

Automated calibration procedure for Raman biomonitoring instruments

5 The hereby described invention discloses an automated calibration procedure using pre-existing Raman instrument data coupled with automated transfer learning techniques.

Technical Field

10 The invention deals with the technological area of Raman biomonitoring.

Background and description of the prior art

15 Raman spectroscopy-based instruments are complex to be calibrated because of the correlation between spectrum data and actual process parameters. It has to be understood that calibration in this context means calibrating the data process which is done by applying digital calibration models which are created using standard model building techniques. The current state of the art in the calibration of these instruments via the
20 application of calibration models is based on the correlation between Raman spectrum data and several parameter concentrations measured by reference instruments, using a specific regression modelling method. For that, at least three batches are required to have enough data at early process stage, and even more batches are required when moving into
25 manufacturing to ensure robust models considering the whole process variabilities.

As soon as there are changes in the process conditions, like instrument, cell culture media, cell line, feeding solution etc, the chemometric models
30 become inaccurate. The good practices in chemometrics at this stage are to rebuild all the previous models using the new process conditions or to feed the existing models including the new variability.

For the instrumental transferability: prior art mentions transfer algorithms such as DS (Direct Standardization), GLS (Generalized Least Squares) or OLS (Ordinary Least Squares). All of those require measuring the exact same “transfer set” samples on both instruments before being able to calculate a transfer matrix. The method is based on the principle that any signal differences would be instrumental since the samples are chemically identical. The algorithms mentioned above also require the user to be an expert to understand how they work and for the user to be capable of using them optimally.

Furthermore in bioprocessing applications, after having developed chemometric models for a particular process, it is important to continue to use those models with the same instrument and experimental conditions for which they have been calibrated. This ensures consistency, reproducibility, and a reliable accurate monitoring of the process. As soon as experimental conditions differ from the model building conditions and its variability was not considered, it can introduce deviations and potentially lead to inaccurate or unreliable predictions. Changes in the experimental conditions, like media, cell line, process conditions, concentration ranges, sample preparation methods etc, may have an impact on the data leading to spectral variations as the relationship between the different variables that constitute a Raman spectrum may also change and are therefore not included in the existing models. Such variabilities can strongly offset the models resulting in inaccurate Raman estimations during monitoring.

In this context, the problem for users is to restart the model building protocol from scratch to develop new models that include the new variability coming from the new process conditions. This leads to the acquisitions of new sets of data that can take a long time to generate, which are also expensive in terms of consumables and resources.

Instrument variability is a point not to be overlooked as well because some biases may occur in the predictive capabilities if a new instrument is used with models calibrated on another. The variations can arise from instrumental factors such as differences in the laser sources, optic fibers, connectors etc. Overall, using the same analyzer and experimental conditions help to maintain the reliability of the chemometric models and consequently, additional efforts are necessary most of the time to adapt a model to a new environment.

Although instrumental variability is a significant challenge in Raman spectroscopy for obtaining good prediction capabilities, several methods have been developed to correct it and to minimize such effects.

From the European Patent Application EP3822717 A1 a method of predicting a parameter of a medium to be observed in a bioprocess based on Raman spectroscopy is known, with the following method steps of acquiring a first series of preparatory Raman spectra of an aqueous medium using a first measuring assembly; normalizing the first series of preparatory Raman spectra based on a characteristic band of water from at least one Raman spectrum acquired with the first measuring assembly; building a multivariate model for the parameter based on the normalized preparatory Raman spectra; acquiring predictive Raman spectra of the medium to be observed during the bioprocess acquired with another measuring assembly; normalizing the predictive Raman spectra based on a characteristic band of water from at least one Raman spectrum acquired with the other measuring assembly; and applying the built model to the predictive Raman spectra for predicting the parameter.

