



[biblio.ugent.be](http://biblio.ugent.be)

The UGent Institutional Repository is the electronic archiving and dissemination platform for all UGent research publications. Ghent University has implemented a mandate stipulating that all academic publications of UGent researchers should be deposited and archived in this repository. Except for items where current copyright restrictions apply, these papers are available in Open Access.

This item is the archived peer-reviewed author-version of:

Title: A batch modeling approach to monitor a freeze-drying process using in-line Raman spectroscopy

Authors: Sarraguca M.C., De Beer T., Vervaet C., Remon J.P., Lopes J.A.

In: Talanta, 83(1), 130-138 (2010)

Optional: link to the article

**To refer to or to cite this work, please use the citation to the published version:**

**Authors (year). Title. *journal Volume(Issue)* page-page. Doi** [10.1016/j.talanta.2010.08.051](https://doi.org/10.1016/j.talanta.2010.08.051)



34 **Abstract**

35 Freeze-drying or lyophilisation is a batch wise industrial process used to remove water  
36 from solutions, hence stabilizing the solutes for distribution and storage. The objective  
37 of the present work was to outline a batch modelling approach to monitor a freeze-  
38 drying process in-line and in real-time using Raman spectroscopy. A 5% (w/v) D-  
39 mannitol solution was freeze-dried in this study as model. The monitoring of a freeze-  
40 drying process using Raman spectroscopy allows following the product behaviour and  
41 some process evolution aspects by detecting the changes of the solutes and solvent  
42 occurring during the process. Herewith, real-time solid-state characterization of the final  
43 product is also possible.

44 The timely spectroscopic measurements allowed the differentiation between batches  
45 operated in normal process conditions and batches having deviations from the normal  
46 trajectory. Two strategies were employed to develop batch models: Partial least squares  
47 (PLS) using the unfolded data and parallel factor analysis (PARAFAC). It was shown  
48 that both strategies were able to developed batch models using in-line Raman  
49 spectroscopy, allowing to monitor the evolution in real-time of new batches. However,  
50 the computational effort required to develop the PLS model and to evaluate new batches  
51 using this model is significant lower compared to the PARAFAC model. Moreover,  
52 PLS scores in the time mode can be computed for new batches, while using PARAFAC  
53 only the batch mode scores can be determined for new batches.

54

55 Keywords: Freeze-drying, D-mannitol, Batch process monitoring, Process analytical  
56 technology, Multi-way analysis, PLS, PARAFAC.

57

## 58 **Introduction**

59

60 Freeze-drying, also called lyophilisation, is a three stages drying process used to convert  
61 solutions of (heat-)labile materials into solids of sufficient stability for distribution and  
62 storage [1]. The initial stage is a freezing step in which water is converted into ice, and  
63 the solutes are crystallized or transformed into an amorphous system. The shelf  
64 temperature in the freezing state is set to ensure that the product is cooled bellow the  
65 glass transition temperature. The second stage is a primary drying step in which the ice  
66 is sublimated under vacuum. The temperature during the primary drying is increased  
67 (but kept under the collapse temperature) to supply energy for ice sublimation. The  
68 process ends with a secondary drying step in which all the unfrozen water is removed  
69 by desorption and/or in which hydrate water is removed [2]. Freeze-drying is a widely  
70 used process for the preservation of microorganisms, food items, biological products  
71 and pharmaceuticals [3-6]. In the pharmaceutical industry, the process provides  
72 improved stability, and/or desired physicochemical properties, such as enhance  
73 dissolution rates and bioavailability [6, 7].

74 Real-time monitoring of freeze-drying processes is essential to reduce costs and to  
75 improve process knowledge and efficiency. Freeze-drying cycles are in many cases set  
76 up by trial and error, herewith only focussing on the final product quality [8]. During  
77 the last decades, several methods based on product temperature and pressure  
78 measurements were developed to monitor freeze-drying processes [7-9]. However, these  
79 methods do not allow the in-line monitoring of all critical process aspects (e.g., product  
80 behaviour).

