

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

2 E. CARRILLO¹, A. MATESE², J. ROUSSEAU³ and B. TISSEYRE^{1*}

3 **Abstract**

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

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

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

28 ¹UMR ITAP, Montpellier SupAgro/IRSTEA, 2 Place Pierre Viala. 34060 Montpellier

29 Cedex 2. France

30 *Corresponding author: bruno.tisseyre@supagro.fr

31 ²Istituto di Biometeorologia (IBIMET – CNR), Consiglio Nazionale delle Ricerche, via

32 Caproni 8, 50145 Firenze. Italy

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

34 **Introduction**

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

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

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

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

63 ii) the second type of error is caused by the variability at the vine level (within-plant
64 variability); bunch weight presents a high within-plant variability (Clingeffer et al.
65 2001). Measuring this yield component is costly and destructive. Classical methods
66 provide an estimation of this parameter using a small number of representative clusters
67 (Clingeffer et al. 2001). To minimize errors of estimation due to the choice of the
68 clusters, different systematic methods have been proposed in the literature (Wulfsohn et
69 al., 2012, Meyers et al., 2011). Other studies propose the use of alternative sources of
70 information to facilitate or automate BW estimates. This is the case of image analysis
71 that has been proposed to detect, count and estimate the weight of clusters (Diago et al.,
72 2012, Reis et al., 2012, Nuske et al., 2011, Serrano et al., 2005, Dunn and Martin,
73 2004,) or to estimate the number and the volume of berries (Grocholsky et al., 2011,

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

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

84 iv) finally, errors can also occur due to within-field variability; Taylor et al. (2005)
85 showed that the coefficient of variation (CV) of the yield is very high in viticulture (CV
86 ~ 50 %) and, what is more important, yield variation is not randomly distributed but
87 presents a strong spatial organization (spatial autocorrelation). Surprisingly, none of the
88 sampling methods proposed in the literature take into account the spatial organization of
89 the yield; indeed, most of the yield estimation methods are based on a random selection
90 of the sampling sites (Clingeffer, 2001, Wolpert and Vilas, 1992).

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

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

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

110 **Materials and Methods**

111 Experimental Site

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

122 vineyard, estimated by pre-dawn leaf water potential measurements (Scholander et al.
123 1965) was between -0.75 MPa in August 2003 (dry year) and -0.60 MPa in August
124 2006 (wet year) (Taylor et al., 2010).

125 Previous work (Acevedo-Opazo et al., 2008; Kazmierski et al., 2011) showed that in
126 this vineyard, water restriction is the main factor affecting the growth and the yield of
127 the vines. The soil variability is the main factor explaining the within-field spatial
128 variability of the vine water status, therefore vigour and estimated vigour through
129 vegetation indices like the NDVI derived from airborne images, were relevant surrogate
130 information to highlight within-field zones of water restriction (Acevedo-Opazo et al.,
131 2008). As a result, NDVI presented a significant temporal stability of the spatial
132 variability at both an intra-annual scale and inter-annual scale (Kazmierski et al., 2011).

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

139 **<Table 1>**

140 Image Acquisition and Processing

141 Two multi-spectral airborne images were taken before veraison in 2008 (31st July) and
142 2009 (1st August). Images of 1m resolution were collected by Avion Jaune (Montpellier,
143 Hérault, France). The spectral regions captured in the images were: i) blue (445-
144 520nm), ii) green (510-600nm), iii) red (632-695nm) and iv) near-infrared (757-

145 853nm). The 1 m square image pixels were aggregated into 3 m square pixels using the
146 methodology outlined in Acevedo-Opazo et al. (2008), which approximates the “mixed
147 pixel” row spacing approach of Lamb et al. (2004). The calculation of NDVI (Rouse et
148 al., 1973) was then made on the 3 m pixels (area of 9 m²). Note that mechanical or
149 chemical weeding was performed over the inter-row spacing; therefore row cover crop
150 did not affect NDVI values.

