

This is a pre print version of the following article:

Minimisation of instrumental noise in the acquisition of FT-NIR spectra of bread wheat using experimental design and signal processing techniques / Foca, Giorgia; Ferrari, Carlotta; N., Sinelli; M., Mariotti; M., Lucisano; R., Caramanico; Ulrici, Alessandro. - In: ANALYTICAL AND BIOANALYTICAL CHEMISTRY. - ISSN 1618-2642. - STAMPA. - 399:6(2011), pp. 1965-1973. [10.1007/s00216-010-4431-z]

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

15/05/2026 23:54



Minimisation of instrumental noise in the acquisition of FT-NIR spectra of bread wheat using experimental design and signal processing techniques

Journal:	<i>Analytical and Bioanalytical Chemistry</i>
Manuscript ID:	ABC-01329-2010
Type of Paper:	Original Paper
Date Submitted by the Author:	20-Jul-2010
Complete List of Authors:	Foca, Giorgia; Universita di Modena e Reggio Emilia, Dipartimento di Scienze Agrarie e degli Alimenti Ferrari, Carlotta; Universita di Modena e Reggio Emilia, Dipartimento di Scienze Agrarie e degli Alimenti Sinelli, Nicoletta; Universita di Milano, Dipartimento di Scienze e Tecnologie Agrarie e Microbiologiche Mariotti, Manuela; Universita di Milano, Dipartimento di Scienze e Tecnologie Agrarie e Microbiologiche Lucisano, Mara; Universita di Milano, Dipartimento di Scienze e Tecnologie Agrarie e Microbiologiche Caramanico, Rosita; CRA, Unità di Ricerca per la Selezione dei Cereali e la Valorizzazione delle varietà vegetali Ulrici, Alessandro; Universita di Modena e Reggio Emilia, Dipartimento di Scienze Agrarie e degli Alimenti
Keywords:	Chemometrics / Statistics, Agriculture, Foods / Beverages, IR spectroscopy / Raman spectroscopy, Spectroscopy/Instrumentation, Software

Minimisation of instrumental noise in the acquisition of FT-NIR spectra of bread wheat using experimental design and signal processing techniques

G. Foca^a, C. Ferrari^a, N. Sinelli^b, M. Mariotti^b, M. Lucisano^b, R. Caramanico^c, A. Ulrici^{a*}

^a Dipartimento di Scienze Agrarie e degli Alimenti, Università degli Studi di Modena e Reggio Emilia, Padiglione Besta, via Amendola 2, 42100 Reggio Emilia, Italy.

^b Dipartimento di Scienze e Tecnologie Agrarie e Microbiologiche, Università di Milano, via Celoria 2, 20133 Milano, Italy.

^c CRA - Unità di Ricerca per la Selezione dei Cereali e la Valorizzazione delle varietà vegetali, Via Forlani 3, 26866 S. Angelo Lodigiano (Lodi), Italy.

* *alessandro.ulrici@unimore.it*

Abstract

Spectral resolution (R) and number of repeated scans (S) have a significant effect on the S/N ratio of FT-NIR spectra, but the optimal values of these two parameters have to be determined empirically for a specific problem, considering separately both the nature of the analysed matrix and the specific instrumental setup. To this aim, the instrumental noise of replicated FT-NIR spectra of wheat samples was modelled as a function of R and S by means of the Doehlert design. The noise amounts in correspondence to different experimental conditions were estimated by analysing the variance spectra derived from replicate measurements with two different signal processing tools, Savitzky-Golay filtering and Fast Wavelet Transform, in order to separate the “pure” instrumental noise from other variability sources, which are essentially connected to sample inhomogeneity. Results confirmed that R and S values leading to minimum instrumental noise can vary considerably depending on the type of analysed food matrix and on the different instrumental setups, and helped in the selection of the optimal measuring conditions for the subsequent acquisition of a wide spectral dataset.

Keywords - FT-NIR, Doehlert design, noise removal, Fast Wavelet Transform, Savitsky-Golay filtering

1
2
3 33
4
5
6
7 34

Introduction

8
9 35 Over recent years, we have worked on the development of classification procedures to predict
10 36 the technological quality of bread wheat as a whole, based on NIR spectra measured on flour
11 37 [1-3]. The results obtained until now, however, highlighted the complexity inherent to this
12 38 kind of application. In fact, many pieces of information reflecting different wheat quality-
13 39 related aspects have to be extracted from the NIR spectra, and the contribution of some of
14 40 these relevant features can be very low and hidden into a relatively large amount of not
15 41 pertinent information and noise. Thus, before proceeding to the sampling of a wide dataset to
16 42 be analysed by means of different multivariate classification methods, including the use of
17 43 feature selection procedures, it is extremely important to define objectively the optimal
18 44 instrumental measuring conditions, in a manner to gain spectra as much informative as
19 45 possible. This aim can be reached to a great extent by improving the S/N ratio, also
20 46 considering that the specific nature of the food matrix to be analysed and the particular
21 47 instrumental setup have their own influence.

22 48 It is well known in the literature [4] that significant variations on the S/N ratio of FT-NIR
23 49 spectra can be induced by varying the spectral resolution (R) and on the number of repeated
24 50 scans (S). The spectral resolution is connected with the frequency of sampling of the
25 51 instrumental signal along its spectral range: the higher is the number of sampled points, the
26 52 higher the resolution of the signal. However, the increase of the number of sampled points is a
27 53 double-edge weapon, since together with the signal resolution, also the sampled instrumental
28 54 noise (which generally has high frequency) increases.