81 In recent years, several methods based on the concept of process analytical technology  
82 (PAT) have emerged in the pharmaceutical industry, the majority of them using

83 spectroscopic techniques [10]. Spectroscopic tools have several advantages over other  
84 analytical methods such as high performance liquid chromatography (HPLC): they can  
85 be non-invasive and non-destructive and can be used in-line hence providing real-time  
86 information. The application of near infrared spectroscopy (NIRS) and Raman  
87 spectroscopy does not only supply information about the chemical and physical  
88 properties of the final product (e.g., physical state, polymorphism), but also about the  
89 chemical and physical changes occurring over time. In previous studies, Raman  
90 spectroscopy and NIRS [2, 3, 9, 11, 12] were evaluated as potential tools for the in-line  
91 and real-time monitoring of freeze-drying processes. Using these methods, the  
92 determination of some process stage end-points as well as the chemical/physical  
93 characterization of the product were achieved. These studies were mainly focussed on  
94 process improvement and the detection of process occurrences (e.g. physical state  
95 transformations) over time. However, and since freeze-drying is a batch wise process,  
96 also the batch-to-batch variation has to be addressed. The differentiation between good  
97 and bad batches in the early process phase is a major concern in the pharmaceutical  
98 industry since batch-to-batch variability can be unpredictable [13]. The unpredictability  
99 of batch variation can lead to quality problems in the final product (e.g. variability in  
100 residual moisture content).

101 The aim of this study was not to focus on the critical freeze-drying process aspects  
102 which can be detected using in-line Raman spectroscopy, as this was done previously  
103 [9, 12]. The objective of this work is to show how freeze-drying process fingerprints  
104 obtained by continuous in-line Raman measurements can be used to model reference  
105 freeze-drying processes (i.e., development of batch models) allowing to evaluate in real-  
106 time whether future new batches are proceeding as the desired reference processes. A  
107 5% (w/v) D-mannitol solution was used as model to freeze-dry [14].

108 Multi-way models have been recognized as useful tools for monitoring batch data since  
109 they improve the process understanding and summarize the process behaviour in a batch  
110 wise manner. Multi-way principal component analysis (MPCA) and multi-way partial  
111 least squares (MPLS) were used to monitor batch wise processes, such as for example,  
112 fluid bed granulation [13, 15]. Other multi-way methods such as parallel factor analysis  
113 (PARAFAC) and Tucker 3 were also used to monitor batches processes, such as wheat  
114 growing experiments using NIRS and polymerization processes [16, 17]. In this study  
115 PLS and PARAFAC were the employed batch modelling strategies. In this particular  
116 case, PLS and not MPLS was used in the work. The data was unfolded and regular PLS  
117 was performed on the data, it is important to refer that regular PLS and MPLS  
118 algorithms are quite distinct [18].

119 A set of nominal batches obtained in normal operational conditions (NOC) were used to  
120 develop the batch (calibration) models. New batches were projected onto these models  
121 to detect any deviation from normal batch trajectories.

122

## 123 **Data Analysis**

124

125 The data obtained from the freeze drying processes were organized in a three-way array  
126  $\underline{\mathbf{X}}$  ( $I \times J \times K$ ) with I batches, J variables (number of spectral variables) and K time points.

127 The PLS was performed using SIMCA P+ 12.01 (Umetrics AB, Umeå, Sweden).

128 PARAFAC modelling was performed using PLS toolbox version 3.5 in Matlab, version  
129 6.5 release 13 (MathWorks, Natick, MA, USA).

130

## 131 **PLS**

132

133 To develop the PLS model, unfolding of the three-way array was done preserving the  
134 variable direction, creating a new mode combining the batch and time mode ( $M = IK$ ).  
135 The row  $m$  of the matrix  $\mathbf{X}$  has the spectrum corresponding to time point  $k$  for the batch  
136  $i$ . The dependent variable vector,  $\mathbf{Y}$ , used for the partial least squares (PLS) regression,  
137 has a length equal to  $M$  and represent batch duration. By performing PLS regression  
138 using time as the dependent variable, the individual observations can be evaluated over  
139 time and batch maturity can be predicted. Moreover, by preserving the variable  
140 direction, the typical tendency of a batch being operated in NOC, can be followed. The  
141 number of PLS components was set by cross-validation using the approach described by  
142 Krzanowski [19] To monitor new batches, and compare their trajectory with the NOC  
143 batches, control charts are developed. After PLS modelling, a score matrix is obtained  
144 ( $M \times T$ ), in which  $T$  is the number of latent variables used to fit the PLS model. To  
145 create the control charts, the scores matrices are rearranged to produce "T" matrices, one  
146 for each latent variable from the PLS model. Row-wise, each of those matrices have  
147 dimension ( $I \times K$ ). From each of these matrices, a vector is estimated ( $1 \times K$ ) with a  
148 standard deviation ( $\sigma$ ) for the corresponding latent variable over the  $K$  time points. The  
149 control limits are set in as  $\pm 3 \times \sigma$ . The essence of this re-ordering principle is that, for  
150 each component of the PLS model, an average trajectory with upper and lower control  
151 limits is obtained. When projecting the new batches into the model the normal  
152 development of these batches can be followed, as well as any deviation from it.

