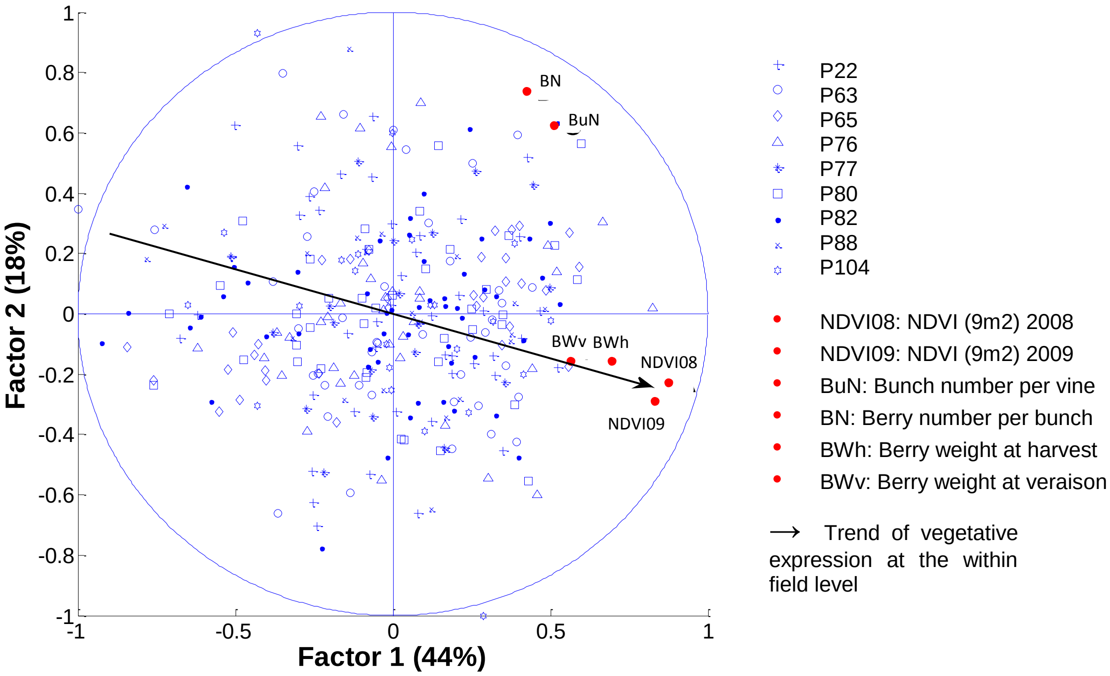

Fig 2, Scattered plot and correlation coefficients of the principal component analysis (first 2 Factors) with data centered and reduced according to a field by field basis.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