29 55 On the other hand, the NIR spectrum acquired by means of a FT spectrophotometer is
30 56 obtained by averaging a number of successive scans of the sample. As a consequence, by
31 57 increasing the number of repeated scans, S, it is generally known that the instrumental noise
32 58 decreases of a factor equal to the square root of S. Unfortunately, the choice of the possible
33 59 highest number of repeated scans is not always convenient, since the time necessary for the
34 60 analysis increases too, hence a proper compromise has to be found.

35 61 Usually, the optimal values of R and S are defined on the basis of the experience and
36 62 sensitivity of the analyst who, for each specific problem, chooses the best values of these
37 63 parameters in an empirical way, fixing them *a priori* or by a fast trial-and-error procedure,
38 64 which is generally based only on the visual inspection of the resulting spectra.

1
2
3 65 At variance, in the present research work, we have decided to define objectively the optimal
4
5 66 values of the acquisition parameters, by extracting the amount of instrumental noise with
6
7 67 proper signal processing techniques and then by modelling it as a function of R and S. In
8
9 68 addition, we have also considered different operative conditions, i.e., different physical forms
10
11 69 under which the sample (bread wheat) has been analysed (i.e., grit, wholemeal flour and
12
13 70 white flour), and different instrumental techniques (i.e., different instrument accessories) for
14
15 71 signal acquisition. In order to find the conditions corresponding to the minimum value of
16
17 72 noise, this was modelled as a function of the factors R and S using the Doehlert experimental
18
19 73 design.

20
21 74 The noise values were estimated by analysing the variance spectra obtained from replicate
22
23 75 measurements. Since different variability sources contributed to the variance spectra,
24
25 76 including not only the high-frequency instrumental noise (whose minimisation was the aim of
26
27 77 the present work), but also lower-frequency components, which are essentially due to sample
28
29 78 inhomogeneities, two different signal processing techniques, i.e., Savitzky-Golay smoothing
30
31 79 filters and Fast Wavelet Transform were used to isolate and to extract the “pure” instrumental
32
33 80 noise component from the “whole” variance spectra. Then, multilinear regression was applied
34
35 81 to model the instrumental noise; this allowed to build response surfaces that were used to
36
37 82 locate, within the experimental domain, the operating conditions (R and S values) leading to
38
39 83 the minimum noise values in correspondence to each kind of matrix analysed by each
40
41 84 instrumental apparatus.

42 85

43 86 Material and methods

44 87 *Sampling and instrumentation*

45
46 88 Two different bread wheat types were analysed in this work, which were selected among 153
47
48 89 different types within the Italian bread wheat production of the harvesting year 2007. The
49
50 90 samples were selected on the basis of their intermediate values for the main technological
51
52 91 parameters that are usually considered to define wheat quality [5]. Selection was performed
53
54 92 with the help of a PCA model calculated on the matrix composed by 322 replicate
55
56 93 measurements of the 153 different bread wheat samples, for which the values of the following
57
58 94 seven parameters have been measured: Hectolitre weight, Falling Number, Protein content,
59
60 95 Alveographic W, Alveographic P/L, Stability and sodium-dodecyl sulphate (SDS)
96
97 96 sedimentation volume. In particular, samples selection was done considering the Q vs. T² plot
97
97 of a 4 PCs model (85.53% explained variance) calculated on autoscaled data (Fig. 1). Two

1
2
3 98 samples were chosen as representative of the whole dataset, belonging to two different quality
4
5 99 categories, within the 95% confidence limits of the PCA model: one sample showing low Q
6
7 100 and intermediate T^2 values, and the other one showing low T^2 and intermediate Q values.

8
9 101 The samples were submitted to NIR analysis under three different physical forms: grit,
10
11 102 wholemeal flour and white flour, in order to consider all the different forms under which the
12
13 103 wheat is usually analysed by NIR spectroscopy in the cereal manufacturing process. The
14
15 104 analyses were performed in two independent laboratories equipped with two
16
17 105 spectrophotometers from the same manufacturer and model, i.e., two Bruker MPA Multi
18
19 106 Purpose FT-NIR Analyzer spectrophotometers. Each laboratory acquired the spectra on a
20
21 107 different sample, in order to verify the possibility to obtain reliable results even on samples
22
23 108 having different compositional characteristics. The two FT-NIR instruments were equipped
24
25 109 both with three sampling tools: Fiber Optic probe (FO), Integrating Sphere (IS) and
26
27 110 Transmission Unit (TU). In order to lower the variability due to possible differences in the
28
29 111 optical path passed by the radiation, the weight of each aliquot of sample under examination
30
31 112 was set at a constant value, between 10.0 and 80.0 g, depending on the physical form of the
32
33 113 analyzed sample and on the sampling tool used for spectral measurements.

34

35 114 *Doehlert experimental design*

36
37 115 Among other possible experimental designs, the Doehlert design [6] was chosen since it
38
39 116 allows to consider different levels for each factor (in particular, in our case five levels for S
40
41 117 and three levels for R in an hexagonally-shaped domain) leading to the possibility to model
42
43 118 the response function with a quadratic surface, based on only seven different experimental
44
45 119 conditions. For our two factors S (number of scans) and R (resolution), the postulated model
46
47 120 is:
48
49 121

50

$$51 122 \text{IN} = b_0 + b_1 S + b_2 R + b_{11} S^2 + b_{22} R^2 + b_{12} S R \quad (\text{eq. 1})$$

52

53 123 where the response IN stands for Instrumental Noise.