153 Another control chart is the residuals chart showing the unmodelled variation, for each  
154 batch. A good batch should evolve in the same way as the reference batches and be  
155 below a critical value set at  $+3 \times \sigma$ , in which the standard deviation is calculated for the  
156 average of the residuals from the calibration batches [20].

157

158 **PARAFAC**

159 PARAFAC is a method for modelling three-way or higher order data. PARAFAC is a  
160 decomposition method that can be compared to the bilinear principal component  
161 analysis (PCA) [21]. In the case of a three-way data set the decomposition is performed  
162 in three components as can be seen in Equation 1.

163

164 
$$x_{i,j,k} = \sum a_{if} b_{jf} c_{kf} + e_{ijk} \quad (1)$$

165

166 In Equation 1,  $x_{ijk}$  is an element of the three-way array  $\underline{\mathbf{X}}$ ; and  $e_{ijk}$  is an element of the  
167 three-way  $\underline{\mathbf{E}}$  of residues. Three ways or modes (a, b and c) are obtained with indices  $i =$   
168  $1, \dots, I$ ,  $j = 1, \dots, J$ , and  $k = 1, \dots, K$ . These indices constitute the loading matrixes  $\mathbf{A}$ ,  $\mathbf{B}$  and  
169  $\mathbf{C}$ . The index  $f$  is the number of PARAFAC components. In matrix notation the  
170 PARAFAC model can be written as,

171 
$$\mathbf{X}_k = \mathbf{A} \mathbf{D}_k \mathbf{B}^T + \mathbf{E}_k, k = 1, \dots, K \quad (2)$$

172 where  $\mathbf{D}_k$  is a diagonal matrix holding the  $k$  row of  $\mathbf{C}$  in its diagonal.

173 The determination of the number of components is one of the major difficulties of a  
174 PARAFAC model. Resampling techniques such as cross-validation or residuals  
175 histograms are some of the techniques that can be use to determine the number of  
176 PARAFAC components. However all of them have some disadvantages such as heavy  
177 computations involved or the difficulty to determine with assurance the optimum  
178 number of components. To overcome the disadvantages, a single diagnostic analysis,  
179 called core consistency, that gives clear differences for different models was created.  
180 The core consistency is always less or equal to 100%, a good trilinear model can be said  
181 to have a core consistency above 90%. Low values of core consistency indicates that



182 elements outside of the super diagonal are significantly different of zero, that the model  
183 is not trilinear and a model such Tucker should be used.[22].

184 After the calibration of the PARAFAC model, new batches can be projected onto the  
185 model. However, only the loadings of mode 1 (in this case the batch mode) are obtained  
186 for the new batches. This fact creates a problem because; no indication on their  
187 behaviour over time is obtained. The residuals statistics can be used to obtain a first  
188 impression on the new batches. If the sum of squares of the residuals values are higher  
189 than the value for the 95% confidence limit, it can be concluded that the predicted batch  
190 had some kind of problem during the process. Nonetheless, no information can be  
191 retrieved regarding where in time the problem occurred. Batch control charts can be  
192 constructed using the Hotellings and residuals statistics by performing the following  
193 procedure [17].

- 194 • A number PARAFAC models were constructed by cutting the batch duration in  
195 expanding time periods, like  $0-K/n$ ,  $0-2K/n$ , ...,  $0-K$  time points, in which  $n$  is the  
196 number of time periods.
- 197 • The prediction batches were projected onto each constructed model.
- 198 • For each model and for each prediction batches the Hotelling and the residuals  
199 sum of squares values were determined.