151 Sampling Sites

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

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

165 Field Measurements

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

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

180 The final data base was a set of 313 sites over the 9 different fields. Each site was then
181 characterized by 5 field parameters (BW_v, BW_h, BN, BuN, and BuW) and 2 remote
182 sensing parameters (NDVI₀₈ and NDVI₀₉), i.e. NDVI values measured in 2008 and
183 2009, respectively.

184 Data Analysis

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

188 Variance-Based Sensitivity Analysis

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

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

$$196 \quad Var(Y) = \sum_{i=1}^3 V_i + \sum_{i<j}^3 V_{ij} + \dots + V_{123} \quad (1)$$

197 where:

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

199 V_i : Variance of each yield component.

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

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

$$204 \quad S_i = \frac{V_i}{Var(Y)}, \text{ with } \sum_{i=1}^3 S_i + \sum_{i<j}^3 S_{ij} + \dots + S_{123} = 1 \quad (2)$$

205 where:

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

208 Proposed Sampling Methods

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

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

$$217 \qquad Y = BuN * BuW \qquad (3)$$

$$218 \qquad \text{with } BuW = BN * BW$$

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

225 <Table 2>

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

232 - Random Sampling (RS)

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

236 - Target Sampling (TS)

237 Sites are chosen according to the distribution of NDVI values. For the field under
238 consideration, NDVI values are divided into $100/n$ % quantiles, with n corresponding to
239 the desired number of sites. Among sites corresponding to each quantile, one site is
240 randomly selected. Therefore, TS is a way to stratify the site selection according to
241 NDVI values. Figure 1 illustrates the TS method with $n=5$ sites. This example led to the
242 consideration of 5 intervals on the NDVI values corresponding to quantiles 0-20 %, 20-
243 40%, 40-60%, 60-80%, and 80-100%. For each interval, one site is randomly chosen.
244 Then, an estimation of the mean field yield is computed following Eq. 3.

245 <Figure 1>

246 - Model Sampling (MS)

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

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

$$259 \quad \widehat{BWh}(s) = a * NDVI(s) + b \quad (4)$$

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

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

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

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

276 Evaluation of Sampling Methods

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

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

283
$$\hat{Y} = \frac{1}{B} \sum_{b=1}^B \hat{Y}_b \quad (5)$$

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

286
$$\widehat{V}(Y) = \frac{1}{B} \sum_{b=1}^B (\hat{Y}_b - \hat{Y})^2 \quad (6)$$

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

289
$$Error (\%) = \frac{\widehat{\sigma}(Y)}{\hat{Y}} \times 100 \quad \text{with} \quad \widehat{\sigma}(Y) = \sqrt{\widehat{V}(Y)} \quad (7)$$

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

296
$$THEO (\%) = \frac{1}{\sqrt{n}} \cdot t \cdot CV \quad (8)$$

297 where :

298 n : is the number of sites

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

300 CV (65%): the coefficient of variation,

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

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

307 **Results and Discussion**

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

312 Yield Spatial Variability at the Within-Field Level

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

317 <Table 3>

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

321 (CV values are above 35%), confirming what Taylor et al. (2005) had already observed
322 for a larger data-base obtained with grape harvesters and yield monitoring systems. In
323 the current case, 4 fields (p65, p76, p77 and p80) presented very high CV values, above
324 70%. This result confirms the significance of the within-field variability for grape yield,
325 and the potential interest of proposing sampling methods adapted to this yield
326 variability.

327 Relationship Between Yield Components and NDVI

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

334 <Figure 2>

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

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

355 <Table 4>

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

365 Sensitivity Analysis

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

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

376 <Table 5>

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

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

393 However, BN is determined by many factors whose incidence is complex.
394 Meteorological conditions and vigour during bunch initiation the previous year as well
395 as meteorological conditions at flowering largely determine BN (Carbonneau and Ollat,
396 1993). Although meteorological conditions can be considered uniform at the within-
397 field level, the complexity of the phenomena involved may explain the lack of clear
398 correlation between NDVI and BN in this experiment.

399 Regarding the sensitivity analysis, note that a very similar analysis was carried out by
400 Clingeleffer et al. (2001). These authors considered the impact of each yield component
401 on the yield variability from one year to another. They showed that BuN explains 60 to
402 70% of the seasonal variation in vine yield. Yield fluctuation over the years was less
403 sensitive to BN (~20%) and less sensitive again to berry weight (~10%). It is interesting
404 to note that the relative importance of the yield components which affect yield
405 variability is rather consistent both spatially and temporally.