54
55 124 The five levels for S were equally spaced in the interval $20 \div 200$ scans (Fig. 2). As it was
56
57 125 previously stated, the higher is the number of scans, the lower the noise; however, we have
58
59 126 chosen 200 scans as the highest level of S, in order to maintain an acceptable acquisition time
60
127 by a practical point of view. The three levels of R were varied over 2, 4 and 8 cm^{-1} for all the
128
129 128 experiments, except for grit and wholemeal flour analysed by TU instrumental setup, for
130
131 129 which our previous experience suggested to use higher values (4, 8 and 16 cm^{-1}). Samples

1
2
3 132 measured as grit were not analysed by FO because it was not possible to guarantee an
4
5 133 acceptable reliability of the obtained measurements, due to the interface between sample and
6
7 134 the fiber surface, whose diameter (3 mm) is too small compared to the grit dimensions and
8
9 135 shape (the analysed sample should be in direct contact with the FO surface). Therefore,
10
11 136 globally the Doehlert design was applied to eight different sample form/instrumental setup
12
13 137 combinations (both TU and IS on grit, wholemeal and white flour and FO on wholemeal and
14
15 138 white flour).

16 139 The NIR spectra were acquired for all the seven experimental conditions (*points*) of the
17
18 140 design, replicated three times in random order. Each spectrum acquisition was repeated three
19
20 141 times by each laboratory (each time with different sample aliquots) for a total of nine spectra
21
22 142 for each point and for each sample form/instrumental setup combination, leading globally to
23
24 143 1008 NIR spectra.

25 144

26 145 *Noise estimate*

27
28 146 The estimation of the noise was obtained starting from the variance spectra of the 3 repeated
29
30 147 measurements. These signals highlight the presence of two kinds of “noise” contributions: a
31
32 148 high-frequency component (ascribable to the “pure” instrumental noise) and a low-frequency
33
34 149 one, which is essentially due to the differences in the specific nature of the sample (e.g.
35
36 150 variability of particle size or moisture content among different sample aliquots). In Figure 3
37
38 151 some examples of variance spectra are reported.

39 152 The differences due to the sample inhomogeneity cannot be eliminated even by properly
40
41 153 adjusting the instrumental conditions by means of the experimental design, hence the low
42
43 154 frequency variance contribution has not to be considered in the noise minimisation process.
44
45 155 High frequency and low frequency variances were separated by means of two different signal
46
47 156 processing techniques: Savitsky-Golay filtering (SG) and Fast Wavelet Transform (FWT).

48 157 The Savitzky-Golay filtering [7, 8] is a widespread smoothing processing, where a polynomial
49
50 158 function of a given order is fitted to the signal in a sliding window having an odd number
51
52 159 width. The polynomial function is moved point by point across the signal and, each time, the
53
54 160 experimentally measured value of the midpoint of the window is substituted with the
55
56 161 corresponding value calculated with the polynomial. By varying the order of the polynomial
57
58 162 and the width of the window it is possible to modulate the extent of the smoothing, taking
59
60 163 care to not distort the signal shape.

164 164 The FWT [9-11] operates on an individual discrete signal by convolving it with a couple of
165
165 165 filters (called the High-pass and Low-pass decomposition wavelet filters, i.e., Hi_D and

1
2
3 166 Low_D, respectively), thus splitting it into two orthogonal subspaces, called vectors of
4 approximations (retaining only the low frequency content of the signal) and of details (which
5 167
6 approximates the high frequency content), respectively. The two wavelet filters are orthogonal,
7 168
8 since frequencies “kept” by the Lo_D are not “kept” by the Hi_D (and *vice versa*), and
9 169
10 complementary, since the original signal can be perfectly reconstructed from the
11 170
12 approximations and details vectors, by applying the proper couple wavelet reconstruction
13 171
14 filters (Lo_R and Hi_R). Then, the decomposition procedure can be repeated to a further
15 172
16 decomposition level, applying the same two filters to the approximations vector. In this way,
17 173
18 sharp and coarse properties of the signal are captured and disjointed in different sub spaces
19 174
20 (vectors of wavelet coefficients, called ‘blocks’), at different levels of resolution. In the
21 175
22 present application, the approximations vector at the lowest level of decomposition has been
23 176
24 kept as the low frequency component of the variance signal, i.e. as the component that is not
25 177
26 related to instrumental noise, but to the variations in the specific nature of the sample.
27 178

28 179 A proper Matlab function, namely *noise_estimate*, was written to subtract the low variance
29 180
30 contribution from the variance spectrum – using both SG filtering and FWT – to obtain the
31 181
32 instrumental noise spectrum. The suitability of the chosen conditions for SG filtering (i.e.,
33 182
34 size of the filtering window and order of the fitting polynomial) and for FWT decomposition
35 183
36 (i.e., type of wavelet and decomposition level) with respect to the convenient noise extraction
37 184
38 was verified by visual inspection for each combination sample form/instrumental setup. The
39 185
40 output figure of *noise_estimate* applied to the NIR spectra measured in the experimental
41 186
42 conditions corresponding to the Doehlert point (0.5; -0.866) of white flour sample analyzed
43 187
44 by the integrating sphere is reported in Figure 4. Figure 4a is an example of a poor extraction
45 188
46 of instrumental noise from the variance spectrum, due to the choice of a not optimal
47 189
48 decomposition level in the wavelet domain, while Figure 4b is an example of good separation
49 190
50 of instrumental noise from low frequency variations. The subtraction of the low variance
51 191
52 contribution from the variance spectra was repeatedly done by means of *noise_estimate* until
53 192
54 satisfactory results (i.e., optimal filtering conditions) have been reached.