200 Batch control charts can be constructed using the Hotelling and the residuals statistics  
201 for the different models constructed and setting as control limits the value  
202 correspondent to the 95% confidence level. The Hotelling statistic gives an indication  
203 on batch variation, or in other words, assesses the statistical significance of the  
204 difference between batches. The residuals statistic is an indication how well each batch  
205 conforms to the model. Consequently, these parameters can be used as indicators of  
206 process consistency.

207

## 208 **Experimental**

209

### 210 **Materials**

211

212 D-mannitol (further abbreviated as mannitol) is one of the most used excipients in  
213 pharmaceutical freeze-drying. It is generally employed as a bulking agent, crystallizing  
214 during lyophilisation, hence providing structural support to the final product.

215 In this study, 5% (w/v) mannitol solutions (3 ml) were used as model for the freeze-  
216 drying process.

217

### 218 **Batches**

219

220 To develop the batch models, six NOC batches were used (process conditions in Table  
221 1). The batch models were evaluated by running three additional batches having  
222 deviating operational conditions, (see Table 1 for process conditions). A NOC batch,  
223 not used in the calibration set, was also used to evaluate the developed batch models.

224

### 225 **Process Description**

226

227 The equipment used was an Amsco FINN-AQUA GT4 (GEA, Köln, Germany) freeze-  
228 dryer. For Raman process monitoring, a Raman probe was built into the freeze-drier  
229 chamber. The probe was placed above a vial, hence allowing to monitor the formulation  
230 top surface without contact between product and probe. The optical fiber cable of the

231 Raman probe was connected through a gap made in the freeze drier chamber door. [2, 9,  
232 12].

### 233 **Raman Spectroscopy**

234

235 A RamanRxn1 spectrometer (Kaiser Optical Systems, Ann Arbor, MI) equipped with an  
236 air-cooled CCD detector (black-illuminated deep depletion design) was used in  
237 combination with a fiber optic non-contact probe to monitor the freeze-drying  
238 processes. As the Raman probe was directly placed above the product to freeze dry, the  
239 glass vial did not interfere with the Raman signal. The laser wavelength was 785 nm  
240 (NIR diode laser). All spectra were collected at a resolution of  $4\text{ cm}^{-1}$  using a laser  
241 power of 400 mW. Data collection and transfer were automated using the  
242 HoloGRAMS<sup>TM</sup> data collection software. A spectrum was collected every two minute  
243 during lyophilisation with thirty-second exposures.

244

### 245 **Results and discussion**

246

247 Two different Raman spectral regions were used to monitor the freeze-drying processes.  
248 Ice produces a Raman signal in the region between  $150\text{ cm}^{-1}$  and  $250\text{ cm}^{-1}$  while  
249 mannitol produces signals between  $1000\text{ cm}^{-1}$  and  $1170\text{ cm}^{-1}$  [9]. Furthermore, the  
250 different polymorphic forms of mannitol can be distinguished in this spectral region.  
251 These two spectral regions were used together (in total 901 spectral variables) to  
252 develop the batch models. During the freezing step, ice formation can be detected by the  
253 appearance of the ice peak at  $215\text{ cm}^{-1}$  (Figure 1a). Shortly after the water solidification,  
254 mannitol starts to crystallize (Figure 1b). During primary drying, the ice is sublimated.  
255 The disappearance of the peak at  $215\text{ cm}^{-1}$  can be seen during this process step (Figure

256 1c). Furthermore, the peaks corresponding to mannitol do not show any visible changes,  
257 indicating that no transformations related to mannitol occurred during primary drying  
258 (Figure 1d). The temperature was raised for the secondary drying step to remove the  
259 hydrate water (i.e., to convert mannitol hemi-hydrate to an anhydrous form [9]). The  
260 Raman signals corresponding to mannitol hemi-hydrate disappear or decrease in  
261 intensity and new Raman peaks corresponding to anhydrous mannitol ( $\alpha$  form) appear at  
262  $1030\text{ cm}^{-1}$  and  $1130\text{ cm}^{-1}$  (Figure 1f).

263 Raman spectra were collected every two minutes to decrease the computational effort,  
264 resulting in a total of 550 spectra per batch. Consequently, the calibration data is  
265 arranged in a three-way array  $\underline{\mathbf{X}}(I \times J \times K)$  of  $I = 6$  batches,  $J = 901$  spectral variables and  
266  $K = 550$  time points.