Furthermore from the European Patent Application EP3276333 A1 a device is known, comprising of one or more processors to obtain a master calibration set associated with a master calibration model of a master instrument, the master calibration set including spectra, associated with a

set of samples, generated by the master instrument; identify a selected set of master calibrants based on the master calibration set, the selected set of master calibrants including spectra associated with a subset of the set of samples; obtain a selected set of target calibrants associated with a target instrument, the selected set of target calibrants including spectra, associated with the subset of the set of samples, generated by the target instrument; create a transfer set based on the selected set of master calibrants and the selected set of target calibrants, the transfer set being associated with the subset of the set of samples; create a target calibration set, corresponding to the master calibration set, based on the transfer set; and generate, based on a support vector regression modeling technique, a transferred calibration model, associated with the target instrument, based on the target calibration set, the transferred calibration model being generated using an optimization technique associated with the transfer set.

However, it's important to mention that these methods only address instrumental variability and not process variability such as changes in cell-line, media, measurement conditions etc... The use of generic models can help in this respect but can also introduce too much variability leading to less accurate results compared to the proposed improvement described below.

The task of this patent application is therefore to develop an approach to reduce the number of required calibration batches by a significant coefficient.

Summary of the invention

This task has been solved by a method for calibrating at least one software-based regression model used for optimizing sensor data of a Raman Spectroscopy Instrument used for measuring a target bioprocess wherein the at least one software-based regression model is run by a computer,

comprising the following steps of Acquiring at least one batch of Raman spectra data of both an initial Raman-monitored bioprocess and the target bioprocess associated with off-line reference values for all required bioprocess parameters and conditions; Creating an optimized software-based model for the target bioprocess adapted for every monitored bioprocess parameter by using dimensionality reduction techniques and applying a machine-learning algorithm using the Raman spectra data and the off-line reference values of the initial bioprocess and the target bioprocess via the computer; and Using the optimized software-based models in the Raman spectroscopy instrument for monitoring the target bioprocess. The invention described here is about an innovative approach to reduce the number of required calibration batches thanks to a transfer learning method, reducing the calibration effort by a significant coefficient. This invention allows to easily transfer models in different process without the need to re-perform all the model steps. The innovation resides in the applicability landscape but most notably in the automation which makes it easily usable by non-experts.

Advantageous and therefore preferred further developments of this invention emerge from the associated subclaims and from the description and the associated drawings.

One of those preferred further developments of the disclosed method comprise that the at least one software-based regression model is a machine learning model or a statistical, in particular a PLS, model. Which kind of model is most preferable depends on the used Raman instrument and the monitored target process.

Another one of those preferred further developments of the disclosed method comprise that the dataset of Raman spectra consists of samples coming from

at least one batch or any set of samples representative of the initial and target bioprocess parameters and conditions.

5 Another one of those preferred further developments of the disclosed method comprise that the Raman spectra data from the initial bioprocess and the target bioprocess is preprocessed using preprocessing tools. This step is optional, but of course a suitable preprocessing enhances the quality of the Raman spectra data and therefore also the resulting quality of the calibration process.

10 Another one of those preferred further developments of the disclosed method comprise that the preprocessing tools include a water normalization, a Savitzky-Golay derivative and smoothing, in particular with the configuration of 1st order derivative, 2nd order polynomial and 5 points window, and a spectral range selection of 350-1775 and/or 2800-3000 cm-1. These are only specifically preferred examples for used preprocessing tools which have shown to provide good results. The invention is not limited to those tools and technically also works without preprocessing at all.

20 Another one of those preferred further developments of the disclosed method comprise that before creating the at least one software-based regression model operating ranges of metabolites in the target bioprocess are analyzed and determined. Furthermore the operating ranges have to be equivalent or in-between those of the initial bioprocess. The reason to do so is that the initial Raman biomonitored bioprocess and the target bioprocess should preferably operate within the same parameter operating ranges. Otherwise the utility of the Raman spectra data of the initial bioprocess is limited and the results of calibration transfer is worse compared to matching operating ranges.

30

Another one of those preferred further developments of the disclosed method comprise that as operating ranges of metabolites in the target bioprocess the parameter of concentration of the metabolites is used. Technically also other bioprocess parameters could be used, but using the concentration of the metabolites has lead to adequate results.

