

1 **Artificial neural network assisted spectrophotometric**
2 **method for monitoring fructo-oligosaccharides production**

3 **Balázs Erdős · Maarten Grachten · Peter**

4 **Czermak · Zoltán Kovács**

5

6 Received: date / Accepted: date

7 **Abstract** Short-chain fructo-oligosaccharides (FOS) are considered as low-calorie car-
8 bohydrates with prebiotic function. They can be produced from sucrose by fructosyl-
9 transferase activity, resulting in a mixture of saccharides with different chain lengths.
10 Current practice for carbohydrate analysis involves the use of time-costly and off-line
11 chromatographic procedures. This study is dedicated to the development of an arti-
12 ficial neural network (ANN) model for predicting carbohydrate composition from the
13 direct measurement of UV spectra. A total of 182 samples were generated by operating
14 an enzyme membrane reactor (EMR) under both optimal and suboptimal settings.
15 The concentration data determined by HPLC and corresponding absorbance readings
16 were used to train a two-layer feedforward neural network. The optimized network was

Balázs Erdős
Department of Food Engineering, Szent István University, Budapest, Hungary

Maarten Grachten
Department of Computational Perception, Johannes Kepler University, Linz, Austria

Peter Czermak
Institute of Bioprocess Engineering and Pharmaceutical Technology, University of Applied
Sciences Mittelhessen, Giessen, Germany
Department of Chemical Engineering, Kansas State University, Manhattan, Kansas, USA
Fraunhofer Institute IME-Bioresources, Giessen, Germany

Zoltán Kovács
Department of Food Engineering, Szent István University, Budapest, Hungary
Tel.: + 36 1 305 7234, E-mail: kovacs.zoltan@etk.szie.hu

17 then validated by using new observations that were not involved in the training. The
18 model explained 98, 97 and 88 percent of the variation in the composition of the new
19 observations regarding the main components sucrose, kestose and glucose with a mean
20 squared error of prediction of 6.59, 3.40 and 2.81, respectively. The results indicate
21 that the proposed UV-ANN method has a great potential to be used for the real-time
22 monitoring of the bioconversion.

23 **Keywords** fructo-oligosaccharides · enzyme membrane reactor · artificial neural
24 network · UV spectrophotometry · prebiotics

25 1 Introduction

26 Food consumption tendencies show an increasing demand for products with enhanced
27 composition that contribute to health improvements (Falguera et al, 2012). Fructo-
28 oligosaccharides (FOS) are recognized as low-digestible carbohydrates with prebiotic
29 attributes (Rastall, 2010). They may alter the colonic microflora toward a healthier
30 composition, reduce constipation, and decrease the serum cholesterol level while sup-
31 plying small amount of energy (Mutanda et al, 2014). FOS are usually added to dairy
32 products, beverages and desserts (Bali et al, 2013).

33 FOS are fructans with a degree of polymerization up to 10 (Corradini et al, 2013).
34 They naturally occur in biological materials, however, in low concentration (Dominguez
35 et al, 2013). They are commonly manufactured either by the enzymatic hydrolysis of
36 inulin (Nguyen et al, 2011; Sarup et al, 2016) or by transfructosylation from sucrose.
37 In the latter case, short-chain FOS are generated by the activity of fructosyltransferase
38 enzymes. The production processes may involve either whole cells or (partially) purified
39 enzymes, in immobilized or free form (Kovács et al, 2014). The product of the enzymatic
40 catalysis is usually a mixture of saccharides including fructosyl-nystose (GF4), nystose
41 (GF3), kestose (GF2), glucose (G), fructose (F), and non-reacting sucrose (GF). Re-
42 cently, continuous production techniques employing packed bed reactors (Tanriseven
43 and Aslan, 2005; Ghazi et al, 2005; Chen et al, 2014), fluidized bed reactors (Lorenzoni
44 et al, 2015), and enzyme membrane reactors with inert (Ur Rehman et al, 2016) and

45 catalytic membranes (Nishizawa et al, 2000; Hicke et al, 2006) have received consider-
46 able attention. In contrast with a conventional batch process utilizing soluble enzymes
47 in a stirred-tank reactor, these techniques allow the production of an enzyme-free prod-
48 uct in a continuous manner by recovering the biocatalysts.

49 A wide range of analytical techniques have been proposed for the identification and
50 quantification of FOS. Among them, high-performance liquid chromatography (HPLC)
51 is the most prevalent method. Various HPLC setups are available, employing refrac-
52 tive index detectors (RID) (Petkova et al, 2014), evaporative light scattering detectors
53 (ELSD) (Li et al, 2013), pulsed amperometric detectors (PAD) (Corradini et al, 2004),
54 and pulsed electrochemical detection (PED) (Borromei et al, 2009). In combination
55 with HPLC, matrix-assisted laser desorption/ionization time-of-flight mass spectrom-
56 etry (MALDI-TOF-MS) is another highly selective technique for the determination
57 of chain length distribution of FOS and inulin (Borromei et al, 2009). Although thin
58 layer chromatography (TLC) represents an available option, it is known to be laborious
59 and costly due to its chemical and material demand (Petkova and Denev, 2015). Sim-
60 ilarly, gas chromatography (GC) is reported to be unfavorable since it requires time
61 consuming sample preparation (Joye and Hoebregs, 2000). The application of Fourier
62 transform infrared spectroscopy (FT-IR) and proton nuclear magnetic resonance (H
63 NMR) spectroscopy to inulin analysis has also been investigated. FT-IR was used to
64 quantify some crystalline fructans obtained from fermentation broth by precipitation
65 (Grube et al, 2002), and H NMR was successfully applied to monitor acidic hydrolysis
66 of inulin conducted in deuterium oxide (Barclay et al, 2012).

67 In industrial practice, continuous processes require a strict supervision in order to
68 ensure a controlled and safe plant operation. HPLC, the standard procedure for saccha-
69 rides analysis, is a sensitive and reproducible method, but it is time-costly, laborious,
70 and requires expensive instrumentation and technically trained personnel. Since HPLC
71 is an off-line technique, there is a considerable long delay between sampling from the
72 production line at the facility and receiving information on the carbohydrate compo-

73 sition from the analytics lab. There is a need for robust, real-time methods to detect
74 possible deviations from an optimal process behavior in FOS manufacturing.