267

### 268 **Batch modelling – PLS**

269

270 The unfolding of the three-way array by preserving the variable direction resulted in a  
271 matrix  $\mathbf{X}$  ( $M \times J$ ) with  $M=3300$  (6 batches with 550 time points) and  $J=901$ . Before PLS  
272 analysis, the spectra were pre-processed using standard normal variate (SNV) and mean  
273 centred. A PLS model was developed and cross-validation was performed, resulting in a  
274 two component model (cumulative variance of  $\mathbf{X}$  of 0.89 (Table 2)). A three component  
275 model didn't significantly improved the explained variance (increase of 0.019)..  
276 Consequently, a two component model was chosen.

277 Analysing the PLS model loadings (Figure 2) it can be seen the spectral variations  
278 described by each PLS component. The loadings correspondent to the first component  
279 are related to the transformations occurred during the freezing and primary drying  
280 stages. The section of the loadings that correspond to the ice signal (Figure 2a) shows

281 the variations that occurred in the band at  $215\text{ cm}^{-1}$ . Comparing the Figure 2b with  
282 Figure 1a and 1d, is clear that the loadings describe the mannitol transformations taking  
283 place during the freezing and primary drying stages. The loadings correspondent to the  
284 second PLS component (Figure 2 c and d) are related to the transformations occurred  
285 during the secondary drying. In the case of the ice signal range (Figure 2c), the main  
286 feature is the appearance of a band at  $240\text{ cm}^{-1}$  (Figure 1e). The loadings correspondent  
287 to the mannitol transformation range (Figure 2d) relate to the appearance and  
288 disappearance of bands during the secondary drying (polymorphic transformation).The  
289 changes occurring during the process can also be detected analysing the scores  
290 evolution over time (Figure 3). Only the scores of calibration batch 2 are depicted for  
291 visualization clarity. The increase of the PLS 1 component scores after 102 minutes (I)  
292 is related to the beginning of the water to ice conversion. Mannitol crystallization can be  
293 detected by the increase of the first PLS component scores and the decrease of the  
294 second PLS component scores at 122 minutes (II). The start of the primary drying (A) is  
295 not followed by any significant changes in the scores. An increase of the scores for both  
296 PLS components at 194 process minutes (III) is attributed to ice sublimation during the  
297 primary drying. The beginning of the secondary drying is accompanied by an increase  
298 of the scores for both PLS components (B). The reason that the secondary drying can be  
299 detected, opposed to the primary drying, is the substantial increase in the temperature  
300 ( $60^{\circ}\text{C}$ ) in the secondary drying stage. The polymorphic transformation between hemi-  
301 hydrate and  $\alpha$  mannitol at minute 1038 (V) can be seen in the increase of the second  
302 PLS component scores.

303 After development of the calibration model, the spectra from the prediction batches  
304 were projected onto the model. To evaluate these new batches, batch control charts  
305 based on the scores (Figure 4) and residuals (Figure 5) from the calibration batches were

306 constructed. The scores from the PLS second component were chosen to construct the  
307 control charts (Figure 4) because they show that information during freeze and primary  
308 stages as the first component scores, but the information associated with the secondary  
309 drying is more visible in the second component as can be seen in Figure 2 d and Figure  
310 3.

311 Prediction batch 1 was a nominal batch, i.e., a batch produced under NOC. However,  
312 when its trajectory was compared to the calibration batches trajectories, significant  
313 deviations could be detected. In the score batch control chart (Figure 4) prediction batch  
314 1 is out of control (above the superior limit) until minute 86, indicating some problem in  
315 the process conditions or spectra acquisition during that time. Looking to the spectra of  
316 prediction batch 1 obtained during the first 86 process minutes (not shown), some  
317 abnormalities could be detected. Since the batch trajectory was within the limits the rest  
318 of the process, it can be concluded that the initial deviation was related with problems  
319 associated with the spectra acquisition. The same conclusion can be drawn by analysing  
320 the residuals control chart.

