

between two to four weeks in a laboratory setting, for example. Further, only limited samples may be available for the analytical instruments (e.g., to ensure that a lab-scale bioreactor maintains a healthy mass of viable cells). In fact, it is not uncommon to have only one or two measurements available each day from in-line or offline analytical instruments. To further exacerbate the situation, the current best practices yield calibration models that are tied to a specific process, the specific formula or profile of the bioreactor media, and the specific operating conditions. Thus, if any of the aforementioned variables were to change, the models may need to be re-calibrated based on new data. In fact, both Raman model calibration and model maintenance require significant resource allocations and are typically performed in an offline setting. While approaches that adapt models to new operating conditions have been proposed (e.g., recursive, moving-window, and time-difference methods), these methods may be unable to adequately handle abrupt process changes.

[0006] There are a number of publications describing generic Raman models based on traditional chemometric methods (e.g., PLS modeling) for multiple molecules. However, these generic models assume that the processes use similar, if not the same, media formulations and/or run process conditions. The media and processes are usually platformed with little or no variation. The drawback of this type of generic model is that once a process deviates from the norm, or if the training dataset contains too wide of a process range in an effort to account for the variations (e.g., media additives, process duration and/or other process changes) between the different molecules, the generic models lose accuracy and precision. Therefore, these “generic” models are only generic within the described strict boundaries. See Mehdizadeh et al., *Biotechnol. Prog.* 31(4): 1004-1013, 2015; Webster et al., *Biotechnol. Prog.* 34(3):730-737, 2018.

SUMMARY OF THE INVENTION

[0007] The term “biopharmaceutical process” refers to a process used in biopharmaceutical manufacturing, such as a cell culture process to produce a desired recombinant protein. Cell culture takes place in a cell culture vessel, such as a bioreactor, under conditions that support the growth and maintenance of an organism engineered to express the protein. During recombinant protein production, process parameters, such as media component concentrations, including nutrients and metabolites (e.g., glucose, lactate, glutamate, glutamine, ammonia, amino acids, Na⁺, K⁺ and other nutrients or metabolites), media state (pH, pCO₂, pO₂, temperature, osmolality, etc.), as well as cell and/or protein parameters (e.g., viable cell density (VCD), titer, cell state, critical quality attributes, etc.) are monitored for control and/or maintenance of the cell culture process.

[0008] To address some of the aforementioned limitations of the current best industrial practices, embodiments described herein relate to systems and methods that improve upon traditional techniques for spectroscopic analysis of biopharmaceutical processes, such as Raman spectroscopy. In particular, a “Just-In-Time Learning” (JITL) platform is used to build and maintain calibration models (e.g., Raman calibration models) in real-time for biopharmaceutical applications. JITL is a nonlinear modeling platform based on local modeling and database sampling technology. Unlike other machine-learning methods, JITL generally assumes that all available observations are stored in a central database, and models are dynamically built in real-time based upon a query, using the most relevant data from the database. This allows for good approximation of complicated process dynamics using relatively simple local models. Under the JITL framework, a library may contain spectral data not only for a single process operating under specific operating conditions, but also data for different processes, different media profiles, and/or different operation conditions. This can significantly reduce the time required to calibrate and maintain models, especially for pipeline drugs that may have little or no past production history.

[0009] The JITL platform maintains a dynamic library that may be updated each time a new analytical measurement is available. Further, to ensure that the local models adapt to new process conditions, the last available analytical measurement (e.g., for the product currently being monitored) may always be included in the training set for local modeling. This allows the local model to more quickly adapt to new conditions, or to new product lines with no history. Using this approach, model

calibration and model maintenance may both be automated, and the time and expense (e.g., material and labor costs) associated with routine calibrations in conventional systems may be greatly reduced. Moreover, the ability to provide credibility bounds (or other confidence indicators, such as confidence scores) around model predictions may allow for robust monitoring and control strategies.

[0010] In some embodiments, Gaussian process models are used for local modeling, within the JITL framework. Gaussian process models are powerful statistical machine-learning models that can efficiently capture complex nonlinear process dynamics, and can readily adapt to virtually any process changes. In contrast to PLS, principal component regression (PCR) and other types of regression models, Gaussian process models are non-parametric methods, and are far more capable of capturing complex correlations between the Raman spectra and the analytical measurements from limited data sets. Moreover, Gaussian process models generally do not require pre-processing filtering of the Raman scans. Accordingly, in some embodiments, the Gaussian process models are instead calibrated on the raw Raman scans (in logarithmic scale), which may save many steps in the model calibration/maintenance process. Furthermore, Gaussian process models provide credibility bounds around the predictions, which can be extremely difficult to obtain using PLS or PCR models. Credibility bounds can be particularly useful for designing optimal sampling strategies for analytical instruments, and/or for implementing closed-loop control (e.g., model-predictive control, or MPC), for instance, to avoid making changes based on unreliable predictions.