55 193 The preliminary tests permitted to set the polynomial order of the SG interpolating function
56 194
57 equal to 2, as the best compromise between eliminating the low frequency variance without
58 195
59 losing spectral information. As for the wavelet functions for FWT decomposition, different
60 196
197 wavelets from the *symlet* wavelet family (*sym*) were tested, since these often resulted the most
198 198
199 appropriate ones to analyse the NIR spectral shapes [3]. Finally the *sym8* wavelet was chosen,
200 199
201 which is a nearly symmetric wavelet whose filters are 16 points long [12].

1
2
3 199 In order to maintain the range of the interpolating function approximately constant with
4
5 200 respect to the wavenumber values of the spectra, the number of points for SG filter varied on
6
7 201 the basis of the R level. For the same reason, also the maximum decomposition level of FWT
8
9 202 varied with the R level (Tab. 1).

10 203 Once *noise_estimate* was punctually applied to the different variance spectra sets, then a
11
12 204 further Matlab function was developed (*noise_estimate_fast*), which was employed to collect
13
14 205 a comprehensive summary of the signal processing conditions found for each variance spectra
15
16 206 set and to calculate the corresponding pure instrumental noise value. In fact,
17
18 207 *noise_estimate_fast.m* provided an output array containing the best SG and FWT conditions, a
19
20 208 figure reporting into three subplots the original signals, the low variance contribution signals
21
22 209 and the noise signals, and finally the mean of the absolute values of noise, that was considered
23
24 210 as the response variable to be minimised.

25 211 Both functions were developed in Matlab language ver. 7.0, using some of the functions
26
27 212 available in the Wavelet Toolbox ver. 3.0.

28 213

31 214 Results and discussion

32
33 215 The first visual inspection of noise behaviour was done by means of 3D scatter plots of the
34
35 216 response values obtained on the 3 replicates of the 7 Doehlert points. For some of the sample
36
37 217 form/instrumental setup combinations (e.g., for grit analysed by integrating sphere, which is
38
39 218 reported Figure 5) consistent differences among replicates were observed. The reason for this
40
41 219 behaviour is probably ascribable to the fact that noise is a completely stochastic variable, so
42
43 220 that its values are not as much reproducible as those of "ordinary" responses, where the
44
45 221 deterministic component is generally much greater than the stochastic one. Thus, even though
46
47 222 the mean value of the three replicate measurements for each experimental condition accounts
48
49 223 for the instrumental noise values that are encountered on the average, it could be more
50
51 224 interesting to estimate the maximum instrumental error that can affect the measurements. In
52
53 225 fact, since noise has to be minimised, the maximum value can be considered as a pessimistic
54
55 226 estimate of the noise amount, therefore allowing to choose the most convenient experimental
56
57 227 conditions even in the worst cases. Based on this considerations, it was decided to create the
58
59 228 response surfaces using both the average value and the maximum value of the three replicates.
60
229 As it is shown in Figure 6, the shapes of the response surfaces can slightly change when using
230 average or maximum values for multilinear regression. However, in all the considered
231 conditions (sample form/instrumental setup) it was observed that the overall behaviour and

1
2
3 232 (most important) the location of the minimum values of instrumental noise within the
4
5 233 analysed experimental domains are coincident when using average or maximum values .
6
7 234 The optimal measuring conditions, i.e. those leading to the minimum instrumental noise
8
9 235 values, were estimated through the computation of the multilinear regression models
10
11 236 described in eq. 1 for all the sample form/instrumental setup combinations, considering
12
13 237 separately the data measured in the two independent laboratories, thus modelling the noise
14
15 238 estimates obtained from the two different instruments of the two laboratories by separate
16
17 239 response surfaces. Then, the eight response surfaces obtained for each sample
18
19 240 form/instrumental setup combination (2 for SG-FWT \times 2 for mean-max \times 2 for lab1-lab2)
20
21 241 were compared to pick out, as the best overall compromise, the experimental condition
22
23 242 (Doehlert point) leading to the minimum of noise (Figure 7). Even if the two laboratories
24
25 243 analysed different samples, the trends of the corresponding response surfaces – obtained on
26
27 244 the same sample form measured with the same instrumental setup – generally converged. In
28
29 245 Table 2, the optimal experimental conditions for spectra acquisition are reported, in terms of
30
31 246 Resolution and number of Scans, obtained for all the sample form/instrumental setup
32
33 247 combinations. Despite the fact that it is usually assumed that the S/N ratio increases with
34
35 248 decreasing the resolution and with increasing the number of scans, it has to be noticed that the
36
37 249 results do not ever confirm this rule. In general, the combinations between the high R level
38
39 250 and the medium-high S level or between the medium R level and the high S level worked as
40
41 251 best.

42 252

43 253 Conclusions

44 254 This work confirmed the convenience of applying experimental design techniques to find the
45
46 255 optimal measuring conditions in FT-NIR analysis. In fact, the obtained results demonstrate
47
48 256 that the amount of noise into the spectra depends not only on the resolution and on the
49
50 257 number of scans with which the signals are acquired, but also on the physical nature of the
51
52 258 sample and on the adopted instrumental setup. In particular, the choice of the optimal
53
54 259 combination of all these variables to produce low noise spectra is not so straightforward and
55
56 260 can not be known *a priori* in a precise way.

57 261 In the end, even if the application of the Doehlert design required some time, it significantly
58
59 262 helped in the selection of the measuring conditions with each instrumental setup, for the
60
263 subsequent acquisition of a spectral dataset of about 15.000 NIR spectra of bread wheat under
264 the forms of grit, wholemeal and white flour. This kind of application is quite novel in the FT-

1
2
3 265 NIR spectroscopy field, where chemometrics is heavily involved, but usually only in the
4
5 266 phase of spectra elaboration, when the time and effort requested to correct or recover rough
6
7 267 spectra are often excessive with respect to the time and effort spent for the spectra acquisition
8
9 268 phase.