75 Spectroscopic analysis in the ultraviolet-visible (UV/vis) spectral region is a com-
76 mon and simple method for acquiring information from sample solutions. Spectropho-
77 tometer setups equipped with optic fiber technology that allow on-line measurement
78 are readily available on the market (Van Den Broeke et al, 2006). The complex nature
79 of spectral data requires adequate data analysis tools, such as multivariate techniques
80 and artificial neural networks (ANN). ANNs are considered as nonlinear statistical
81 modeling tools capable of estimating complex functions. ANN models have found sev-
82 eral applications in food engineering, biotechnology and related fields (e. g. (Mateo
83 et al, 2011; Prabhu and Jayadeep, 2017; Lin et al, 2011; Cámara et al, 2010)).

84 Molecular spectroscopy represents a potential candidate for the rapid detection of
85 carbohydrates. Dias et al (2009) have first shown that UV-spectroscopy combined with
86 chemometrics methods, such as partial least-squares regression (PLSR) and artificial
87 neural networks (ANN), could serve as a fast alternative method for monitoring of
88 galacto-oligosaccharides (GOS) production. They have successfully constructed PLSR
89 and ANN models trained on UV data to predict the concentration of two fractions (lac-
90 tose and total GOS) in samples taken from fermentation broth. The chemical analytics
91 of metabolites, that may appear in the matrix as a result of microbial activity and
92 may interfere with UV readings, were not reported. The proposed method involved the
93 filtration (for cell removal) and dilution of the samples prior to UV analysis of a total of
94 53 samples. Recently, Veloso et al (2012) have published a book chapter on UV spec-
95 trophotometry methods for dietary sugars. Based on some preliminary experiments,
96 the authors point out that such methods might potentially be applied for monitoring
97 the production of FOS in fermentative processes, given that real samples are used in
98 the future to validate this hypothesis. To the best of our knowledge, this hypothesis
99 has never been proven, and our study is the first of its kind to explore the applicability
100 of UV spectroscopy for FOS monitoring in depth.

101 This study investigates artificial neural networks dedicated to the real-time quan-
102 tification of FOS fractions to mono- and disaccharides concentrations based on UV
103 spectral data. First, a large number of samples were generated by operating an en-
104 zyme membrane reactor (EMR). Concentrations of the individual saccharides fractions
105 and UV spectra of the enzyme-free samples were determined by HPLC analysis and
106 UV spectroscopy, respectively. Then, an artificial neural network model was trained
107 to predict the concentration of saccharides in the samples based on their UV spectra.
108 Finally, the optimized network was validated by using new observations, which were
109 not included in the training data.

110 2 Materials and methods

111 2.1 Materials

112 Food-grade sucrose manufactured by 1. MCM Kft. (Kaposvár, Hungary) was purchased
113 from a local food store. Pectinex Ultra SP-L, a commercial enzyme preparation from
114 *Aspergillus aculeatus*, was supplied by Novozyme A/S (Bagsværd, Denmark). The re-
115 action liquor, used in all the experiments, consisted of 40 w/w% food-grade sucrose in
116 deionized water. Its pH was set to 5.7 ± 0.2 with a 0.1 M HCl solution. A spiral-wound
117 ultrafiltration (UF) membrane module with a filtration area of 0.37 m^2 and a cut-off
118 of 20 kDa was purchased from Synder Filtration Inc. (San Francisco, USA).

119 2.2 Enzyme membrane reactor

120 The schematic drawing of the lab-scale EMR employed for generating samples is shown
121 in Fig. 1.

122 The enzymatic catalysis took place in the stirred-tank reactor using soluble en-
123 zymes. The feed pump circulated the enzyme-containing reaction liquid through the
124 UF module. The cross-flow rate of the pump was adjusted to $0.16\text{ m}^3\text{h}^{-1}$ with the
125 frequency drive of the pump. The temperature of the solution was controlled by an ex-
126 ternal bath thermostat and kept at $50\pm 0.5^\circ\text{C}$. The applied UF membrane completely

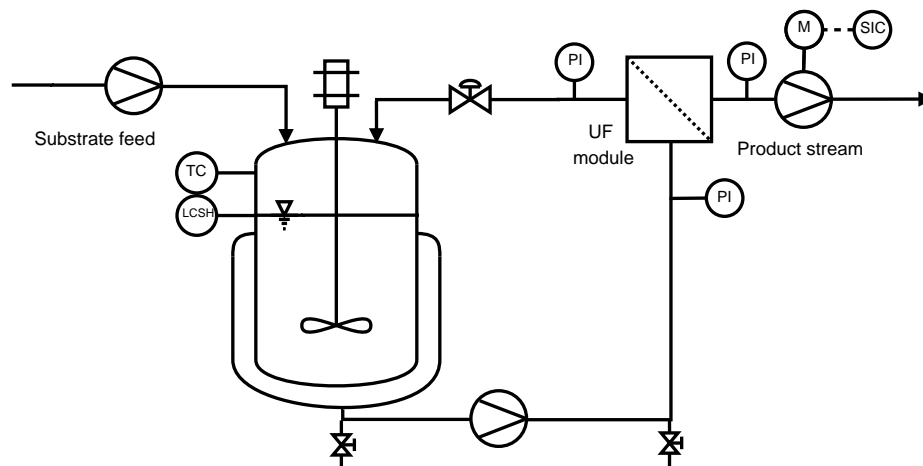


Fig. 1 Schematic drawing of the enzyme membrane reactor

127 rejected the enzymes but allowed the passage of carbohydrates. The retentate-side
 128 pressure was adjusted manually with the control-valve to 1 bar. The residence time
 129 (i. e. the ratio of reactor volume to product flow-rate) was kept constant during
 130 process runs. This was achieved by ensuring that the volumetric flow-rate of the permeate
 131 (i. e. product stream) is equal to that of the feed (i. e. supply stream of the fresh
 132 saccharose solution). For controlling the permeate flow-rate, a pump on the permeate-
 133 line was employed. The rotation of this pump was adjusted to the desired value to
 134 ensure a constant permeate flow-rate. An automatic filling of fresh saccharose solution
 135 to a constant volume was realized by employing a supply pump and a conductive level
 136 switch.