321 Prediction batches 2 to 4 were subjected to different process conditions (Tables 1). For  
322 prediction batch 2, the primary drying step was longer compared to the NOC batches  
323 and the shelf temperature during the secondary drying stage was first set at 25 °C during  
324 first 100 secondary drying minutes instead of 40 °C. In the score control chart,  
325 corresponding to the second PLS component (Figure 4) a deviation occurred in the end  
326 of the process indicating the difference in behaviour of this batch during the secondary  
327 drying. In the residuals control chart (Figure 5) this batch also deviates from the model  
328 at the end of the process where the difference in process conditions compared to the  
329 NOC batches is more significant.

330 Prediction batch 3 and 4, have very different process conditions compared to the  
331 reference batches (Table 1): the primary drying starts later, the set shelf temperature  
332 during primary drying is higher and no secondary drying was done. It is expected that  
333 these two batches are out of trajectory during the entire process time. In fact, the score  
334 control charts (Figure 4) show that the trajectory is completely different. The residuals  
335 control chart (Figure 5) shows that almost the complete trajectory of prediction batches  
336 3 and 4 is above the imposed limit.

337

### 338 **Batch modelling – PARAFAC**

339

340 To develop the PARAFAC calibration model, a three-way array  $\underline{\mathbf{X}}$  ( $I \times J \times K$ ) with  $I = 6$   
341 (number of batches),  $J = 901$  (number of spectral variables) and  $K = 550$  (number of  
342 time points) was used. The spectra were pre-processed using SNV and centred before  
343 PARAFAC analysis. The number of PARAFAC components was chosen based on the  
344 core consistency criterion. For a number of PARAFAC components between 1 and 4  
345 the core consistency and the percentage of explained variance was determined (Table  
346 3). A model with 3 components was chosen with a core consistency value of 94.3% and  
347 an explained variance of 44.0%.

348 The loadings from the third mode (time) of the PARAFAC model can be seen as the  
349 average batch trajectory for the calibration batches (Figure 6). The changes occurring  
350 during the process can be seen in the three component loadings. The water to ice  
351 conversion around 100 minutes (I) followed by the mannitol crystallization (II) are  
352 clearly seen in the three component loadings. The ice sublimation occurring around  
353 minute 200 (III) can be seen in the first component. The polymorphic transformation

354 occurred at the end of the process (IV) can be followed by an increase in the first  
355 component loading and a decrease in the second component loading.

356 The beginning of the primary drying (A) stage can be seen in the loadings of the first  
357 and second component. A decrease in the third component loading around 845 minutes  
358 (B) is an indication of the beginning of the secondary drying.

359 The loadings for the second mode (spectral variables) and for the three PARAFAC  
360 components are shown in Figure 7. By comparing the loadings with the spectra  
361 presented in Figure 1 it can be concluded (as was the case of the PLS loadings) that they  
362 are related to the spectral changes occurred during the three process stages. The  
363 loadings correspondent to the first, second and third component are associated with the  
364 freezing, primary and secondary drying stages, respectively.

365 After the calibration model was developed, the four prediction batches were projected  
366 onto the model. The residuals sum-of-squares for the prediction batches (not shown)  
367 confirm that batches 3 and 4 are clearly different from the NOC ones. Batches 1 and 2  
368 are above the 99% confidence limit but below the 95% confidence limit. These statistics  
369 provide an indication of problematic batches. However, no indication is given in which  
370 part of the process trajectory the problem occurred. For this reason, the procedure  
371 explained in section PARAFAC was done to get an indication on the process phases  
372 during which the problems occurred (Figure 8 and 9). A total of 22 models were  
373 constructed with expanding time periods.

374 The Hotelling control chart (Figure 8) shows prediction batch 1 with abnormal  
375 behaviour in the first 50 minutes of the process, which is in accordance with what has  
376 already been explained in section Batch modelling - PLS model). However, the  
377 residuals (Figure 9) indicate that this batch is above the control limit until minute 350,  
378 and very near to the control limit the rest of the process. Prediction batch 2 is always



379 within the control limit in the Hotelling control chart. Only the last model for this batch  
380 is above the control limit in the residuals control chart, indicating that the problem in  
381 this batch is in the end of the process as was already discussed above (section Batch  
382 modelling - PLS). Prediction batches 3 and 4 are above the control limit for the first 150  
383 minutes of the process time according to the Hotelling chart. The plot regarding the  
384 residual statistics shows that both batches have a residual value higher than the imposed  
385 control limit.