[0011] Although JITL is a nonlinear modeling framework, and although the approach described above provides some adaptability by updating the dynamic library with recent analytical measurements, JITL alone may not be sufficiently adaptive to account for time-varying process conditions (e.g., abrupt changes to the set-point or other process conditions). In particular, local models that are calibrated using JITL may fail to make use of recent samples. For example, and particularly if there has been a recent and abrupt change in process conditions, the recent samples may fail to satisfy a similarity criterion that is based purely on “spatial” similarity (e.g., similarity of the Raman scans). Modified JITL techniques that can better leverage the information offered by recent samples (irrespective of spatial similarity), and therefore can better adapt to time-varying process changes, are also described herein. In particular, “adaptive” JITL (A-JITL) and “spatiotemporal” JITL (ST-JITL) techniques for model calibration and maintenance are described herein.

[0012] Real-time model maintenance, in which local models can learn from the latest analytical measurements and thereby adapt quickly to time-varying conditions, can be important to the success of JITL techniques. However, frequent access to analytical instruments/measurements (e.g., analyzing offline samples) tends to be highly resource-intensive. To minimize such resource usage, without overly degrading model performance, a performance-based model maintenance protocol may be implemented in which the system schedules/triggers an analytical measurement in response to determining that the current model performance is unacceptable/unreliable.

[0012a] According to a first aspect of the present invention, there is provided a computer-implemented method for monitoring and/or controlling a biopharmaceutical process, the method comprising:

determining, by one or more processors, a query point associated with scanning of the biopharmaceutical process by a spectroscopy system;

querying, by the one or more processors, an observation database containing a plurality of observation data sets associated with past observations of biopharmaceutical processes, wherein each of the observation data sets includes spectral data and a corresponding actual analytical measurement, and wherein querying the observation database includes selecting as training data, from among the plurality of observation data sets, observation data sets that satisfy one or more relevancy criteria with respect to the query point;

calibrating, by the one or more processors and using the selected training data, a local model specific to the biopharmaceutical process, the local model being trained to predict analytical measurements based on spectral data inputs; and

predicting, by the one or more processors, an analytical measurement of the biopharmaceutical process, wherein predicting the analytical measurement of the biopharmaceutical process includes using the local model to analyze spectral data that the spectroscopy system generated when scanning the biopharmaceutical process.

[0012b] According to a second aspect of the present invention, there is provided a bioreactor system comprising:

- a bioreactor chamber configured for containing a biopharmaceutical process;
- one or more spectroscopy probes collectively configured to (i) deliver source electromagnetic radiation to the biopharmaceutical process and (ii) collect electromagnetic radiation while the source electromagnetic radiation is delivered to the biopharmaceutical process;
- one or more memories collectively storing an observation database containing a plurality of observation data sets associated with past observations of biopharmaceutical processes, wherein each of the observation data sets includes spectral data and a corresponding actual analytical measurement; and
- one or more processors configured to
 - determine a query point associated with scanning of the biopharmaceutical process by the spectroscopy system,
 - query the observation database, at least by selecting as training data, from among the plurality of observation data sets, observation data sets that satisfy one or more relevancy criteria with respect to the query point,
 - calibrate, using the selected training data, a local model specific to the biopharmaceutical process, the local model being trained to predict analytical measurements based on spectral data inputs, and
 - predict an analytical measurement of the biopharmaceutical process, at least by using the local model to analyze spectral data that the spectroscopy system generated when scanning the biopharmaceutical process with the one or more spectroscopy probes.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] The skilled artisan will understand that the figures, described herein, are included for purposes of illustration and are not limiting on the present disclosure. The drawings are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the present disclosure. It is to be understood that, in some instances, various aspects of the described implementations may be shown exaggerated or enlarged to facilitate an understanding of the described implementations. In the drawings, like reference characters throughout the various drawings generally refer to functionally similar and/or structurally similar components.

[0014] FIG. 1 is a simplified block diagram of an example Raman spectroscopy system that may be used to predict analytical measurements of biopharmaceutical processes.