137 2.3 Sample preparation

138 Experimental samples containing mono-, di-, and oligomers in various concentrations
 139 were generated by operating the EMR in continuous fashion. Ten experimental runs
 140 were conducted by varying the enzyme load (i. e. the ratio of the mass of the enzyme
 141 to that of the solution) in the range of 5 to 50 g kg⁻¹ and the residence time in the
 142 range of 1 to 4 h. The EMR was operated until steady-state was reached, typically for

143 6-8 hours. Samples were taken from the permeate at several time points during the
144 process run-time.

145 2.4 Chemical analysis

146 Samples taken from the permeate were diluted with a dilution factor of 5 prior to HPLC
147 analysis. An HPLC system by Thermo Fisher Scientific Inc. (Waltham, MA, USA) was
148 used as reference method for carbohydrate analysis. It included an SCM1000 degasser,
149 a P200 gradient pump, and an AS100 autosampler equipped with a built-in column
150 thermostat. An RNM Carbohydrate 8% Na+ 300x7.8 (Phenomenex Inc., Torrance,
151 USA) analytical column together with a guard column was used at 50°C at 0.2 mL
152 min^{-1} with pre-filtered ($2\mu\text{m}$) DI water as mobile phase. Peak detection and inte-
153 gration was performed by a Shodex R-101 refractive index detector manufactured by
154 Showa Denko Europe GmbH (Munich, Germany), an N2000 Chromatography Data
155 System interface and a N2000 Photographic Data Workstation software package sup-
156 plied by Science Technology Inc. (Hangzhou, China). The concentration of individual
157 carbohydrate fractions was calculated by integrating the respective peak areas and then
158 expressed in relative mass percentage. The relative mass percentage [%] is the ratio
159 of the mass of a substance to the total mass of carbohydrates present in the solution
160 multiplied by 100.

161 The UV spectrophotometry analysis was carried out with a Spectronic GENESYS 5
162 UV-Visible Spectrophotometer manufactured by Thermo Electron Corporation (Waltham,
163 MA, USA). Absorbance of samples at 34 wavelengths, ranging from 201 to 300 nm, at
164 intervals of 3 nm was measured without dilution at room temperature.

165 2.5 Artificial neural networks

166 The purpose of the present study is to predictively model the relationship between the
167 UV absorbance measurements and the carbohydrate concentrations, as obtained by the

168 HPLC measurements, such that given a measurement of UV absorbances over different
 169 wavelengths, we can reliably predict the corresponding carbohydrate composition.

170 We use artificial neural networks (ANNs) for this modeling task. ANNs are non-
 171 linear models suitable for approximating complex scalar or vector-valued functions.
 172 They consist of information processing units that mimic the behavior of neuron cells in
 173 living organisms, in the sense that they output a non-linear response to a weighted sum
 174 of their inputs. These units are arranged into layers, where each unit receives informa-
 175 tion from all the units in the previous layer. The lowest layer of this network—receiving
 176 the input—is called the *input layer* \mathbf{l}_1 , the highest layer—producing the output—is
 177 the *output layer* \mathbf{l}_K , and any intermediate layers are called *hidden layers* $\mathbf{l}_2, \dots, \mathbf{l}_{K-1}$.
 178 Thus, given UV absorbances \mathbf{x}_i , an ANN approximates the relative mass percentage of
 179 the individual carbohydrate fractions \mathbf{y}_i according to the following recursive equations:

$$\hat{\mathbf{y}}_i = \mathbf{l}_K \quad (1)$$

$$\mathbf{l}_k = f_k \left(\mathbf{W}_k^T \mathbf{l}_{k-1} + \mathbf{w}_k \right) \quad (2)$$

$$\mathbf{l}_1 = \mathbf{x}_i \quad (3)$$

182 where $\theta = \{ \mathbf{W}_k, \mathbf{w}_k \mid k \in [2, K] \}$ are the parameters of the model, and f_k is
 183 a function applied elementwise to its arguments, i. e. the non-linear *sigmoid* function
 184 for hidden layers (f_2, \dots, f_{K-1}), and in our case of real-valued output variables, the
 185 *identity* function for f_K .

186 In a procedure referred to as *training*, the model parameters are adjusted based
 187 on a set of data pairs $(\mathbf{x}_i, \mathbf{y}_i)$ ($i \in [1, N]$). More specifically, the optimal parameters $\tilde{\theta}$
 188 are chosen to minimize the *objective function* involving the data, based on the *mean*
 189 *squared error* (MSE) between $\hat{\mathbf{y}}_i$ and \mathbf{y}_i :

$$\tilde{\theta} = \underset{\theta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \|\mathbf{y}_i - \hat{\mathbf{y}}_{i;\theta}\|_2^2 \quad (4)$$

190 where $\|\cdot\|_2$ is the L_2 norm, and $\hat{\mathbf{y}}_{i;\theta}$ is the model approximation of \mathbf{y}_i using
191 parameters θ . The optimal parameters $\tilde{\theta}$ are estimated using an iterative procedure in
192 which an initial guess for the parameter values θ_0 is repeatedly updated according to
193 the gradient of objective function (Equation (4)) with respect to the parameters.

194 To facilitate training of an ANN, it is convenient to scale the inputs and targets of
195 the model to lie in the same range. We do so by scaling all input and target variables
196 to the interval $[0, 1]$ using the formula:

$$x_{scaled} = \frac{x_{orig} - x_{min}}{x_{max} - x_{min}} \quad (5)$$

197 where x_{min} and x_{max} are the minimum and maximum of variable x as encountered
198 in the training data.

199 The ANN model was implemented in Matlab 8.5 (R2015a) using the Neural Net-
200 work toolbox. This tool is known for the efficient implementation of the Levenberg-
201 Marquardt algorithm to neural network training, which appears to be the fastest
202 method for training moderate-sized feedforward neural networks up to several hun-
203 dred weights (Hagan et al, 2014). The collected raw data was divided into three sets
204 for training (70%), validation (15%) and test (15%) purposes. The training set is used
205 for computing the gradient and updating the network weights. The validation set pro-
206 vides the basis for optimizing the network setup, and to guard against overfitting the
207 weights on the training set. The test set is used to give an independent verification of
208 the model design (Demuth et al, 2009). To find $\tilde{\theta}$ given the training data, the `lmtrain`
209 optimization method provided with the toolbox was used. This method often finds bet-
210 ter (local) optima than regular gradient-descent (and in less iterations) by computing
211 an approximation of the second derivative of the objective function with respect to
212 the parameters, to estimate the stationary points of the gradient, corresponding to the
213 minima of the objective function.