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

417 Results of the Sampling Strategies

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

420 **<Figure 3>**

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

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

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

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

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

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

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

465

<Table 6>

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

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

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

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

484 iii) Incidence of the variety as well as training systems must be investigated. In
485 particular, there is little work on the effect of these parameters on the relationship
486 between BuN, BN and the NDVI. On non-pruned vineyards or mechanically pruned
487 vineyards, a better correlation between NDVI and BuN is expected (Rousseau et al.,
488 2013). Therefore, significant gain in yield estimation may be observed using sampling
489 strategies based on NDVI in these training systems. BuN and BN are yield components
490 which impact significantly the yield variance at the within-field level. If, for a given

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

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

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

515 < Figure 4 Near here>.

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

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

534 **Conclusions**

535 This study, based on a database from nine different fields, showed the value of NDVI
536 information to optimize yield sampling. NDVI presents the highest correlation with berry
537 weight (BW) which, unfortunately, is the yield component with the lowest impact on yield
538 variability. As a result, depending on the field considered, sampling based on NDVI provides
539 marginal improvements on yield estimation. On average, yield estimation can be improved by
540 10%. Therefore target sampling based on NDVI can be recommended to the wine industry.
541 Using NDVI as auxiliary information is particularly interesting to stratify sampling (target

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

546 **Acknowledgements**

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

550 **References**

- 551 Acevedo-Opazo, C., Tisseyre, B., Guillaume, S., & Ojeda, H. (2008). The potential of
552 high spatial resolution information to define within-vineyard zones related to
553 vine water status. *Precision Agriculture*, 9(5), 285-302.
- 554 Baldwin, J. (1964). The relation between weather and fruitfulness of the sultana vine.
555 *Australian Journal of Agricultural Research*, 15, 920-928.
- 556 Blom, P.E. and Tarara, J.M. (2009). Trellis Tension Monitoring Improves Yield
557 Estimation in Vineyards. *HortScience*, 44(3), 678-685.
- 558 Bramley R.G.V., Proffitt A.P.B., Hinze C.J., Pearse B. and Hamilton R.P., (2005).
559 Generating benefits from precision viticulture through differential harvest. In:
560 Stafford, J. V. (Ed.) *Proceedings of the 5th European Conference on Precision*
561 *Agriculture*, Wageningen Academic Publishers, The Netherlands. pp. 891-898.
- 562 Carbonneau, A. and Ollat, N. (1993). Etude de la coulure et maîtrise de la
563 production.(study on vines shattering and yield control) *Progress. Agricole et*
564 *Viticole*. 110 (15-16), 331-340.
- 565 Champagnol, F. (1984). *Elements of the physiology of the vine and of general*
566 *viticulture*: Impr: Déhan, Montpellier, France.

- 567 Clingeleffer, P. R., Martin, S., Krstic, M., & Dunn, G. M. (2001). *Crop Development,*
568 *Crop Estimation and Crop Control to Secure Quality and Production of Major*
569 *Wine Grape Varieties: A National Approach: Final Report to Grape and Wine*
570 *Research & Development Corporation: Grape and Wine Research &*
571 *Development Corporation [CSIRO and NRE: Victoria, Australia, 148].*
- 572 Coulouma, G., Tisseyre, B., & Lagacherie, P. (2010). Is a systematic two
573 dimensional EMI soil survey always relevant for vineyard production
574 management? A test on two pedologically contrasting Mediterranean vineyards
575 (Chap. 24). In: R. A. Viscarra-Rossel, A. B. McBratney, & B. Minasny
576 (Eds.), Proximal soil sensing. Progress in soil science series . Heidelberg,
577 Germany: Springer (in press). ISBN 978-90-481-8858-1.
- 578 Cristofolini F. & Gottardini E. (2000). Concentration of airborne pollen of *Vitisvinifera*
579 L. and yield forecast: a case study at S.Michele all'Adige, Trento, Italy.
580 *Aerobiologia*, 16, 171-216.
- 581 Diago M.P., Correa C., Millán B., Barreiro P., Valero C., Tardaguila J. (2012).
582 Grapevine Yield and Leaf Area Estimation Using Supervised Classification
583 Methodology on RGB Images Taken under Field Conditions. *Sensors*, 12,
584 16988-17006.
- 585 Dunn, G.M.; Martin, S.R. (2004). Yield prediction from digital image analysis: A
586 technique with potential for vineyard assessments prior to harvest. *Australian*
587 *Journal of Grape and Wine Research*, 10, 196–198.
- 588 Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of*
589 *Statistics*, 7(1), 1-26.