WHAT IS CLAIMED:

1. A computer-implemented method for monitoring and/or controlling a biopharmaceutical process, the method comprising:
 - determining, by one or more processors, a query point associated with scanning of the biopharmaceutical process by a spectroscopy system;
 - querying, by the one or more processors, an observation database containing a plurality of observation data sets associated with past observations of biopharmaceutical processes, wherein each of the observation data sets includes spectral data and a corresponding actual analytical measurement, and wherein querying the observation database includes selecting as training data, from among the plurality of observation data sets, observation data sets that satisfy one or more relevancy criteria with respect to the query point;
 - calibrating, by the one or more processors and using the selected training data, a local model specific to the biopharmaceutical process, the local model being trained to predict analytical measurements based on spectral data inputs; and
 - predicting, by the one or more processors, an analytical measurement of the biopharmaceutical process, wherein predicting the analytical measurement of the biopharmaceutical process includes using the local model to analyze spectral data that the spectroscopy system generated when scanning the biopharmaceutical process.
2. The computer-implemented method of claim 1, wherein the spectroscopy system is a Raman spectroscopy system.
3. The computer-implemented method of claim 1 or 2, wherein:
 - determining a query point includes determining the query point based at least in part on a spectral scan vector, the spectral scan vector being generated by the spectroscopy system when scanning the biopharmaceutical process; and
 - selecting as training data the observation data sets that satisfy one or more relevancy criteria with respect to the query point includes comparing the spectral scan vector on which determination of the query point was based to spectral scan vectors associated with the past observations of the biopharmaceutical processes.
4. The computer-implemented method of claim 3, wherein:
 - determining a query point further includes determining the query point based on a sample number associated with the spectral scan vector; and
 - selecting as training data the observation data sets that satisfy one or more relevancy criteria with respect to the query point includes (i) comparing the spectral scan vector on which determination of the query point was based to spectral scan vectors associated with the past observations of the biopharmaceutical processes, and (ii) comparing the sample number associated with the query point to sample numbers associated with the past observations of the biopharmaceutical processes.
5. The computer-implemented method of claim 4, wherein selecting as training data the observation data sets that satisfy one or more relevancy criteria with respect to the query point includes:
 - selecting the most recent k observation data sets for inclusion in the training data.
6. The computer-implemented method of any one of claims 3 through 5, wherein predicting the analytical measurement of the biopharmaceutical process includes:
 - using the local model to analyze the spectral scan vector on which determination of the query point was based.

7. The computer-implemented method of any one of claims 3 through 6, wherein selecting as training data the observation data sets that satisfy one or more relevancy criteria with respect to the query point includes:

calculating distances between (i) the spectral scan vector on which determination of the query point was based and (ii) the spectral scan vectors associated with the past observations of the biopharmaceutical processes; and

selecting as the training data any of the spectral scan vectors associated with the past observations that are within a threshold distance of the spectral scan vector on which determination of the query point was based.

8. The computer-implemented method of any one of claims 1 through 7, wherein determining a query point includes:

determining the query point based at least in part on one or both of (i) a media profile associated with the biopharmaceutical process, and (ii) one or more operating conditions under which the biopharmaceutical process is analyzed.

9. The computer-implemented method of any one of claims 1 through 8, wherein calibrating a local model specific to the biopharmaceutical process includes:

calibrating a Gaussian process machine-learning model specific to the biopharmaceutical process.

10. The computer-implemented method of any one of claims 1 through 9, wherein calibrating a local model specific to the biopharmaceutical process includes:

calibrating a model that is a function of both spectral data and sample number of a given observation data set.

11. The computer-implemented method of any one of claims 1 through 10, wherein predicting an analytical measurement of the biopharmaceutical process includes:

using the local model to determine a confidence indicator associated with the predicted analytical measurement of the biopharmaceutical process.

12. The computer-implemented method of any one of claims 1 through 11, further comprising:

controlling, by the one or more processors and based at least in part on the predicted analytical measurement of the biopharmaceutical process, at least one parameter of the biopharmaceutical process.

13. The computer-implemented method of any one of claims 1 through 12, wherein the predicted analytical measurement of the biopharmaceutical process is

1) a media component concentration, a media state, a viable cell density, a titer, a critical quality attribute, or a cell state.

2)) a concentration of glucose, lactate, glutamate, glutamine, ammonia, amino acids, Na⁺, or K⁺; or

3) pH, pCO₂, pO₂, temperature or osmolality.

14. The computer-implemented method of any one of claims 1 through 13, further comprising:

obtaining, by an analytical instrument, an actual analytical measurement of the biopharmaceutical process; and

causing, by the one or more processors, (i) spectral data that the spectroscopy system generated when the actual analytical measurement was obtained, and (ii) the actual analytical measurement of the biopharmaceutical process, to be added to the observation database.