214 **3 Results and discussion**

215 3.1 Experimental data and problem-solving approach

216 A total of 182 samples were collected from the permeate (i. e. product) stream of the
217 EMR during 10 independent experimental runs. The EMR was operated by varying
218 the residence time and the enzyme load under fixed temperature, pH, and sucrose feed
219 concentration. The total concentration of all carbohydrate fractions was 40 w/w% in
220 the collected samples, containing GF3, GF2, GF, G and F in various composition.
221 The chemical composition and UV spectral data of the samples were obtained via the
222 chromatographic and spectroscopic methods, as described in Sect. “2.4”.

223 The formation of individual compounds in the EMR is governed by certain phys-
224 ical and chemical relationships, and thus, depends on the reaction and operational
225 conditions employed in the EMR. It is to be noted that investigating the influence of
226 operational settings of EMR on FOS production is beyond the scope of this study. Our
227 goal is to obtain a dataset in which the carbohydrate composition varies in a broad
228 range in order to simulate possible scenarios that may occur in real-life applications
229 operating under both optimal and suboptimal conditions. Other reactor configurations,
230 such as packed bed reactors, in which an enzyme-free product stream is produced, are
231 also suitable to generate such a dataset. In this sense, the choice of EMR is arbitrary.

232 Fig. 2 shows the relative mass percentage of individual saccharides in the samples
233 collected during the experimental phase, sorted in descending order according to the
234 sucrose concentration.

235 Obviously, an UV absorbance value measured at a single wavelength is not suit-
236 able for estimating the relative concentrations of the individual saccharide fractions.
237 However, absorbance values recorded for a range of wavelengths may contain valuable
238 information regarding chemical composition in such a multi-component system. Fig. 3
239 presents the UV spectra of two illustrative samples and their corresponding composi-
240 tion measured by HPLC.

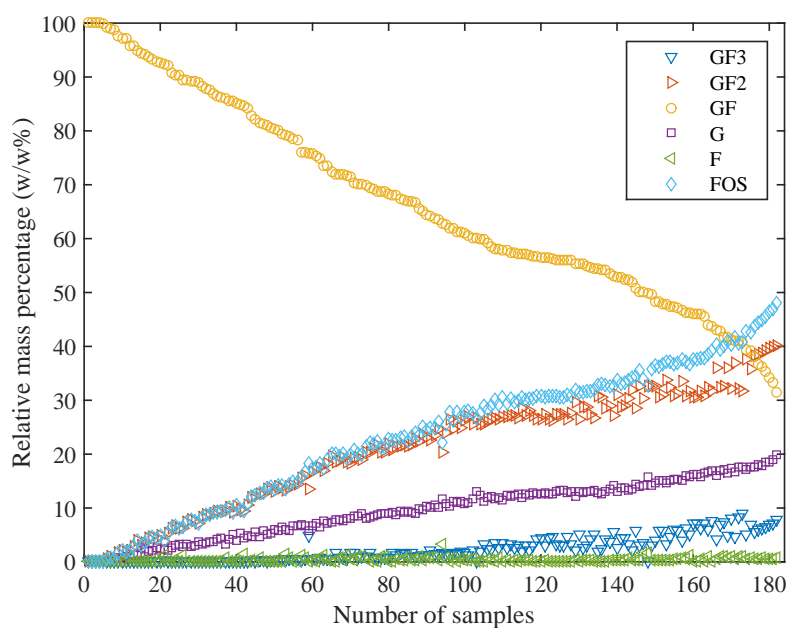


Fig. 2 Relative mass percentage of individual carbohydrate fractions in samples collected from the permeate stream of the enzyme membrane reactor at several settings of residence time and enzyme load. Operation conditions: pH 5.7 ± 0.2 , $50 \pm 0.5^\circ\text{C}$, 40 w/w% sucrose feed, 20 kDa UF membrane.

241 The large deviation observed in the absorption profiles is attributed to the relative
242 ratio of mono-, di-, and oligosaccharides present in the samples. The aim of this study
243 was to find the fundamental relation between the spectra and corresponding saccha-
244 rides composition. In other words, we want to use the easy-to-measure spectrographs,
245 i. e. absorbance values measured at a given range of wavelengths, to predict the amount
246 of different saccharides in the samples. For this purpose, we employ artificial neural
247 networks.

248 3.2 Determination of ANN hyperparameters

249 In terms of hyperparameters, the early stopping criterion and the hidden layer size (i.e.
250 number of neurons in the hidden layer) were investigated for the optimal setting. In
251 preliminary tests we have found, that deeper models (models with multiple hidden lay-

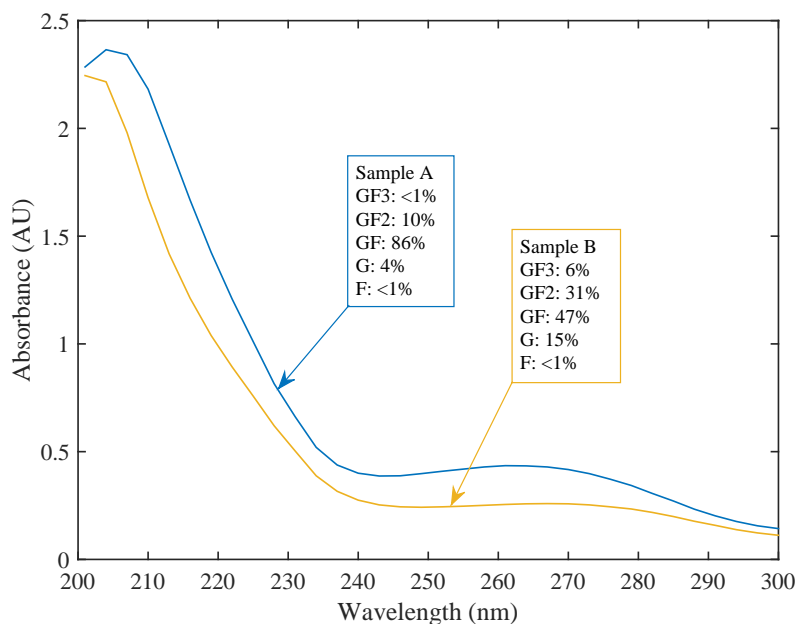


Fig. 3 UV spectra of two illustrative samples and corresponding chemical compositions determined by HPLC.