- 590 Grocholsky B., Nuske S., Aasted M., Achar S. and Bates T. (2011). A Camera and
591 Laser System for Automatic Vine Balance Assessment. *ASABE Technical*
592 *Library*, Paper no. 11-11651, ASABE, St Joseph, MI, USA.
- 593 Hall, A., Lamb, D.W., Holzapfel, B.P., & Louis, J.P. (2011). Within-season temporal
594 variation in correlations between vineyard canopy and winegrape composition
595 and yield. *Precision Agriculture*, 12, (1), 103-117.
- 596 Hengl, T., Gruber, S., & Shrestha, D. (2003). Digital terrain analysis in ILWIS. Lecture
597 notes: *International Institute for Geo-Information Science and Earth*
598 *Observation (ITC) Enschede, The Netherlands*,
599 https://www.itc.nl/library/Papers_2003/misca/hengl_digital.pdf. Last access
600 [22/04/2015].
- 601 Kazmierski, M., Glemas, P., Rousseau, J., & Tisseyre, B. (2011). Temporal stability of
602 within-field patterns of NDVI in non irrigated Mediterranean vineyards. *Journal*
603 *international des sciences de la vigne et du vin*, 45(2), 61-73.
- 604 Lamb, D. W., Weedon, M., & Bramley, R. (2004). Using remote sensing to predict
605 grape phenolics and colour at harvest in a Cabernet Sauvignon vineyard: Timing
606 observations against vine phenology and optimising image resolution.
607 *Australian Journal of Grape and Wine Research*, 10(1), 46-54.
- 608 Lesch, S. (2005). Sensor-directed response surface sampling designs for characterizing
609 spatial variation in soil properties. *Computers and Electronics in Agriculture*,
610 46(1-3), 153-179.
- 611 Lesch, S. M., Strauss, D. J., & Rhoades, J. D. (1995). Spatial prediction of soil salinity
612 using electromagnetic induction techniques: 1. Statistical prediction models: A
613 comparison of multiple linear regression and cokriging. *Water Resources*
614 *Research*, 31, (2), 373-386.

- 615 Martinez-Casasnovas J.A. & Bordes X. (2005). Viticultura de precisión: predicción de
616 cosecha a partir de variables del cultivo e índices de vegetación (Precision
617 viticulture: yield prediction from crop variables and vegetation indices). *Revista*
618 *de la Asociación Española de Teledetección*, 24, 67-71.
- 619 Meyers, J.M., Sacks, G.L., van Es, H.M. & Vanden Heuvel, J.E. (2011). Improving
620 vineyard sampling efficiency via dynamic spatially-explicit optimisation.
621 *Australian Journal of Grape and Wine Research*, 17, 306-315.
- 622 Nuske, S., Achar, S., Bates, T., Narasimhan, S., & Singh, S. (2011). Yield estimation in
623 vineyards by visual grape detection. In: IEEE/RSJ International Conference on
624 Intelligent Robots and Systems (IROS 2011), [2352-2358]
625 doi:10.1109/iros.2011.6048830.
- 626 Rabatel G. and Guizard C. (2007). Grape berry calibration by computer vision using
627 elliptical model fitting. In: J.V.Stafford (Ed.) Precision Agriculture '07,
628 Proceedings of the 6th European Conference on Precision Agriculture,
629 Wageningen Academic Publishers, The Netherlands, pp 581-587
- 630 Reis M. J. C. S., Morais R., Peres E., Pereira C., Contente O., Soares S. (2012).
631 Automatic detection of bunches of grapes in natural environment from color
632 images. *Journal of Applied Logic*, 10, 285-290.
- 633 Robbez-Masson J. M., and Foltete J.C.(2005). Localising missing plants in squared
634 grid patterns of discontinuous crops from remotely sensed imagery. *Computers*
635 *& Geosciences* 31, 900-912.
- 636 Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring
637 vegetation systems in the Great Plains with ERTS. In: S. C. Freden & M. A.
638 Becker (Eds.), Third ERTS Symposium (pp. 309–317). Greenbelt, MD: NASA
639 Goddard Space Flight Center