Another one of those preferred further developments of the disclosed method comprise that for creating an optimized at least one software-based regression model, in particular a chemometric model, for every bioprocess parameter by applying the machine-learning transfer-algorithm on the Raman and the reference data of the initial bioprocess and the target bioprocess the following method steps are performed by the computer, respective software: Computing for each possible value of a Singular Value Decomposition (SVD) from 1 to I, with I being set by the software by default to at least 10, of a Dynamic Orthogonal Projection (DOP) algorithm, a PLS regression using the initial process data as input and the target bioprocess data as validation; Calculating for each 1 to I of the SVD decomposition, and each 1 to j, with j being set by the software by default to at least 10, the statistical criteria of performance, in particular PLS latent variables, RMSEP, Bias, R^2 , Intercept, Slope, standard deviation differences, Coefficient of Variation RMSEP (CVRMSEP) and Bias Corrected RMSEP (BCRMSEP); Creating statistical tables with lines I and columns j filled with the computed values for each statistical criteria of performance; Calculating via the software for each statistical criteria of performance a maximum threshold value where above the condition $i*j$ is not acceptable; Determining via the software which $i*j$ conditions satisfy the maximum threshold value simultaneously on all considered tables and computing the DOP based on the $i*j$ conditions which summed equals to the lowest cumulative number while satisfying all previous conditions to avoid overfitting; and Creating via rendering an optimized PLS model pre-computed with the optimal $i*j$ parameters. This is a very specific approach which is one possible and preferred embodiment of the disclosed invention.

Others do exist and the invention is not limited to this preferred embodiment.

Another one of those preferred further developments of the disclosed
5 method comprise that the maximum threshold value depends on the input
data and the bioprocess parameter for which the respective model is
created. The threshold value can determined in different ways. Possible
approaches are that it can be fix and coupled directly to the bioprocess
parameter, maybe with different values regarding the input data. Or it can
10 be calculated by the software ith a formula using bioprocess parameter and
input data.

Another one of those preferred further developments of the disclosed
method comprise that If no i^*j condition satisfies simultaneously all
15 statistical criteria thresholds conditions, then the software uses estimated
values based on implemented intelligence, wherein the implemented
intelligence is based on experience data in form of statistical parameters
from the initial Raman-monitored bioprocess which is used by the software
by applying an algorithm to determine which set of parameters from the
20 initial Raman-monitored bioprocess is most suitable to be used with the
model calibrated for the target bioprocess. This algorithm can use different
approaches how to determine what is considered suitable. One possible
approach might be the mentioned operating ranges for the bioprocess
parameters. The closer they are for initial Raman-monitored bioprocess and
25 target bioprocess the more suitable it is to use respective sets of
paramters from the initial Raman-monitored bioprocess. Experience data
can also help to determine suitable parameter sets.

Another one of those preferred further developments of the disclosed
30 method comprise that the bioprocess conditions include the used media,
cell-line, number of batches, number of samples, feeding, clone integration
time, Raman Analyzer Instrument, scale or location, while the monitored

bioprocess parameters include glucose, lactate and VCD. Also here the invention is not limited to these bioprocess conditions, even if they are preferable to use. Other conditions might work as well.

5 Another one of those preferred further developments of the disclosed method comprise that the acquiring of the at least one batch of Raman spectra data of the target bioprocess is done by using samples of the target process measured either at-line or in-line, within one same batch or several batches. These are at least the most promising approaches.

10

Another one of those preferred further developments of the disclosed method comprise that the method steps are performed by the software alone in a complete automated way, manually by a human user or both partially divided by the human user and the software performed by the computer. Which approach is most suitable depends on the target

15 bioprocess, the automatization performance of the available software and hardware and the abilities and experience of the human users.