10 269

13 270 Acknowledgements

15
16 271 Regione Lombardia (Italy) is gratefully acknowledged for financial support (research project
17
18 272 AUTOSTOC). Dr. Maria Corbellini (CRA – S. Angelo Lodigiano) and Dr. Paolo Belloni
19
20 273 (Bruker Optrics Italia) are also gratefully acknowledged for technical support.

22 274

25 275 References

- 27
28 276 [1] Cocchi M, Corbellini M, Foca G, Lucisano M, Pagani MA, Tassi L, Ulrici A (2005)
29
30 277 Anal. Chim. Acta 544: 100-107.
- 31 278 [2] Foca G, Ulrici A, Corbellini M, Pagani MA, Lucisano M, Franchini GC, Tassi L
32
33 279 (2007) J. Sci. Food Agr. 87(5): 839-846.
- 34
35 280 [3] Foca G, Cocchi M, Li Vigni M, Caramanico R, Corbellini M, Ulrici A (2009)
36
37 281 Chemom. Intell. Lab. Syst. 99: 91-100.
- 38 282 [4] Furukawa Y (2002) In: Near Infrared Spectroscopy – Principles, Instruments,
39
40 283 Applications, Siesler HW, Ozaki Y, Kawata S, Eise HM (eds) Wiley-VCH,
41
42 284 Weinheim, p. 92.
- 43
44 285 [5] Borasio E, Proceedings of the GranoItalia Symposium, Bologna, pp. 59-61 (1997).
- 45 286 [6] Ferreira SLC, dos Santos WNL, Quintella CM, Neto BB, Bosque Sendra JM (2004)
46
47 287 Talanta 63: 1061-1067.
- 48
49 288 [7] Brown SD (2006) In: Practical guide to chemometrics – 2nd edition, Gemperline P
50
51 289 (ed) CRC Press, Boca Raton, p. 403
- 52
53 290 [8] Savitzky A, Golay MJE (1964) Anal. Chem. 36: 1627-1639.
- 54 291 [9] Mallet Y, de Vel O and Coomans D (2000) In: Wavelets in chemistry, Walczak B
55
56 292 (ed), Elsevier, Amsterdam, p. 74.
- 57
58 293 [10] Mallat S, A Wavelet Tour of Signal Processing - 2nd Ed., Academic Press, New York,
59
60 294 US (1999).
- 295 [11] Barclay VJ, Bonner RF, Hamilton IP (1997) Anal. Chem. 69: 78-90.

- 1
2
3 296 [12] Misiti M, Misiti Y, Oppenheim G, Poggi JM (2004) Wavelet Toolbox, The
4
5 297 MathWorks, Natick, US.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1
2
3
4 298 Figures captions
5

6
7 299 **Fig. 1.** Q vs. T² plot from the PCA model based on the experimental parameters measured on
8
9 300 the bread wheat samples. The selected samples are represented with black squares.

10 301 **Fig. 2.** R and S level values used in the Doehlert design for the different sample
11
12 302 form/instrumental setup combinations. The numbers reported in the grey circles correspond to
13
14 303 the coordinates of Doehlert points in the experimental matrix.

15
16 304 **Fig. 3.** Examples of some variance spectra, where the high and the low frequency
17
18 305 contributions (sharp and coarse variations, respectively) can be observed.

19 306 **Fig. 4.** Example of an output figure from *noise_estimate*, which was obtained by applying to
20
21 307 the variance spectrum the FWT processing with a sym8 wavelet at decomposition levels 3 (a)
22
23 308 and 6 (b).

24 309 **Fig. 5.** 3D scatter plot of noise values of the three replicate experimental measurements
25
26 310 acquired for each one of the seven experimental conditions of the Doehlert design, obtained
27
28 311 for grit analysed by IS.

29
30 312 **Fig. 6.** 3D response surfaces for grit analysed by IS by laboratory 1 computed on average (a)
31
32 313 and maximum (b) noise values.

33 314 **Fig. 7.** 2D response surfaces obtained on grit analysed by IS for all the laboratory/sample
34
35 315 form/instrumental setup combinations.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

316

R (cm ⁻¹)	TU			IS			FO	
	Grit	Whole meal	White flour	Grit	Whole meal	White flour	Whole meal	White flour
2			(107)/(5)	(107)/(6)	(107)/(6)	(107)/(6)	(53)/(5)	(53)/(5)
4	(53)/(4)	(53)/(4)	(53)/(4)	(53)/(5)	(53)/(5)	(53)/(5)	(27)/(4)	(27)/(4)
8	(27)/(3)	(27)/(3)	(27)/(3)	(27)/(4)	(27)/(4)	(27)/(4)	(13)/(3)	(13)/(3)
16	(13)/(2)	(13)/(2)						

317

318 **Table 1.** (Number of points for SG filtering)/(maximum FWT decomposition level), as they
 319 vary with the R level.