252 ers) do not give substantially better results, thus a simple model with a single hidden
 253 layer was used in this work for reasons of parsimony. The rest of the model parameters
 254 were set to the default values provided in the Neural Network Toolbox version 8.3. The
 255 hyperparameters including the training parameters of the Levenberg-Marquardt algo-
 256 rithm are shown in Table 1. Training is stopped when either of the following scenarios
 257 take place: the maximum number of training iterations (Max epochs) is reached, the
 258 performance goal is achieved, the performance gradient falls below the specified min-
 259 imum performance gradient, the dampening parameter μ exceeds the specified maxi-
 260 mum μ , or the validation error increases for the number of iterations specified by the
 261 early stopping criterion.

262 The early stopping criterion terminates the training of the network after the error
 263 on the validation set increases for a specified number of iterations. This prevents the
 264 model from adapting to variation in the training set that is due to measurement error

Table 1 Hyperparameters of the final model including the training parameters of the Levenberg-Marquardt algorithm

Parameter	Value
Number of hidden neurons	15
Max epochs	1000
Performance goal	0
Early stopping criterion	50
Minimum performance gradient	1e-7
Initial adaptive value (μ)	0.001
μ decrease factor	0.1
μ increase factor	10
Maximum μ	1e10

265 and other extraneous influences, and tends to produce models that better generalize
 266 beyond the training set. After terminating training, the weights at the minimum of the
 267 validation error are returned. Since the decrease of the validation error during training
 268 may be non-monotonic, a certain tolerance to increase of validation error is necessary.
 269 In our experiments, we found that a tolerance of 50 iterations was enough to make sure
 270 that the training was not stopped before the minimum validation error was reached.

271 Training performance of the models averaged at a determination coefficient of 0.99
 272 and MSE of <10 indicating that all of the tested models were capable of learning the
 273 underlying function of the training data. To choose the best neural network architecture
 274 in terms of the number of neurons in the hidden layer, networks with hidden layer sizes
 275 of 1 neuron to 20 neurons were created. Choice of the number of neurons to be examined
 276 was based on recommendations from the literature (Panchal et al, 2011; Huang, 2003).
 277 At each configuration the network was trained 80 times with random divisions of the
 278 data into training and validation. The MSE on the validation set was recorded for
 279 every network. The optimal setup was chosen by the average error on the validation
 280 set achieved at the different hidden layer sizes. Fig. 4 shows the resulting MSE values
 281 on a logarithmic scale for each hidden layer size.

282 The hidden layer size appeared to have little impact on the validation error in the
 283 range 4 to 20 neurons. Nonetheless, the best results were achieved with 15 neurons in

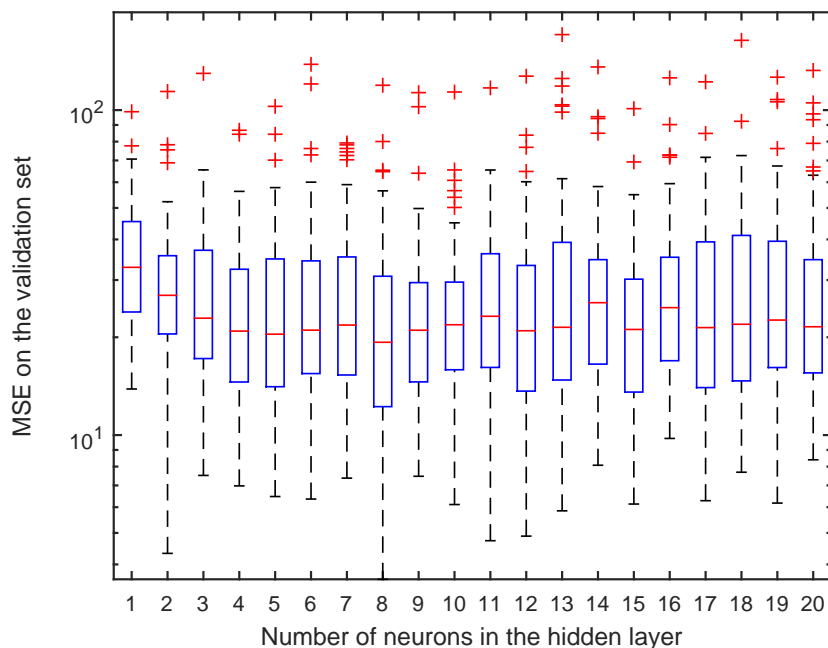


Fig. 4 Boxplot of the MSE on the validation set at different hidden layer sizes on a logarithmic scale.

284 the hidden layer. Thus, this network architecture was retained to be used for model
 285 validation and testing.

286 3.3 Model performance

287 The model was validated on the test set, a randomly selected 15% of the data, that
 288 was not used in training the network in any way. A determination coefficient of 0.98
 289 with a MSE value of 19.63 was achieved for the test set. Predicted concentration of
 290 individual saccharide fractions for the test set (expressed in relative mass percentage)
 291 is plotted against the reference values in Fig. 5.

292 The performance of the model for the test set is shown in detail in Table 2 includ-
 293 ing the prediction error (MSE), goodness of fit (slope, intercept, R^2) and the F-test
 294 (i. e. test for a significant linear regression relationship between the response variable
 295 and the predictor variables) results for each component.