Another one of those preferred further developments of the disclosed method comprise that if the operating ranges of the target bioprocess parameters differs from the operating ranges of the initial bioprocess parameter, the method will only be applied to transfer correctly within common range of both bioprocesses but will not be validated outside of the common range. This is a safeguard to prevent miscalibrated Raman

20 instruments which can happen very easily if the invented method is applied to non-matching bioprocesses without a capable human user available who is able to correct the resulting deviations.

Another solution for the given task is a system for calibrating a Raman

30 spectroscopy instrument for a target bioprocess, comprising of the Raman spectroscopy instrument, a computer performing an optimized software-based model, data input and output means connected to the computer, at

least one database for storing the Raman spectra data and the associated off-line reference values saved on a memory accessible to the computer, wherein the system is configured to perform the previously described method. The invention also includes a respective system with the required hardware to apply the invented method.

Detailed description of the invention

The method according to the invention and functionally advantageous developments of those are described in more detail below with reference to the associated drawings using at least one preferred exemplary embodiment. In the drawings, elements that correspond to one another are provided with the same reference numerals.

The drawings show:

- Figure 1: a schematical representation of the methodology
- Figure 2: a schematical representation of a transfer batch for the automated algorithm optimization
- Figure 3: the results obtained for the second real-time monitoring batch
- Figure 4: the results obtained for the third real-time monitoring batch
- Figure 5: a schematic workflow of the first mode of operation

In the following the invented method is described in form of a preferred embodiment. It is not limited to this embodiment though, but shows just a specifically preferred embodiment. The invented transfer methodology can be used in more versatile ways such as instrumental variability and/or media difference and/or cell-line difference. Figure 1 shows a schematical representation of the methodology in a generic illustrative example.

Assuming the users have pre-existing Raman data coming from an initial bioprocess and you want to use it in new variability conditions, the method

consists in acquiring one or two batches of Raman spectra data associated to off-line reference values for all the parameters the users want to model on the new process conditions. Then, the first step is to analyze the typical operating ranges, i.e. the concentration, of the metabolites in the target process. The operating ranges are preferably equivalent or in-between those of the initial process.

The second step consists of preprocessing the Raman data coming both from the initial process and the target process using any preprocessing tools. In this case, water SNV (Standard Normal Variate) normalization was used followed by Savitzky-Golay (1st order derivative; 2nd order polynomial; 5 points window) and spectral range selection (350-1775; 2800-3000 cm⁻¹).

In the third step, the Raman and the reference data of the initial process and the Raman and reference data of the target process are uploaded into a developed optimization code, performed by a software hosted by suitable computer, e.g. a local personal or industrial computer or a remote server which can be cloud-based or local. The computer respective the software then automatically computes and optimizes the transfer learning based on the following steps:

1. For each possible value of SVD (Singular Value Decomposition) from 1 to i (i being set by the software by default to 10 but not limited to this value) of the public DOP (Dynamic Orthogonal Projection) algorithm, computes a PLS regression using the initial process data as input and the target process data as validation.
2. For each 1 to i of the SVD decomposition, and each 1 to j (j being set by the software by default to 10 but not limited to this value) of the PLS latent variables, calculate RMSEP, Bias, R^2 , Intercept, Slope, standard deviation differences, CVRMSEP (Coefficient of Variation RMSEP) and BCRMSEP (Bias Corrected RMSEP).

3. Fill out statistical tables of size (lines (I) * columns (J)) with the computed values for each statistical criteria of performance.
4. For each statistical criteria of performance, the software calculates a threshold value below which the condition I*J is "acceptable" and above
5 which the condition I*J is not acceptable. This threshold depends on the input data & the parameter (metabolite) for which we are building models.
5. Then, the software automatically determines which I*J conditions satisfies the threshold conditions simultaneously on all considered tables (if exists). Then, the software computes the DOP based on the I*J conditions
10 which summed equals to the lowest cumulative number while satisfying all previous conditions to avoid overfitting.

If there is not a single I*J conditions which satisfies simultaneously all statistical criteria thresholds conditions, then the software will make
15 compromises to deliver based on implemented intelligence in the code.