386

### 387 **Batch monitoring with PLS and PARAFAC**

388

389 The two approaches used to create the batch models gave similar results and  
390 conclusions. The conversion of water into ice, mannitol crystallization, ice sublimation  
391 at the surface and polymorphic transformation (hydrate removal) were clearly detected  
392 by following the scores over time for both methods (Figures 3 and 6). Batch control  
393 charts were constructed and used to evaluate new batches running under normal and non  
394 normal process conditions. Both methods detected that prediction batch 1, thought to be  
395 a nominal batch, deviated from the normal trajectory in the beginning of the process  
396 (Figure 4, 5, 8 and 9). Prediction batch 2 was subjected to different process conditions  
397 (see Table 1). Hence, it was expected to have a different behaviour, in particularly in the  
398 end of the process. The residuals batch control charts (Figure 5 and 9) showed a few  
399 deviations, particularly in the end of the process. For Prediction batch 3 and 4, both  
400 methods considered them out of limit during the first 150 minutes (Figure 4 and 8). The  
401 residuals control charts for both methods (Figure 5 and 9) supported these conclusions.  
402 The batch control charts only give information regarding a process disturbance; no  
403 information is obtained about the cause of the disturbances. A solution to this problem

404 is the use of contribution plots. By using such plots the contribution of each process  
405 variable can be evaluated and control limits can be introduced in the contribution plots.  
406 This procedure allows the unveiling of the process variables that show different  
407 behaviour compared with the NOC batches [23]. The use of contribution plots is an easy  
408 concept when dealing with few process variables, but with spectroscopic data, the use  
409 and analysis of these plots is not straightforward. Firstly, the number of variables is very  
410 high (wavelengths) and secondly, these variables are highly correlated. To construct  
411 contribution plots with spectroscopic data an initial variable reduction should be  
412 performed. This possibility is undoubtedly worth of exploration in a future work.

413

## 414 **Conclusions**

415

416 The objective of this work is to show how freeze-drying process fingerprints obtained  
417 by continuous in-line Raman measurements can be used to model reference freeze-  
418 drying processes (i.e., development of batch models) allowing to evaluate in real-time  
419 whether future new batches are proceeding as the desired reference processes. Two  
420 chemometric batch modelling approaches were used and tested: PLS and PARAFAC.  
421 The main product transformations occurring during the freeze-drying process can be  
422 successfully evaluated during the monitoring of new batches. The PLS and PARAFAC  
423 control charts were able to detect non nominal batches and gave similar results for both  
424 methods. It can hence be concluded that PLS and PARAFAC perform equally well.  
425 However, the computational effort is less for PLS, compared to PARAFAC, which is  
426 important for the real-time evaluation of new batches.  
427 Future work can be performed in order to include contribution plots in the process  
428 monitoring.

429

430

431

432

433

434

435

436

437 **Acknowledgements**

438 Mafalda C. Sarraguça acknowledges the financial support from the Fundação para a

439 Ciência e Tecnologia (FCT), Portugal, (PhD grant, ref: SFRH/BD/32614/2006).

440 **References**

441

442

443 [1] M.J. Pikal, in: J. Swarbrick (Ed.) Encyclopedia of Pharmaceutical Technology,  
444 Marcel Dekker, New York, 2002, pp. 1807-1833.

445 [2] T.R.M. De Beer, M. Wiggenghorn, R. Veillon, C. Debacq, Y. Mayeresse, B. Moreau,  
446 A. Burggraeve, T. Quinten, W. Friess, G. Winter, C. Vervaet, J.P. Remon, W.R.G.  
447 Baeyens, Anal. Chem. 81 (2009) 7639-7649.

448 [3] M. Brulls, S. Folestad, A. Sparen, A. Rasmuson, Pharm. Res. 20 (2003) 494-499.

449 [4] V. Rindler, S. Luneberger, P. Schwindke, I. Heschel, G. Rau, Cryobiology 38 (1999)  
450 2-15.

451 [5] J. Teixeira, Genet. Eng. Biotechnol. News 25 (2005) 52-54.

452 [6] W. Wang, Int. J. Pharm. 203 (2000) 1-60.

453 [7] J.S. Liu, Pharm. Dev. Technol. 11 (2006) 3-28.

454 [8] H. Gieseler, T. Kramer, M.J. Pikal, J. Pharm. Sci. 96 (2007) 3402-3418.