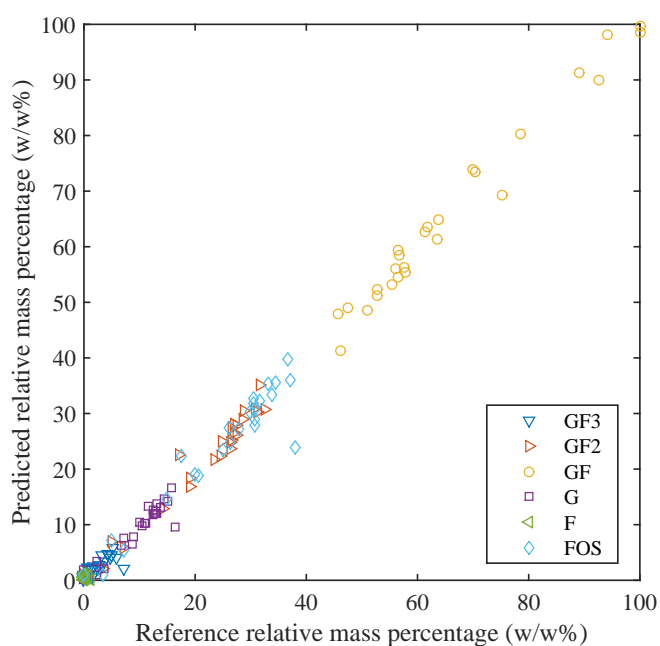


Fig. 5 Predicted versus reference concentration of individual saccharide fractions in the test set

296 The results indicate that the developed ANN model provides satisfactory perfor-
297 mance for the prediction of GF2, GF, G and total FOS. As shown in Table 2, smaller
298 R-squared values were obtained for components in low concentrations (GF3 and F).
299 The reason of the poorer prediction likely originates from the specific features of the
300 reference method. Note that samples were diluted prior to HPLC to avoid exceeding
301 the upper measurement limit of the RI detector. Thus, the main components (GF,
302 GF2, and G), which are present in high concentrations, are measurable. However,
303 the investigated range of concentration for GF3 and F tends to be close to the lower
304 detection limit of the HPLC, resulting relatively imprecise (but otherwise accurate)
305 measurements. For F and GF3, the measurement error relative to the variance of the
306 concentration is much larger, and since the measurement error cannot (and should not)
307 be learned by the model, less accurate predictions are made for those components.

308 A common way of improving the analysis of compounds in low concentration is
309 to repeat HPLC tests without dilution. This would be, however, a laborious task for

310 such a large dataset (182 observations). We would like to highlight that the reported
 311 ANN predicts F poorly only in relative terms, as measured by the R-squared value.
 312 The model predicts its concentrations well in absolute terms, i. e. adequately estimates
 313 that its concentration is very low (around 1 w/w%). From the technological point of
 314 view, this latter information is more important. As a compromise, we may sacrifice
 315 model fit for compounds in low concentration in order to reduce experimental efforts
 316 required for model building.

Table 2 Performance of the retained network on the test set by compounds

component	GF3	GF2	GF	G	F	FOS
MSE	1.66	3.40	6.59	2.81	0.32	10.79
slope	0.65	0.98	1.01	0.90	0.06	0.96
intercept	0.81	0.22	-1.29	0.44	0.34	0.56
R^2	0.62	0.97	0.98	0.88	0.08	0.92
F-test	38.3	672	1040	184	0.05	288
p-value	0.00	0.00	0.00	0.00	0.82	0.00

317 The reference and estimated relative mass percentage values for the test set are
 318 shown on Fig 6. The reference values and the estimated values are represented by open
 319 and closed symbols, respectively.

320 Note that the UV-ANN method may also function as an alarm system to detect
 321 unwanted disturbances during process run (e. g. microbial contamination, pH shift,
 322 temperature and concentration disturbances, etc.) that may occur in real-life appli-
 323 cations. However, the system is not able to provide any specific information on the
 324 type of disturbance, unless it is previously calibrated for such unwanted scenarios. The
 325 here proposed method is general in the sense that it could potentially be applied for
 326 any reactor configurations, given that experimental data is available for the specific
 327 process environment, and thus, the network can be re-trained following the procedure
 328 described in this paper.

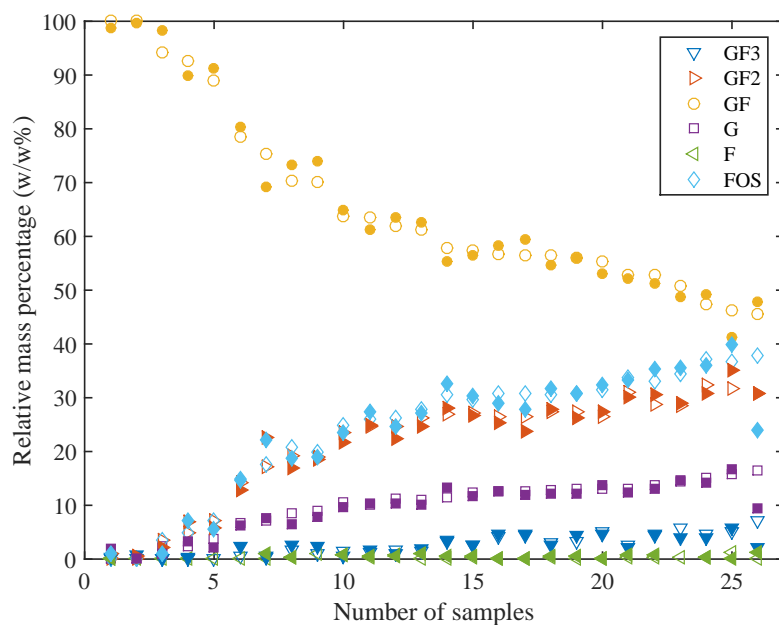


Fig. 6 Reference (open symbols) and predicted (closed symbols) concentrations of individual saccharide fractions in the test set.

329 4 Conclusions

330 In conclusion, our results suggest that the proposed UV-ANN method has a great
331 potential to be applied as an on-line monitoring tool in FOS manufacturing processes.
332 It provides satisfactory estimation of the main components found in the product stream
333 of enzymatic FOS production. The estimation accuracy is suitable for tracing the course
334 of the bioconversion. It may be implemented at certain critical points of the FOS
335 production line to monitor the progress of the reaction and to detect possible failures
336 and disturbances in the operation.

337 Acknowledgements

338 The preparation of this work was supported by the Marie Skłodowska-Curie grant of
339 the EU FP7 Framework Programme (PCIG11-GA-2012-322219) and the Bolyai Schol-
340 arship Programme of the Hungarian Academy of Sciences.