6. Finally, the Software will automatically render a PLS model pre-computed with the optimal I*J parameters and ready to be used in the new process conditions.

- 20 All the steps mentioned above are seamless for the user as the user would click on one button and the software would deliver the final step. Similarly to when computing one PLS-1 models for each bioprocess parameter of interest, the methodology needs to be done as many times as there are parameters to model, for example: One transfer for glucose, another one
25 for lactate, another one for IgG titer, etc..

In the following the results of some working examples for the preferred embodiment of the invention are presented.

- 30 The spectral data were acquired with a Raman PAT Platform using a process A with the following process conditions:

- CHO-ZN cell-line with EX-CELL® Advanced HD Perfusion Medium on instrument A with 30 seconds integration time.
- Cellvento® 4Feed COMP / EX-CELL® Advanced CHO Feed 1 as feeding solution.
- 5 • ProCellics™ Single Channel Unit 1 (S/N: PC-1002625)

Then, the developed models were used on a target process B:

- 10 • CHO-DP12 cell-line with HyClone™ ActiPro™ cell culture media on instrument B with 20 seconds integration.
- No feeding strategy.
- ProCellics™ Single Channel Unit 2 (S/N: PC-1002629)

15 All the process conditions are summarized in the data table below:

Dataset	N. samples	N. batches	Integration time (s)	Media	Clone	Feeding
Process A (initial)	171	8	30	EX-CELL® HD Perfusion	CHO-ZN	Cellvento® 4Feed COMP / EX-CELL® CHO Feed1
20 Process B (target)	44	3	20	HyClone™ ActiPro™	CHO-DP12	No feeding

Three parameters were assessed: glucose, lactate and VCD. The spectral data of both datasets was pre-processed using water SNV normalization
25 followed by Savitzky-Golay (1st order derivative; 2nd order polynomial; 5 points window) and spectral range selection (350-1775; 2800-3000 cm⁻¹).

Firstly, without using any transfer algorithm and trying to use the Process A data to build PLS models to monitor Process B. The results obtained are
30 presented in the data table below where the first batch of Process B was used as validation set:

Parameter	N _{calibration}	Calibration range	RMSEC	RMSECV	N _{validation}	Validation Range	RMSEP	Relative error
Glucose (g/L)	165	0.1 – 10.4	0.20	0.26	16	0 – 4.9	1.78	36%
Lactate (g/L)	170	0 – 9.1	0.12	0.14	16	0.3 – 3.4	1.33	39%
VCD (Mcells/mL)	166	0.4 – 34.1	1.78	1.82	16	0.6 – 6.4	4.66	73%

5 Based on the three relative errors which are above 30%, it appears that the PLS models are performing poorly in such conditions and are not suitable for predicting the different parameters accurately.

10 Secondly, using our transfer learning methodology, the same input data from Process A was used to build the PLS models but this time, the input spectral data were modified by using the first batch of Process B as a “transfer batch” for the automated algorithm optimization, as described in Figure 2.

15 Once the new PLS regressions adapted to Process B were obtained, the new PLS models were then imported into Bio4C® PAT Raman Software Version 6.0 to be able to perform real-time quantitative monitoring on new batches. The results obtained for the second real-time monitoring batch are presented in Figure 3 and the following data table set:

20

Parameter	N _{validation}	Validation Range	RMSEP	Relative error (%)
Glucose (g/L)	14	0 – 5.5	0.52	9%
Lactate (g/L)	14	0.3 – 3.5	0.21	6%
VCD (Mcells/mL)	14	0.5 – 6	0.70	12%

25

The models were capable to measure with satisfying accuracy the second batch of the target process.

30 The results obtained for the third real-time monitoring batch are presented in Figure 4 and the following data table set:

Parameter	N _{validation}	Validation Range	RMSEP	Relative error (%)
Glucose (g/L)	14	0 – 5.3	0.31	6%
Lactate (g/L)	14	0.3 – 3.5	0.33	10%
VCD (Mcells/mL)	14	0.6 – 8.1	1.10	14%

5

The same models were also capable to measure with satisfying accuracy the third batch of the target process.