- 640 Rousseau J., Dupin S., Acevedo-Opazo C., Tisseyre B., Ojeda H. (2008). L'imagerie
641 aérienne : application à la caractérisation des potentiels viticoles et oenologiques
642 [Airborne imagery: application to the characterization of viticultural and
643 oenological potential], *Bulletin de l'organisation internationale de la vigne et du*
644 *vin*, 81, 507-517.
- 645 Rousseau, J., Pic, L., Carbonneau, A. and Ojeda, H. 2013. Incidence of minimal pruning
646 on wine quality. *Acta Horticulturae*. (ISHS) 978:309-316
- 647 Santesteban, L., Guillaume, S., Royo, J., & Tisseyre, B. (2013). Are precision
648 agriculture tools and methods relevant at the whole-vineyard scale? *Precision*
649 *Agriculture*, 14(1), 2-17.
- 650 Scholander, P. F., Bradstreet, E. D., Hemmingsen, E., & Hammel, H. (1965). Sap
651 pressure in vascular plants negative hydrostatic pressure can be measured in
652 plants. *Science*, 148(3668), 339-346.
- 653 Serrano, E.; Roussel S., Gontier L., Dufourcq T. (2005). Estimation précoce du
654 rendement de la vigne: corrélation entre le volume de la grappe de vitis vinifera
655 en cours de croissance et son poids à la récolte (Early grape yield estimation:
656 correlation between the volume of the cluster of vitis vinifera during growth and
657 harvest weight), In: Proceeding of the Groupe Européen d'Etude des Systèmes
658 de Conduite de la Vigne, Geisenheim/Schultz, Hans (ed), 311-318.
- 659 Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models.
660 *Mathematical Modelling and Computational Experiments*, 1(4), 407-414.
- 661 Stein, A., and Ettema, C. (2003). An overview of spatial sampling procedures and
662 experimental design of spatial studies for ecosystem comparisons. *Agriculture*,
663 *Ecosystems and Environment*, 94, 31-47.

- 664 Taylor, J., Acevedo-Opazo, C., Ojeda, H., & Tisseyre, B. (2010). Identification and
665 significance of sources of spatial variation in grapevine water status. *Australian*
666 *Journal of Grape and Wine Research*, 16(1), 218-226.
- 667 Taylor, J., Tisseyre, B., Bramley, R., Reid, A. (2005). A comparison of the spatial
668 variability of vineyard yield in European and Australian production systems. In:
669 J.V.Stafford (Ed.) Precision Agriculture '05, Proceedings of the 5th European
670 Conference on Precision Agriculture, Wageningen Academic Publishers, The
671 Netherlands, pp. 907-914.
- 672 Wolpert, J., and Vilas, E. (1992). Estimating vineyard yields: Introduction to a simple,
673 two-step method. *American Journal of Enology and Viticulture*, 43(4), 384-388.
- 674 Wulfsohn D, Aravena-Zamora F, Potin-Téllez C, Zamora I, García-Fiñana M (2012)
675 Multilevel systematic sampling to estimate total fruit number for yield forecasts.
676 *Precision Agriculture*, 13(2), 256-275
- 677

678

List of Tables

679 **Table 1.** Description of the physical condition and management practices over the 9
680 fields used in the study.

681 **Table 2.** Sampling strategies

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

683 **Table 4.** Correlation coefficients (r) between NDVI (2008) and each yield component
684 for the 9 fields of the experiment.

685 **Tables 5.** Sensitivity analysis by Sobol's method.