341 References

- 342 Bali V, Panesar PS, Bera MB, Panesar R (2013) Fructo-oligosaccharides: Pro-
343 duction, purification and potential applications. *Critical reviews in food science*
344 and nutrition 8398(June 2014):37–41, DOI 10.1080/10408398.2012.694084, URL
345 <http://www.ncbi.nlm.nih.gov/pubmed/24915337>
- 346 Barclay T, Ginic-Markovic M, Johnston MR, Cooper PD, Petrovsky N (2012)
347 Analysis of the hydrolysis of inulin using real time ¹H NMR spectroscopy.
348 *Carbohydrate Research* 352:117–125, DOI 10.1016/j.carres.2012.03.001, URL
349 <http://dx.doi.org/10.1016/j.carres.2012.03.001>
- 350 Borronei C, Careri M, Cavazza A, Corradini C, Elviri L, Mangia A, Merusi C (2009)
351 Evaluation of Fructooligosaccharides and Inulins as Potentially Health Benefiting
352 Food Ingredients by HPAEC-PED and MALDI-TOF MS. *International journal of*
353 *analytical chemistry* 2009:530,639, DOI 10.1155/2009/530639
- 354 Cámara M, Torrecilla JS, Caceres JO, Cortes Sánchez Mata M, Fernandez-Ruiz V
355 (2010) Neural network analysis of spectroscopic data of lycopene and β -carotene
356 content in food samples compared to HPLC-UV-Vis. *Journal of Agricultural and*
357 *Food Chemistry* 58(1):72–75, DOI 10.1021/jf902466x
- 358 Chen SC, Sheu DC, Duan KJ (2014) Production of fructooligosaccharides us-
359 ing β -fructofuranosidase immobilized onto chitosan-coated magnetic nanoparticles.
360 *Journal of the Taiwan Institute of Chemical Engineers* 45(4):1105–1110, DOI
361 10.1016/j.jtice.2013.10.003, URL <http://dx.doi.org/10.1016/j.jtice.2013.10.003>
- 362 Corradini C, Bianchi F, Matteuzzi D, Amoretti A, Rossi M, Zanoni S (2004) High-
363 performance anion-exchange chromatography coupled with pulsed amperometric de-

- 364 tection and capillary zone electrophoresis with indirect ultra violet detection as pow-
365 erful tools to evaluate prebiotic properties of fructooligosaccharides and inulin. Jour-
366 nal of Chromatography A 1054(1-2):165–173, DOI 10.1016/j.chroma.2004.07.109
- 367 Corradini C, Lantano C, Cavazza A (2013) Innovative analytical tools to character-
368 ize prebiotic carbohydrates of functional food interest. Analytical and Bioanalytical
369 Chemistry 405(13):4591–4605, DOI 10.1007/s00216-013-6731-6
- 370 Demuth H, Beale M, Hagan M (2009) Neural Network ToolboxTM 6 User's Guide,
371 Revised for Version 6.0.3 (Release 2009b)
- 372 Dias LG, Veloso ACA, Correia DM, Rocha O, Torres D, Rocha I, Ro-
373 drigues LR, Peres AM (2009) UV spectrophotometry method for
374 the monitoring of galacto-oligosaccharides production. Food Chem-
375 istry 113(1):246–252, DOI 10.1016/j.foodchem.2008.06.072, URL
376 <http://dx.doi.org/10.1016/j.foodchem.2008.06.072>
- 377 Dominguez AL, Rodrigues LR, Lima NM, Teixeira JA (2013) An Overview of the
378 Recent Developments on Fructooligosaccharide Production and Applications. Food
379 and Bioprocess Technology pp 1–14, DOI 10.1007/s11947-013-1221-6
- 380 Falguera V, Aliguer N, Falguera M (2012) An integrated approach to current
381 trends in food consumption : Moving toward functional and organic prod-
382 ucts? Food Control 26(2):274–281, DOI 10.1016/j.foodcont.2012.01.051, URL
383 <http://dx.doi.org/10.1016/j.foodcont.2012.01.051>
- 384 Ghazi I, Gómez De Segura A, Fernández-Arrojo L, Alcalde M, Yates M, Rojas-
385 Cervantes ML, Plou FJ, Ballesteros A (2005) Immobilisation of fructosyltransferase
386 from *Aspergillus aculeatus* on epoxy-activated Sepabeads EC for the synthesis of
387 fructo-oligosaccharides. Journal of Molecular Catalysis B: Enzymatic 35(1-3):19–27,
388 DOI 10.1016/j.molcatb.2005.04.013
- 389 Grube M, Bekers M, Upite D, Kaminska E (2002) Infrared spectra of
390 some fructans. Spectroscopy 16(3-4):289–296, DOI 10.1155/2002/637587, URL
391 <http://www.hindawi.com/journals/jspec/2002/637587/abs/>

- 392 Hagan M, Demuth H, Beale M, De Jesús O (2014) Neural network design. Martin
393 Hagan
- 394 Hicke HG, Becker M, Paulke BR, Ulbricht M (2006) Covalently coupled nanoparticles in
395 capillary pores as enzyme carrier and as turbulence promoter to facilitate enzymatic
396 polymerizations in flow-through enzyme-membrane reactors. *Journal of Membrane*
397 *Science* 282(1-2):413–422, DOI 10.1016/j.memsci.2006.05.051
- 398 Huang GB (2003) Learning capability and storage capacity of two-hidden-layer feed-
399 forward networks. *IEEE Transactions on Neural Networks* 14(2):274–281, DOI
400 10.1109/TNN.2003.809401
- 401 Joye D, Hoebregs H (2000) Determination of oligofructose, a soluble dietary fiber,
402 by high-temperature capillary gas chromatography. *Journal of AOAC International*
403 83(4):1020–1025
- 404 Kovács Z, Benjamins E, Grau K, Ur Rehman A, Ebrahimi M, Czermak P (2014)
405 Recent Developments in Manufacturing Oligosaccharides with Prebiotic Functions.
406 In: Zorn H, Czermak P (eds) *Biotechnology of Food and Feed Additives*, Springer
407 Berlin Heidelberg, Berlin, Heidelberg, pp 257–295, DOI 10.1007/10.2013.237, URL
408 http://dx.doi.org/10.1007/10_2013_237
- 409 Li J, Liu X, Zhou B, Zhao J, Li S (2013) Determination of fructooligosaccharides in
410 burdock using HPLC and microwave-assisted extraction. *Journal of Agricultural and*
411 *Food Chemistry* 61(24):5888–5892, DOI 10.1021/jf400534n
- 412 Lin H, Zhao J, Sun L, Chen Q, Zhou F (2011) Freshness measurement of eggs using
413 near infrared (NIR) spectroscopy and multivariate data analysis. *Innovative Food*
414 *Science and Emerging Technologies* 12(2):182–186, DOI 10.1016/j.ifset.2011.01.008,
415 URL <http://dx.doi.org/10.1016/j.ifset.2011.01.008>
- 416 Lorenzoni ASG, Aydos LF, Klein MP, Ayub MAZ, Rodrigues RC, Hertz PF (2015)
417 Continuous production of fructooligosaccharides and invert sugar by chitosan immo-
418 bilized enzymes: Comparison between in fluidized and packed bed reactors. *Journal*
419 *of Molecular Catalysis B: Enzymatic* 111:51–55, DOI 10.1016/j.molcatb.2014.11.002