For Peer Review

320

		Optimal R (cm ⁻¹)	Optimal S (# scans)
Transmission Unit	Grit	16	155
	Wholemeal	8	200
	White flour	4	200
Integrating Sphere	Grit	8	155
	Wholemeal	8	155
	White flour	8	155
Fiber Optic	Wholemeal	4	200
	White flour	8	155

321

322

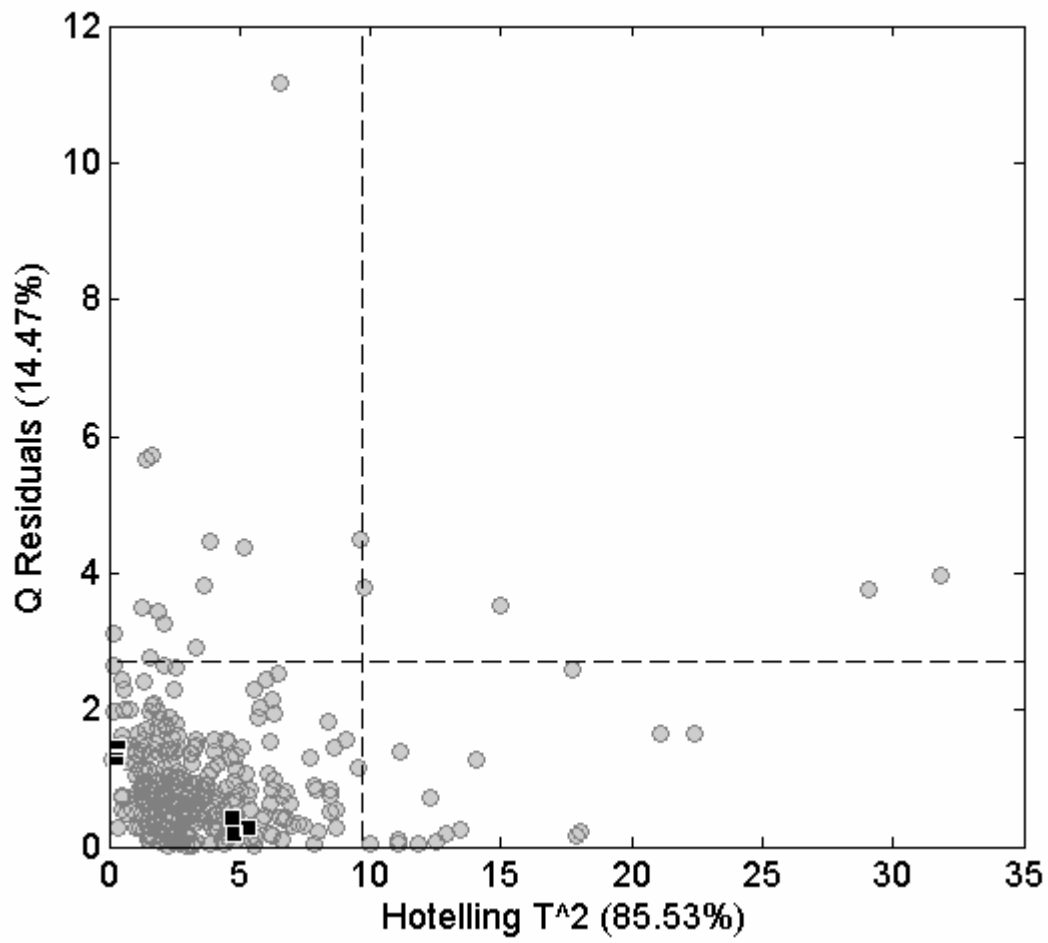
323

Table 2. Optimal R and S levels for spectral acquisition obtained by using the experimental design.