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

688

689

690

List of Figures

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

693 **Figure 2.** Scatter plot and correlation coefficients of the principal component analysis
694 (first 2 Factors) with data centered and reduced according to a field by field basis.

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

700 **Figure 4.** Mean error (%) from the nine fields of the database obtained for RTM and
701 RMM as a function of the number of sampling sites for the yield component A and for
702 yield component B independently. 5p, 6p and 7p correspond respectively to 5, 6 and 7
703 sampling sites for the component B. RTM: Random-Target, RMM: Random-Model.

704

705

706 **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

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

708

709

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)

710

711

712 **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

713

714

715 **Table 4.** Correlation coefficients (r) between NDVI (2008), and each yield component, and the yield for
716 the nine fields of the experiment. (BWh, Berry weight at harvest, BN; Berry number, BuN; Bunch
717 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**

718

719 **Table 5.** Sensitivity analysis by Sobol's method.

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

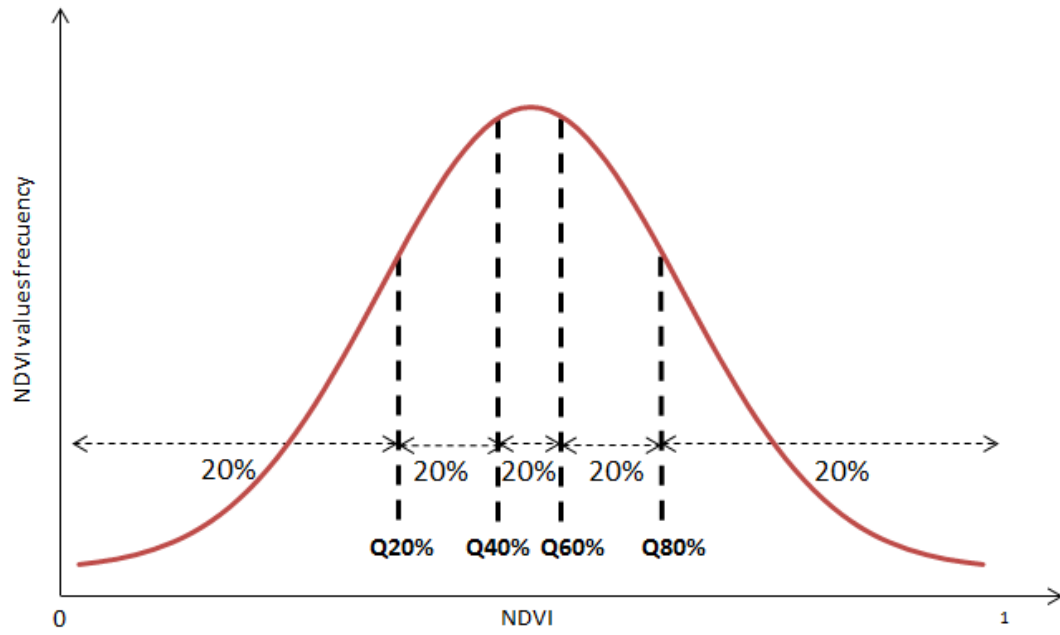
720

721

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

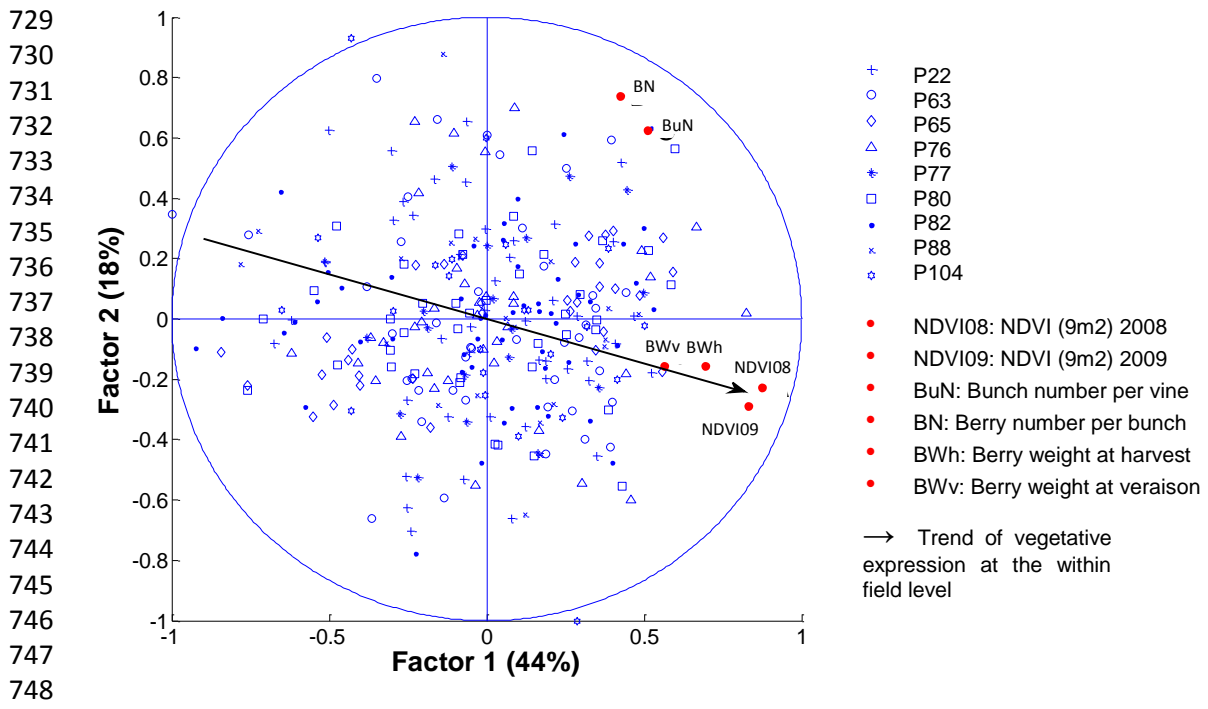
Field (Id)	RM %	TM %	MM %
P22	24.90	25.15	25.97
P63	26.88	25.43	25.61
P65	35.03	15.42	20.48
P76	34.80	30.69	28.63
P77	33.51	26.68	25.61
P80	32.85	27.17	25.77
P82	27.37	25.12	22.72
P88	11.96	9.66	8.78
P104	17.84	19.46	19.50