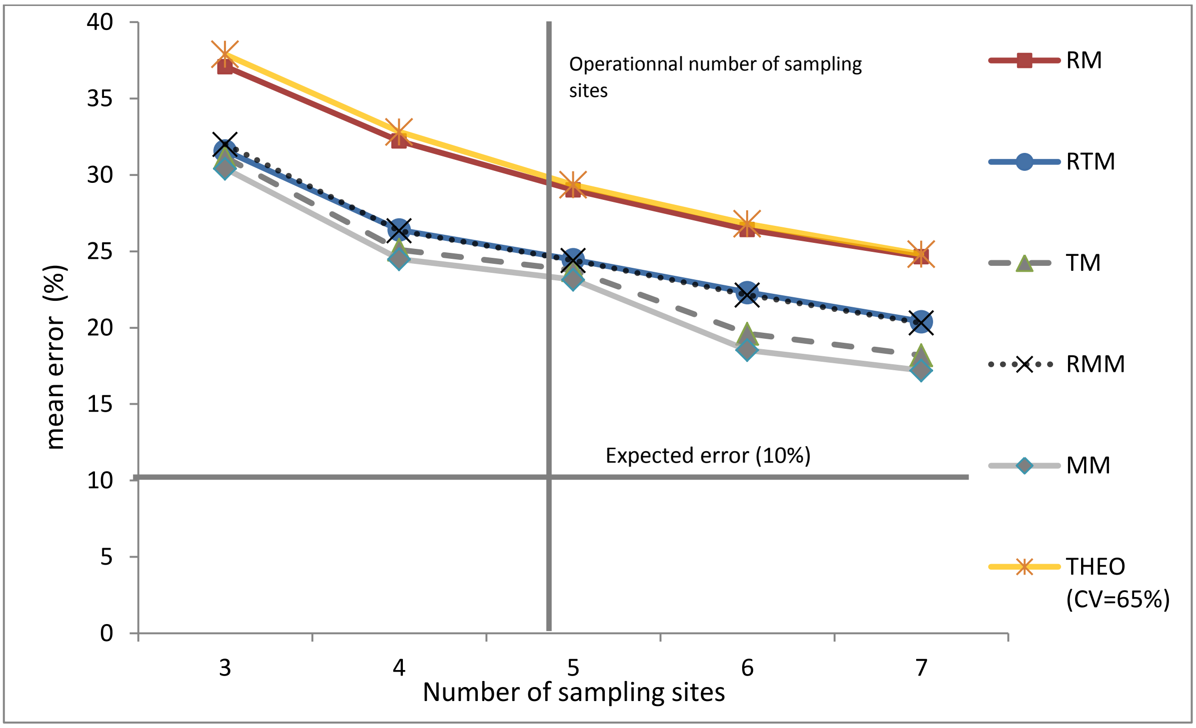

Fig 3. Mean field error (%) of the different sampling strategies in relation to the number of sampling sites. Mean error is computed over the nine fields of the experiment for the different sampling approaches: RTM: Random-Target, TM: Target, RMM: Random-Model, MM: Model, THEO: random sampling for a theoretical normal distribution corresponding to CV = 65% (mean CV of yield observed on the data base).

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.74-92.
The original publication is available at http://link.springer.com
Doi: 10.1007/s11119-015-9407-8

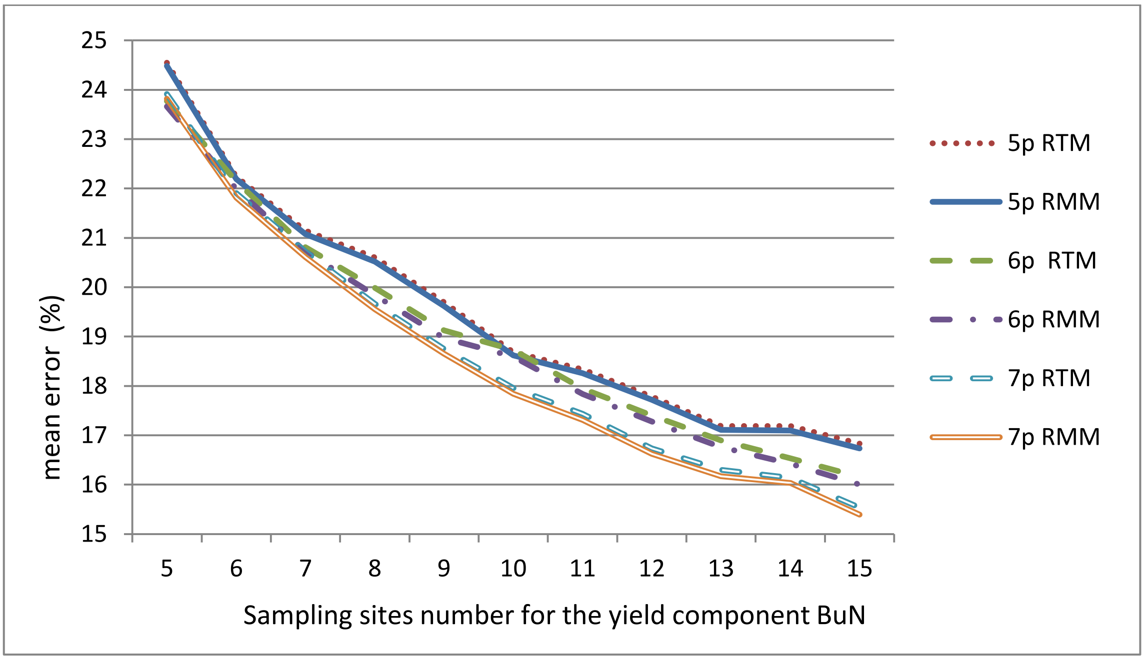

Fig 4. Mean error (%) from the nine fields of the database obtained for RTM and RMM as a function of the number of sampling sites for BuN (Bunch Number) and for BuW (bunch weight) independently. 5p, 6p and 7p correspond respectively to 5, 6 and 7 sampling sites for BuW. RTM: Random-Target, RMM: Random-Model.