



(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2023.05.12

(51) Int. Cl. **G01N 21/65** (2006.01)
G01N 21/84 (2006.01)

(21) Номер заявки
202191101

(22) Дата подачи заявки
2019.10.23

(54) АВТОМАТИЧЕСКАЯ КАЛИБРОВКА И АВТОМАТИЧЕСКОЕ ОБСЛУЖИВАНИЕ РАМАНОВСКИХ СПЕКТРОСКОПИЧЕСКИХ МОДЕЛЕЙ ДЛЯ ПРЕДСКАЗАНИЙ В РЕАЛЬНОМ ВРЕМЕНИ

(31) 62/749,359; 62/833,044; 62/864,565

(32) 2018.10.23; 2019.04.12; 2019.06.21

(33) US

(43) 2021.08.10

(86) PCT/US2019/057513

(87) WO 2020/086635 2020.04.30

(71)(73) Заявитель и патентовладелец:
ЭМДЖЕН ИНК. (US)

(72) Изобретатель:
Тулсиан Адитиа (US)

(74) Представитель:
Медведев В.Н. (RU)

(56) SU QING-LIN ET AL.: "Just-in-Time-Learning based Extended Prediction Self-Adaptive Control for batch processes", JOURNAL OF PROCESS CONTROL, OXFORD, GB, vol. 43, 10 May 2016 (2016-05-10), pages 1-9, XP029553248, ISSN: 0959-1524, DOI: 10.1016/J.PROCONT.2016.04.009 p. 3, r. col., 1. 12 - p. 4 r. col. 1. 15 p. 7, 1. col., 1. 27, abstract title, figure 7

CHEN LINGYI ET AL.: "A modified recursive locally weighted NIR modeling for fermentation process", 2017 6TH INTERNATIONAL SYMPOSIUM ON

ADVANCED CONTROL OF INDUSTRIAL PROCESSES (ADCONIP), IEEE, 28 May 2017 (2017-05-28), pages 559-564, XP033126064, DOI: 10.1109/ADCONIP.2017.7983841 [retrieved on 2017-07-18], p. 559, 1. col. par. 4 - r. col., par. 1 paragraphs [000I], [00II], [00III], abstract title, figures 1, 2, 6

YI LIU ET AL.: "Just-in-Time Kernel Learning with Adaptive Parameter Selection for Soft Sensor Modeling of Batch Processes", INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH vol. 51, no. 11, 8 March 2012 (2012-03-08), pages 4313-4327, XP055673439, ISSN: 0888-5885, DOI: 10.1021/ie201650u, abstract title, figure 1

US-A1-2018291329

ZHIQIANG GE ET AL.: "A comparative study of just-in-time-learning based methods for online soft sensor modeling", CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS, vol. 104, no. 2, 25 September 2010 (2010-09-25), pages 306-317, XP055673639, NL, ISSN: 0169-7439, DOI: 10.1016/j.chemolab.2010.09.008, table 1

GE ZHIQIANG ET AL.: "Data Mining and Analytics in the Process Industry: The Role of Machine Learning", IEEE ACCESS, vol. 5, 26 September 2017 (2017-09-26), pages 20590-20616, XP011672147, DOI: 10.1109/ACCESS.2017.2756872 [retrieved on 2017-10-24], page 20603, right-hand column

(57) Способ отслеживания биофармацевтического процесса и/или управления им включает определение точки запроса, связанной со сканированием процесса системой спектроскопии (например, системой рамановской спектроскопии), и запрос базы данных наблюдений, содержащей наборы данных наблюдений, связанные с прошлыми наблюдениями биофармацевтических процессов. Каждый из наборов данных наблюдений содержит спектральные данные и соответствующее фактическое аналитическое измерение. Запрос базы данных наблюдений включает выбор в качестве обучающих данных из наборов данных наблюдений тех наборов данных наблюдений, которые удовлетворяют одному или нескольким критериям релевантности относительно точки запроса. Способ также включает использование выбранных обучающих данных для калибровки локальной модели, характерной для биофармацевтического процесса. Локальная модель (например, модель на основе гауссовского процесса) обучается для предсказания аналитических измерений на основе входных спектральных данных. Способ также включает использование локальной модели для предсказания аналитического измерения биофармацевтического процесса.

Перекрестная ссылка на родственные заявки

Испрашивается приоритет предварительной заявки на патент США № 62/749359, поданной 23 октября 2018 г., предварительной заявки на патент США № 62/833044, поданной 12 апреля 2019 г., и предварительной заявки на патент США № 62/864565, поданной 21 июня 2019 г., каждая из которых включена в данный документ посредством ссылки во всей своей полноте.

Область техники, к которой относится изобретение

Настоящая заявка в целом относится к отслеживанию биофармацевтических процессов и/или управлению ими с использованием спектроскопических методик, таких как рамановская спектроскопия, и, в частности, относится к оперативной калибровке и обслуживанию моделей предсказания.