724

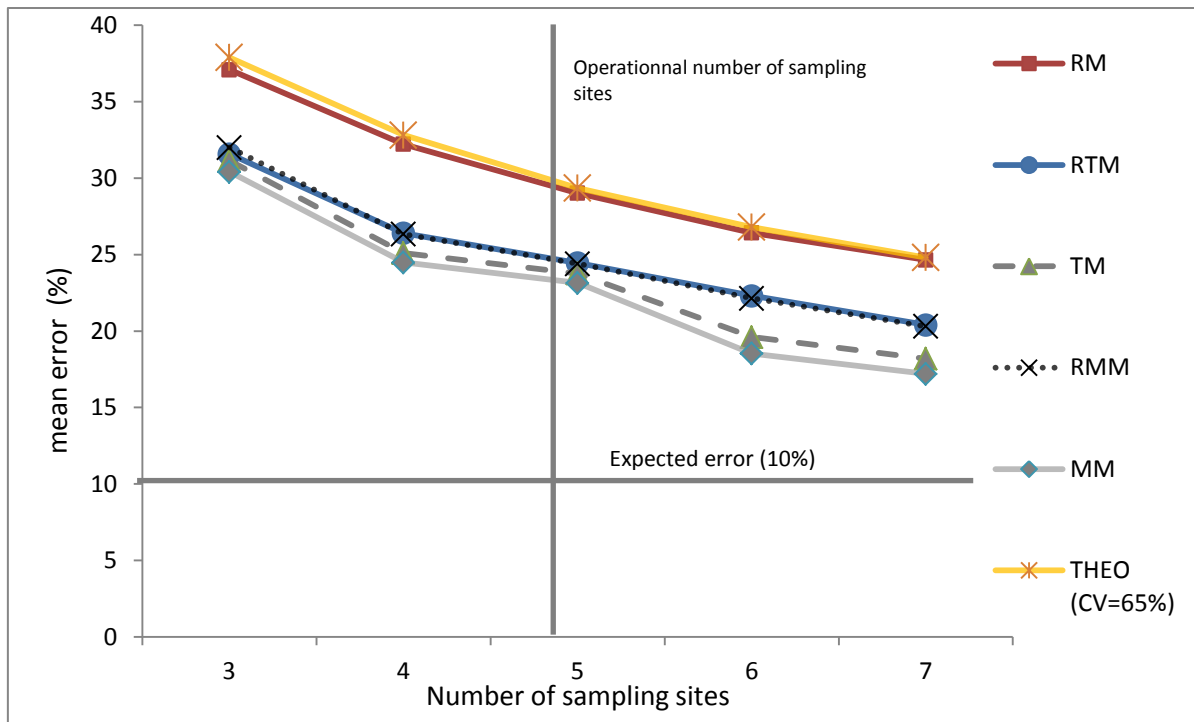


725
726 **Fig. 1** Example of 20% percentile division of NDVI distribution to define intervals used to perform
727 targeted sampling for 5 sites.

728



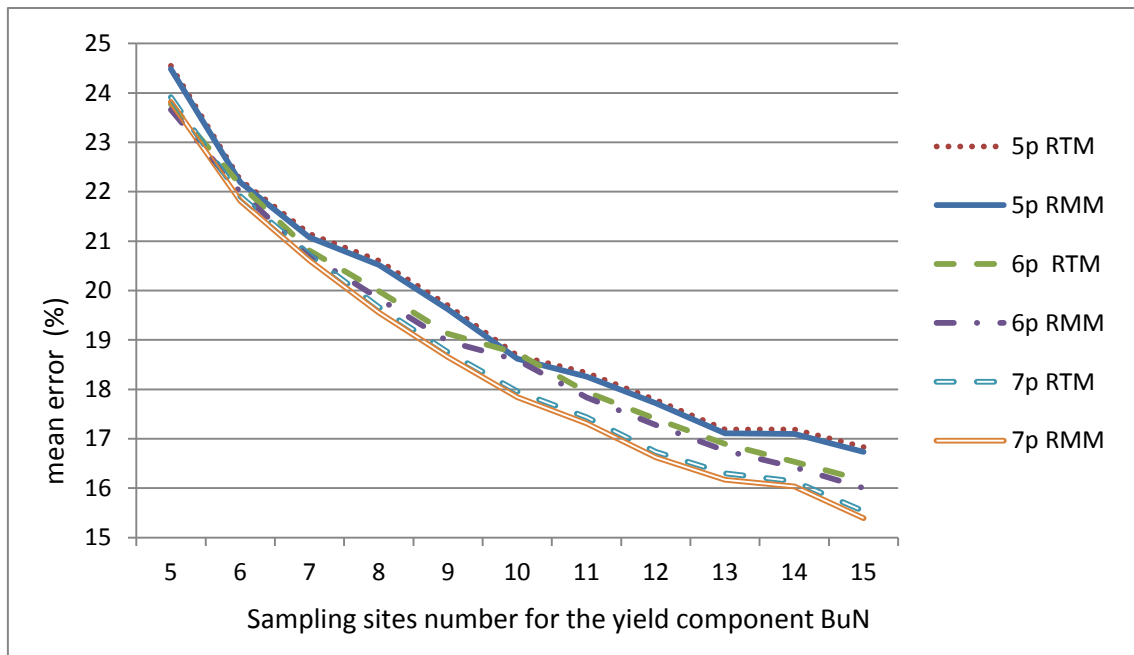
749 **Fig 2**, Scattered plot and correlation coefficients of the principal component analysis (first 2 Factors) with
 750 data centered and reduced according to a field by field basis.
 751



752

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

758



759

760 **Fig 4.** Mean error (%) from the nine fields of the database obtained for RTM and RMM as a function of
761 the number of sampling sites for BuN (Bunch Number) and for BuW (bunch weight) independently. 5p,
762 6p and 7p correspond respectively to 5, 6 and 7 sampling sites for BuW. RTM: Random-Target, RMM:
763 Random-Model.