10

For this preferred exemplary embodiment there are three different Mode of Operations. Figure 5 shows the First Mode while the other two Modes are describes in the following two chapters:

Mode of operation 2:

15

The user is working since a while with Raman on cell-line A, media A and instrument A. and then acquires a lot of data on these process conditions and doesn't want to lose them. Therefore he wants to transition into different conditions, like cell-line B, media B and instrument B, but he doesn't want to go through the whole model building phase again as it takes a lot of time and resources. By using our invention would allow user to transfer his models easily with one or two batch of transfer data acquired in the new process conditions by keeping all the data already acquired.

20

Mode of operation 3:

25

Here the user has pre-built models embedded within the system, which means he would need to run only 1 or 2 batches to adapt the models to the respective process. This greatly facilitates the implementation and reduce the lengthy model building phase which often keeps customers from using Raman technology.

30

Improvements:

5 This approach greatly benefits especially during the pre-sale evaluations of the system when users often don't have much time and resources dedicated for model building. Then, proposing this solution would greatly improve the success of such evaluations and prove the capabilities of the Raman PAT Platform more easily.

Further improvements over the prior art comprise that:

- 10 - No need to have the exact same samples in the transfer set measured on both instruments.
- Only 1 or 2 batches, around 10 - 15 samples over the entire process range, of the target process conditions are enough to have reliable transferability.
- 15 - The data can be transferred from an old system obsolete which is not working anymore.
- The method works well with cumulative variability, whereas prior art methods work well only with instrumental variability.
- No need to be an expert to make it work as everything is automated.
- 20 - Significant saves in time, consumables, and resources as there is no need to re-acquire several batches of data in the new process conditions to re-build relevant models.

25 Advantages of this method in case of instrumental transferability include that the transfer set can be measured only on the new instrument, meaning it can transfer models from an old system that is obsolete and not working anymore and only one or two batches are required for effective transfer resulting in around 10 - 15 samples over the entire process range.

30 Advantages of this method in case of process transferability include the it can transfer models to a process with a different cell-line and/or different media and/or different feed and so on.

Advantages of this method in both cases include that the method doesn't require any expertise as all transfer parameters are automatically optimized to give the best results and the new models are automatically built -> "one button auto-transfer" for ready-to-use models.

5

10

15

20

25

30

Patent claims

1. Method for calibrating at least one software-based regression model
used for optimizing sensor data of a Raman Spectroscopy Instrument
5 used for measuring a target bioprocess wherein the at least one
software-based regression model is run by a computer, the following
steps comprising:
 - Acquiring a dataset of Raman spectra data associated with off-
line reference values for an initial bioprocess and the target
10 bioprocess for all required bioprocess parameters and conditions;
 - Creating at least one optimized software-based regression model
for the target bioprocess adapted for every monitored bioprocess
parameter by using dimensionality reduction techniques and
15 applying a machine-learning algorithm using the Raman spectra
data and the off-line reference values of the initial bioprocess and
the target bioprocess via the computer;
 - Using the at least one optimized software-based regression
models in the Raman spectroscopy instrument for monitoring the
20 target bioprocess.
2. Method according to claim 1, wherein
the at least one software-based regression model is a machine
learning model or a statistical, in particular a PLS, model.
- 25 3. Method according to any of the previous claims, wherein
the dataset of Raman spectra consists of samples coming from at
least one batch or any set of samples representative of the initial and
target bioprocess parameters and conditions.
- 30 4. Method according to any of the previous claims, wherein

the initial bioprocess and the target bioprocess have been previously monitored by an Raman spectroscopy instrument.