15. The computer-implemented method of claim 14, further comprising:
determining, by the one or more processors, that at least the predicted analytical measurement does not satisfy one or more model performance criteria,
wherein obtaining the actual analytical measurement is performed in response to determining that at least the predicted analytical measurement does not satisfy the one or more model performance criteria.
16. The computer-implemented method of claim 15, wherein determining that at least the predicted analytical measurement does not satisfy the one or more model performance criteria includes:
generating a credibility interval associated with the predicted analytical measurement; and
comparing the credibility interval to a pre-defined threshold.
17. The computer-implemented method of any one of claims 1 through 16, wherein the biopharmaceutical process is a cell culture process.
18. A non-transitory computer-readable medium storing instructions for monitoring and/or controlling a biopharmaceutical process, wherein the instructions, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 1 through 17.
19. A bioreactor system comprising:
a bioreactor chamber configured for containing a biopharmaceutical process;
one or more spectroscopy probes collectively configured to (i) deliver source electromagnetic radiation to the biopharmaceutical process and (ii) collect electromagnetic radiation while the source electromagnetic radiation is delivered to the biopharmaceutical process;
one or more memories collectively storing an observation database containing a plurality of observation data sets associated with past observations of biopharmaceutical processes, wherein each of the observation data sets includes spectral data and a corresponding actual analytical measurement; and
one or more processors configured to
determine a query point associated with scanning of the biopharmaceutical process by the spectroscopy system,
query the observation database, at least by selecting as training data, from among the plurality of observation data sets, observation data sets that satisfy one or more relevancy criteria with respect to the query point,
calibrate, using the selected training data, a local model specific to the biopharmaceutical process, the local model being trained to predict analytical measurements based on spectral data inputs, and
predict an analytical measurement of the biopharmaceutical process, at least by using the local model to analyze spectral data that the spectroscopy system generated when scanning the biopharmaceutical process with the one or more spectroscopy probes.
20. The bioreactor system of claim 19, wherein the one or more processors are configured to:
determine the query point based at least in part on a spectral scan vector, the spectral scan vector being generated when scanning the biopharmaceutical process via the one or more spectroscopy probes; and
select the training data at least by comparing the spectral scan vector on which determination of the query point was based to spectral scan vectors associated with the past observations of the biopharmaceutical processes.

21. The bioreactor system of claim 20, wherein the one or more processors are configured to:
 determine the query point based in part on a sample number associated with the spectral scan vector; and
 select as training data the observation data sets that satisfy one or more relevancy criteria with respect to the query point in part by (i) comparing the spectral scan vector on which determination of the query point was based to spectral scan vectors associated with the past observations of the biopharmaceutical processes, and (ii) comparing the sample number associated with the query point to sample numbers associated with the past observations of the biopharmaceutical processes.

22. The bioreactor system of any one of claims 19 to 21, wherein the local model is a Gaussian process machine-learning model.

23. The bioreactor system of any one of claims 19 to 22, wherein the local model is a function of both spectral data and sample number of a given observation data set.

24. The bioreactor system of any one of claims 19 to 23, wherein the one or more processors are further configured to:
 use the local model to determine a confidence indicator associated with the predicted analytical measurement of the biopharmaceutical process.

25. The bioreactor system of any one of claims 19 to 24, wherein the one or more processors are further configured to:
 control, based at least in part on the predicted analytical measurement of the biopharmaceutical process, at least one parameter of the biopharmaceutical process.

26. The bioreactor system of any one of claims 19 to 25, wherein the predicted analytical measurement of the biopharmaceutical process is

- 1) a media component concentration, a media state, a viable cell density, a titer, a critical quality attribute, or a cell state.
- 2) a concentration of glucose, lactate, glutamate, glutamine, ammonia, amino acids, Na^+ , or K^+ , or
- 3) pH, pCO_2 , pO_2 , temperature, or osmolality

27. The bioreactor system of any one of claims 19 to 26, further comprising:
 an analytical instrument configured to obtain an actual analytical measurement of the biopharmaceutical process,
 wherein the one or more processors are further configured to cause (i) spectral data that the spectroscopy system generated when the actual analytical measurement was obtained, and (ii) the actual analytical measurement of the biopharmaceutical process, to be added to the observation database.

28. The bioreactor system of claim 27, wherein the one or more processors are further configured to:
 determine that at least the predicted analytical measurement does not satisfy one or more model performance criteria;
 and

obtain the actual analytical measurement from the analytical instrument in response to determining that at least the predicted analytical measurement does not satisfy the one or more model performance criteria.

29. The bioreactor system of claim 28, wherein the one or more processors are configured to determine that at least the predicted analytical measurement does not satisfy the one or more model performance criteria at least by:
generating a credibility interval associated with the predicted analytical measurement; and
comparing the credibility interval to a pre-defined threshold.