- 420 Mateo F, Gadea R, Mateo EM, Jiménez M (2011) Multilayer perceptron neu-
421 ral networks and radial-basis function networks as tools to forecast accu-
422 mulation of deoxynivalenol in barley seeds contaminated with *Fusarium cul-*
423 *morum*. *Food Control* 22(1):88–95, DOI 10.1016/j.foodcont.2010.05.013, URL
424 <http://dx.doi.org/10.1016/j.foodcont.2010.05.013>
- 425 Mutanda T, Mokoena MP, Olaniran AO, Wilhelmi BS, Whiteley CG (2014) Microbial
426 enzymatic production and applications of short-chain fructooligosaccharides and in-
427 ulooligosaccharides: Recent advances and current perspectives. *Journal of Industrial*
428 *Microbiology and Biotechnology* 41(6):893–906, DOI 10.1007/s10295-014-1452-1
- 429 Nguyen QD, Rezessy-Szabó JM, Czukor B, Hoschke Á (2011) Con-
430 tinuous production of oligofructose syrup from Jerusalem arti-
431 choke juice by immobilized endo-inulinase. *Process Biochemistry*
432 46(1):298–303, DOI <http://dx.doi.org/10.1016/j.procbio.2010.08.028>, URL
433 <http://www.sciencedirect.com/science/article/pii/S1359511310003399>
- 434 Nishizawa K, Nakajima M, Nabetani H (2000) A forced-flow membrane reac-
435 tor for transfructosylation using ceramic membrane. *Biotechnology and Bio-*
436 *engineering* 68(1):92–97, DOI 10.1002/(SICI)1097-0290(20000405)68:1<92::AID-
437 BIT11>3.0.CO;2-1
- 438 Panchal G, Ganatra A, Kosta Y, Panchal D (2011) Behaviour analysis of
439 multilayer perceptrons with multiple hidden neurons and hidden layers. *Inter-*
440 *national Journal of Computer Theory and Engineering* 3(2):332–337, URL
441 <http://www.ijcte.org/papers/328-L318.pdf>
- 442 Petkova N, Denev P (2015) Methods for determination of inulin. In: Hadzhiev B (ed)
443 *Monograph of 4rd European Young Engineers Conference 2015, Food Research &*
444 *Development Institute, Plovdiv, UFT Academic Publishing House, Plovdiv, Plovdiv,*
445 *Bulgaria, ISSN 2367-6213, p 135*
- 446 Petkova N, Radka V, Denev P, Ivanov I, Pavlov A (2014) HPLC-RID method for
447 determination of inulin and fructooligosaccharides. *Acta Scientifica Naturalis* 1:99–
448 107

- 449 Prabhu AA, Jayadeep A (2017) Optimization of enzyme-assisted improve-
450 ment of polyphenols and free radical scavenging activity in red rice bran:
451 A statistical and neural network-based approach. *Preparative Biochemistry
452 and Biotechnology* 47(4):397–405, DOI 10.1080/10826068.2016.1252926, URL
453 <https://www.tandfonline.com/doi/full/10.1080/10826068.2016.1252926>
- 454 Rastall RA (2010) Functional Oligosaccharides: Application and Manufac-
455 ture. *Annual Review of Food Science and Technology* 1:305–339, DOI
456 10.1146/annurev.food.080708.100746
- 457 Sarup R, Pal R, Kennedy JF (2016) International Journal of Biolog-
458 ical Macromolecules Recent insights in enzymatic synthesis of fruc-
459 tooligosaccharides from inulin. *International Journal of Biological
460 Macromolecules* 85:565–572, DOI 10.1016/j.ijbiomac.2016.01.026, URL
461 <http://dx.doi.org/10.1016/j.ijbiomac.2016.01.026>
- 462 Tanriseven A, Aslan Y (2005) Immobilization of pectinex ultra sp-1 to pro-
463 duce fructooligosaccharides. *Enzyme and Microbial Technology* 36(4):550
464 – 554, DOI <http://dx.doi.org/10.1016/j.enzmictec.2004.12.001>, URL
465 <http://www.sciencedirect.com/science/article/pii/S014102290400362X>
- 466 Ur Rehman A, Kovacs Z, Quitmann H, Ebrahimi M, Czermak P (2016) Enzy-
467 matic production of fructo-oligosaccharides from inexpensive and abundant sub-
468 strates using a membrane reactor system. *Separation Science and Technology*
469 6395(April):01496,395.2016.1167,740, DOI 10.1080/01496395.2016.1167740, URL
470 <http://www.tandfonline.com/doi/full/10.1080/01496395.2016.1167740>
- 471 Van Den Broeke J, Langergraber G, Weingartner A (2006) On-line and in-situ UV/vis
472 spectroscopy for measurements: a brief review. *Spectroscopy Europe* 1 18(4):4–7
- 473 Veloso ACA, Rodrigues LR, Dias LG, Peres AM (2012) Chapter 14 uv spectrophotome-
474 try method for dietary sugars. In: *Dietary Sugars: Chemistry, Analysis, Function and
475 Effects*, The Royal Society of Chemistry, pp 229–248, DOI 10.1039/9781849734929-
476 00229, URL <http://dx.doi.org/10.1039/9781849734929-00229>