5. Method according to any of the previous claims, wherein
5 the Raman spectra data from the initial bioprocess and the target bioprocess is preprocessed using preprocessing tools which include a water normalization, a Savitzky-Golay derivative and smoothing, in particular with the configuration of 1st order derivative, 2nd order polynomial and 5 points window, and a spectral range selection of 350-1775 and/or
10 2800-3000 cm⁻¹.
6. Method according to any of the previous claims, wherein
before creating the at least one software-based regression model
operating ranges of metabolites in the target bioprocess are analyzed and
15 determined and the operating ranges have to be equivalent or in-between those of the initial bioprocess.
7. Method according to claim 6, wherein
as operating ranges of metabolites in the target bioprocess the
20 parameter of concentration of the metabolites is used.
8. Method according to any of the previous claims, wherein
for creating an optimized at least one software-based regression model,
in particular a chemometric model, for every bioprocess parameter by
25 applying the machine-learning transfer-algorithm on the Raman and the reference data of the initial bioprocess and the target bioprocess, the following method steps are performed by the computer:
 - Computing for each possible value of a Singular Value
Decomposition (SVD) from 1 to I, with I being set by the software
30 by default to at least 10, of a Dynamic Orthogonal Projection (DOP) algorithm, a PLS regression using the initial process data as input and the target bioprocess data as validation.

- 5
- Calculating for each 1 to I of the SVD decomposition, and each 1 to j, with j being set by the software by default to at least 10, the statistical criteria of performance, in particular PLS latent variables, RMSEP, Bias, R^2 , Intercept, Slope, standard deviation differences, Coefficient of Variation RMSEP (CVRMSEP) and Bias Corrected RMSEP (BCRMSEP);
 - Creating statistical tables with lines I and columns j filled with the computed values for each statistical criteria of performance;
 - 10
 - Calculating via the software for each statistical criteria of performance a maximum threshold value where the condition $i*j$ is not acceptable;
 - Determining via the software which $i*j$ conditions satisfy the maximum threshold value simultaneously on all considered tables and computing the DOP based on the $i*j$ conditions which summed equals to the lowest cumulative number while satisfying
 - 15
 - all previous conditions to avoid overfitting;
 - Creating via rendering an optimized PLS model pre-computed with the optimal $i*j$ parameters .
- 20
9. Method according to claim 8, wherein the maximum threshold value depends on the input data and the bioprocess parameter for which the respective optimized at least one software-based regression model is created.
- 25
10. Method according to claim 8 or claim 9, wherein
- 30
- If no $i*j$ condition satisfies simultaneously all statistical criteria thresholds conditions, then the computer uses estimated values based on implemented intelligence, wherein the implemented intelligence is based on experience data in form of statistical parameters from the initial bioprocess which is used by the computer by applying an algorithm to determine which set of parameters from the initial

bioprocess is most suitable to be used with the model calibrated for the target bioprocess.

- 5 11. Method according to any of the previous claims, wherein the bioprocess conditions include the used media, cell-line, number of batches, number of samples, feeding, clone integration time, Raman Analyzer Instrument, scale or location, while the monitored bioprocess parameters include glucose, lactate and VCD.
- 10 12. Method according to any of the previous claims, wherein the acquiring of the at least one batch of Raman spectra data of the target bioprocess is done by using samples of the target process measured either at-line or in-line, within one same batch or several batches.
- 15 13. Method according to any of the previous claims, wherein the method steps are performed by the computer alone in a complete automated way, manually by a human user or both partially divided by the human user and the computer.
- 20 14. Method according to any of the previous claims, wherein if the operating ranges of the target bioprocess parameters differs from the operating ranges of the initial bioprocess parameter, the method will only be applied to transfer correctly within common range of both bioprocesses but will not be validated outside of the common range.
- 25 15. System for calibrating a Raman spectroscopy instrument for a target bioprocess, comprising of the Raman spectroscopy instrument, a computer performing at least one optimized software-based chemometric model, data input and output means connected to the computer, at least one database for storing the Raman spectra data and the associated off-line reference values saved on a memory
- 30

accessible to the computer, wherein the system is configured to perform the method of claims 1 to 14.

5

10

15

20

25

30