324

325

326



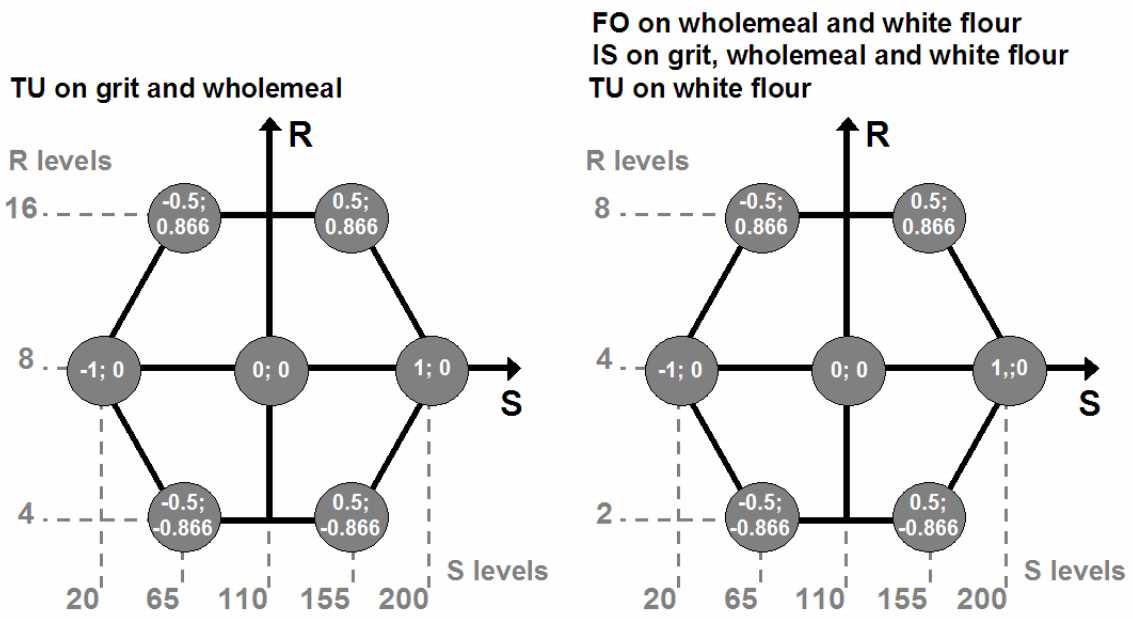
327

328

Figure 1

view

329



330

331

Figure 2

Peer Review

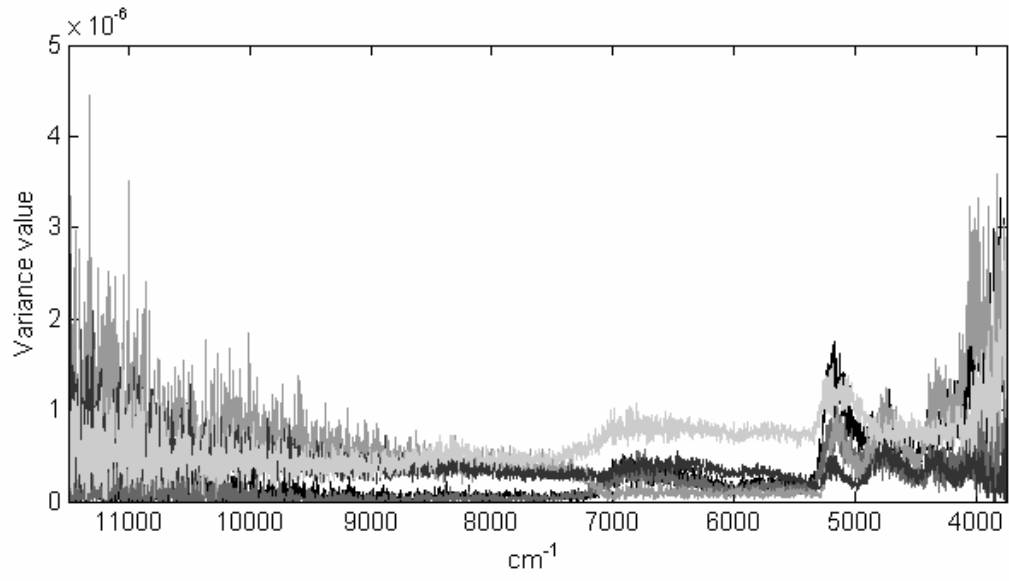


Figure 3

Peer Review

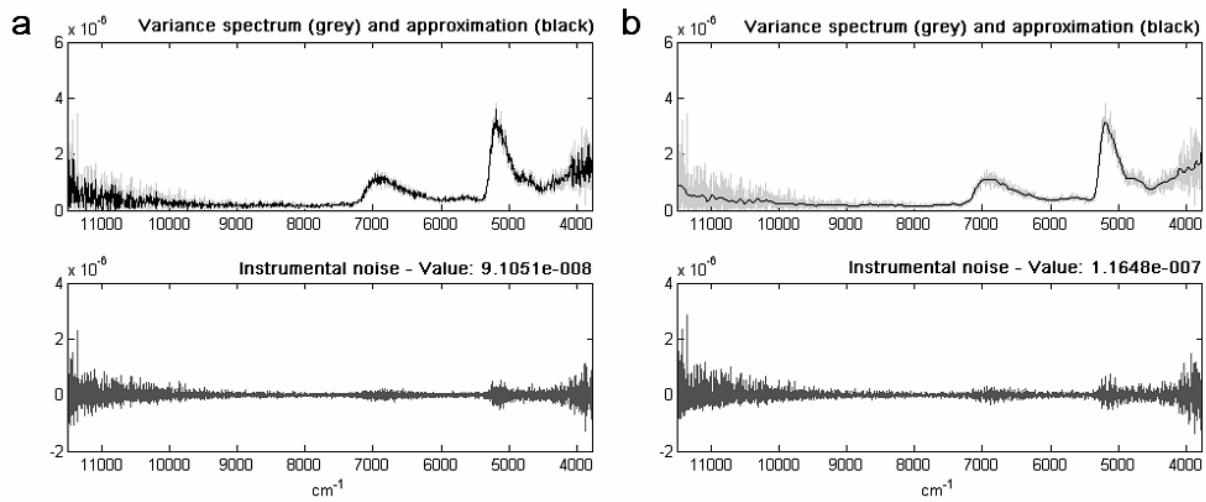