455 [9] T.R.M. De Beer, M. Alleso, F. Goethals, A. Coppens, Y.V. Heyden, H.L. De Diego,  
456 J. Rantanen, F. Verpoort, C. Vervaet, J.P. Remon, W.R.G. Baeyens, Anal. Chem. 79  
457 (2007) 7992-8003.

458 [10] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, J. Pharm.  
459 Biomed. Anal. 44 (2007) 683-700.

460 [11] W.J. Cao, C. Mao, W. Chen, H. Lin, S. Krishnan, N. Cauchon, J. Pharm. Sci. 95  
461 (2006) 2077-2086.

462 [12] T.R.M. De Beer, P. Vercruyssen, A. Burggraeve, T. Quinten, J. Ouyang, X. Zhang,  
463 C. Vervaet, J.P. Remon, W.R.G. Baeyens, J. Pharm. Sci. 98 (2009) 3430-3446.

464 [13] S. Matero, S. Poutiainen, J. Leskinen, S.P. Reinikainen, J. Ketolainen, K. Jarvinen,  
465 A. Poso, *Chemometrics Intell. Lab. Syst.* 96 (2009) 88-93.

466 [14] R. Chakkittakandy, J.A.W.M. Corver, P.C.M. Planken, *J. Pharm. Sci.* 99 (2010)  
467 932-940.

468 [15] T. Kourti, *Journal of Chemom.* 17 (2003) 93-109.

469 [16] P. Geladi, M. Manley, *Journal of Chemom.* 22 (2008) 247-251.

470 [17] D.J. Louwse, A.K. Smilde, *Chem. Eng. Sci.* 55 (2000) 1225-1235.

471 [18] R. Bro, A.K. Smilde, S. de Jong, *Chemometrics Intell. Lab. Syst.* 58 (2001) 3-13.

472 [19] H.T. Eastment, W.J. Krzanowski, *Technometrics* 24 (1982) 73-77.

473 [20] L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikström, S. Wold,  
474 *Multi- and Megavariate Data Analysis Part I + Part II*, Umetrics AB, Umea, 2001.

475 [21] R. Bro, *Chemometrics Intell. Lab. Syst.* 38 (1997) 149-171.

476 [22] R. Bro, H.A.L. Kiers, *Journal of Chemom.* 17 (2003) 274-286.

477 [23] J.A. Westerhuis, S.P. Gurden, A.K. Smilde, *Chemometrics Intell. Lab. Syst.* 51  
478 (2000) 95-114.

479

480

481

482

483

484

485

486

487

488

489 **Figure Captions**

490

491 Figure 1 Raman spectra corresponding to the three process steps for the two studied  
492 spectral ranges. A - ice signal range and B - mannitol signal range.

493

494

495 Figure 2 First and second PLS component loadings for the two studied spectral ranges.  
496 A – ice signal range and B – mannitol signal range.

497

498 Figure 3 Evolution over time of the first and second PLS component scores for  
499 calibration batch 2. I – Ice solidification; II – Mannitol crystallization; III – Ice  
500 sublimation; IV – Polymorphic transformation; A – Beginning of the primary drying; B-  
501 Beginning of the secondary drying.

502

503 Figure 4 Evolution over time of the predicted scores for the second PLS component –  
504 Scores batch control chart.

505

506

507 Figure 5 Evolution over time of the normalized distance correspondent to the PLS  
508 residuals for the prediction batches – Residuals batch control chart.

509

510 Figure 6 PARAFAC model loadings for the third mode (time) for the three components.  
511 I – Ice solidification; II – Mannitol crystallization; III – Ice sublimation; IV –  
512 Polymorphic transformation; A – Beginning of primary drying; B- Beginning of  
513 secondary drying.

514

515

516 Figure 7 Loadings for the second PARAFAC mode (spectral variables) for the three  
517 components and for the two spectral ranges studied. A – ice signal range and B –  
518 mannitol signal range.

519

520 Figure 8 Hotelling  $T^2$  statistics for the prediction batches as a function of the modelled  
521 time intervals for a PARAFAC model with 3 components.

522

523 Figure 9 Residuals statistics for the prediction batches as a function of the modelled  
524 time intervals for a PARAFAC model with 3 components.