Figure 4

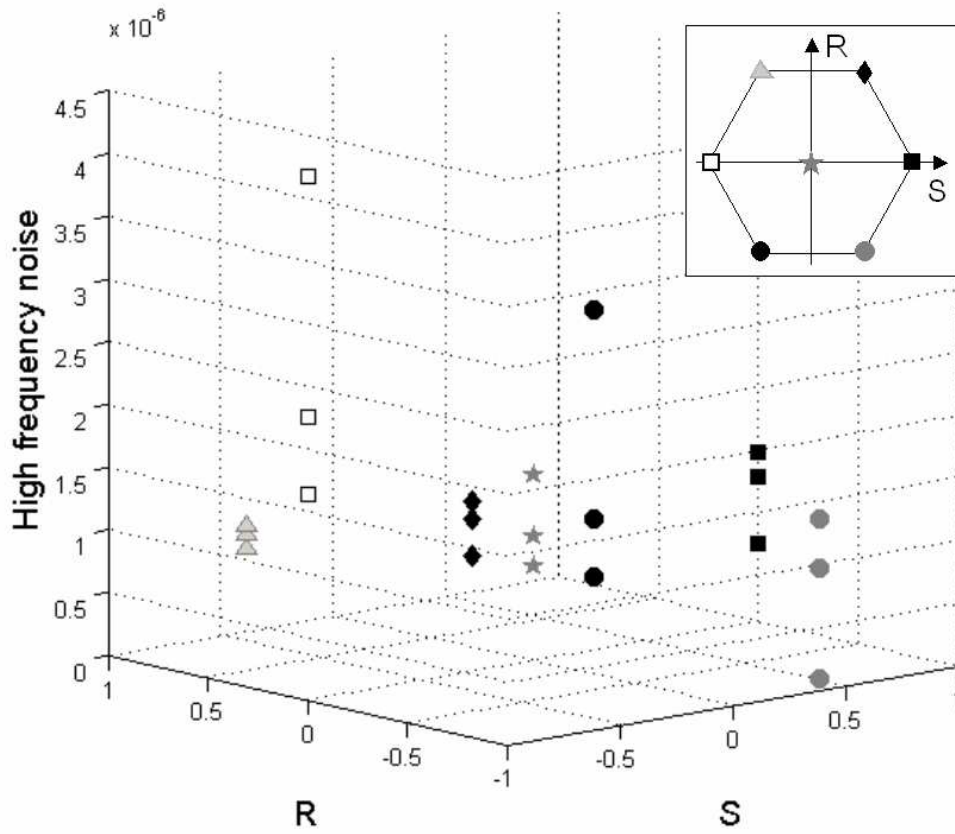
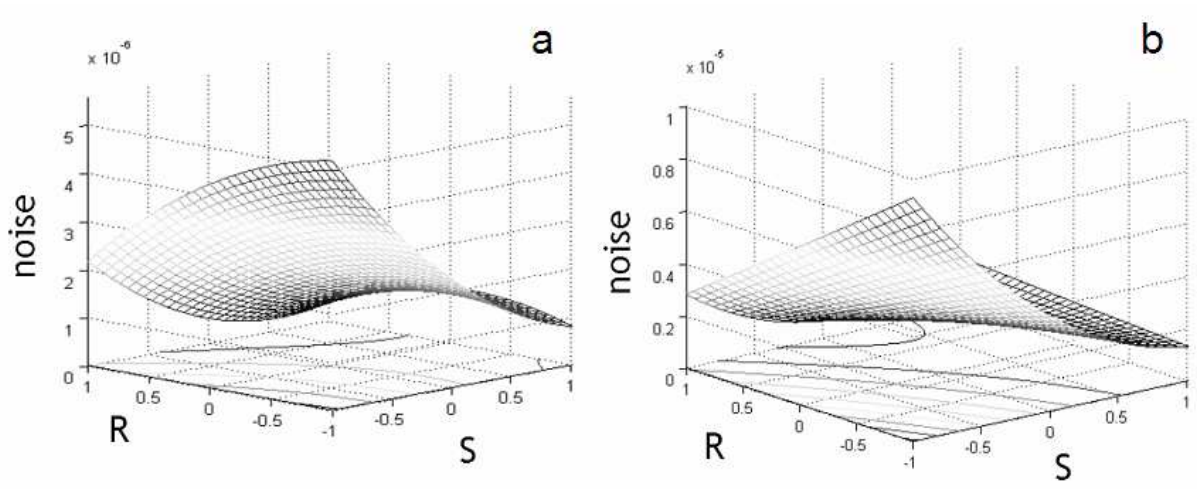


Figure 5



340
341
342

Figure 6

For Peer Review

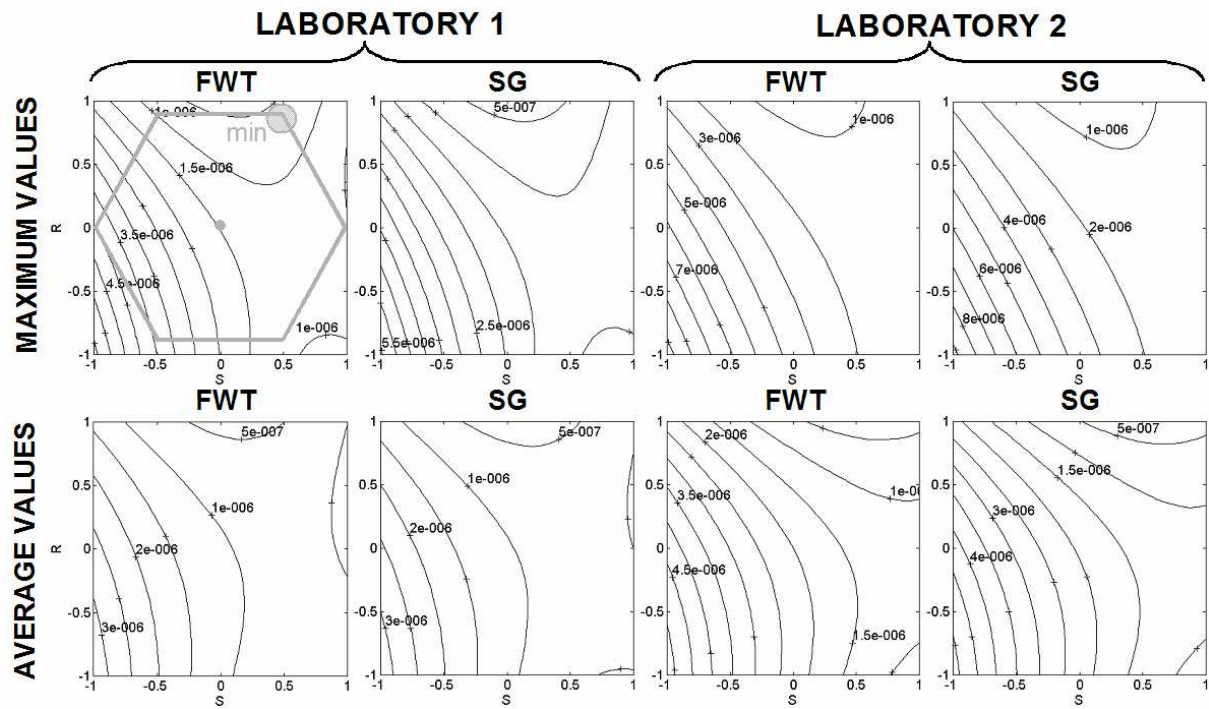


Figure 7