Предпосылки изобретения

Устойчивое производство биотерапевтических белков при помощи биофармацевтического процесса в целом требует поддержания биореактором сбалансированных и согласованных параметров (например, концентраций клеточных метаболитов), что, в свою очередь, требует строгого отслеживания процесса и управления им. Для удовлетворения этих требований все больше применяются средства технологии анализа процессов (PAT). Некоторыми примерами традиционных средств PAT, которые используются в системах управления с обратной связью, являются оперативное отслеживание pH, растворенного кислорода и температуры клеточной культуры. В последние годы были исследованы и развернуты другие встраиваемые в процесс зонды для непрерывного отслеживания более сложных вещей, таких как плотность жизнеспособных клеток (VCD), глюкоза, лактат и другие критические клеточные метаболиты, аминокислоты, титр и критические показатели качества.

Рамановская спектроскопия является популярным средством PAT, широко используемым для оперативного отслеживания в биопроизводстве. Она представляет собой оптический метод, позволяющий выполнять неразрушающий анализ химического состава и молекулярной структуры. В рамановской спектроскопии падающее лазерное излучение рассеивается неупруго вследствие наличия мод молекулярных колебаний. Разность частот между падающими и рассеянными фотонами называется "рамановским сдвигом", и вектор рамановского сдвига в зависимости от уровней интенсивности (называемый в данном документе "рамановским спектром", "результатом рамановского сканирования" или "вектором рамановского сканирования") можно проанализировать с целью определения химического состава и молекулярной структуры образца. Применения рамановской спектроскопии в анализе полимеров, фармацевтическом, биопроизводственном и биомедицинском анализе значительно увеличились в последние три десятилетия по мере совершенствования технологий лазерной выборки и детекторов. Вследствие этих технологических достижений рамановская спектроскопия в настоящее время представляет собой практическую методику анализа, используемую как внутри, так и вне лаборатории. С тех пор, как о применении в биопроизводстве рамановских измерений in-situ сообщили впервые, они были введены в использование для обеспечения оперативных предсказаний в реальном времени некоторых ключевых состояний процессов, таких как глюкоза, лактат, глутамат, глутамин, аммиак, VCD и т.д. Эти предсказания обычно основаны на калибровочной модели или модели мягкого датчика, построенной в автономных условиях на основе аналитических измерений из аналитического прибора. Для установления корреляции рамановского спектра с аналитическими измерениями обычно используют методы моделирования частично наименьших квадратов (PLC) и множественной линейной регрессии. Эти модели обычно требуют фильтрации путем предварительной обработки результатов рамановского сканирования перед калибровкой относительно аналитических измерений. После обучения калибровочной модели ее реализуют в условиях реального времени для обеспечения измерений in-situ с целью отслеживания процесса и/или управления им.

Калибровка рамановской модели для биофармацевтических применений является нетривиальной, так как биофармацевтические процессы обычно действуют в условиях жестких ограничений и норм. В текущем известном уровне техники подход к калибровке рамановской модели в биофармацевтической промышленности заключается в том, что, в первую очередь запускается несколько исследований производственного цикла для генерирования релевантных данных, которые используются для установления корреляции рамановского спектра с аналитическим измерением (измерениями). Эти исследования являются как дорогостоящими, так и трудоемкими, так как каждый производственный цикл может длиться в лабораторных условиях, например, от двух до четырех недель. Кроме того, для аналитических приборов может являться доступным лишь ограниченное количество выборок (например, для того чтобы убедиться, что лабораторный биореактор сохраняет значительную массу жизнеспособных клеток). Фактически нередко случается, что при поточных или автономных аналитических измерениях ежедневно доступно лишь одно или два измерения. Ситуацию дополнительно усугубляет то, что лучшие современные применения дают калибровочные модели, привязанные к конкретному процессу, конкретной формуле или профилю питательной среды биореактора и конкретным рабочим условиям. Поэтому при изменении любой из вышеупомянутых переменных может потребоваться повторная калибровка моделей на основе новых данных. Фактически как калибровка рамановских моделей, так и обслуживание моделей требуют выделения значительных ресурсов и обычно выполняются в автономных условиях. И хотя были предложены подходы, в которых модели адаптируют к новым рабочим условиям (например, рекурсивный ме-

тод, метод скользящего окна и разностный временной метод), эти способы могут быть неспособны адекватно справляться с резкими изменениями в процессе.

Существует несколько публикаций, в которых описаны обобщенные рамановские модели на основе традиционных хемометрических методов (например, моделирование PLS) для множества молекул. Однако в этих обобщенных моделях предполагается, что в процессах используются подобные, если не одинаковые, составы питательной среды и/или рабочие условия процесса. Питательные среды и процессы обычно являются платформенными, с небольшими изменениями или без изменений. Недостатком обобщенной модели указанного типа является то, что, как только процесс отклоняется от нормы, или если набор обучающих данных содержит слишком широкий технологический диапазон в попытке учесть изменения (например, добавки к питательной среде, длительность процесса и/или другие изменения в процессе) между разными молекулами, обобщенные модели теряют точность и достоверность. Поэтому эти "обобщенные" модели являются обобщенными лишь в пределах описанных жестких рамок. См. Mehdi-zahedi и др., *Biotechnol. Prog.* 31(4): 1004-1013, 2015; Webster и др., *Biotechnol. Prog.* 34(3):730-737, 2018.

Краткое описание

Термин "биофармацевтический процесс" относится к процессу, используемому в биофармацевтическом производстве, такому как процесс культивирования клеток с целью получения требуемого рекомбинантного белка. Культивирование клеток происходит в сосуде для культивирования клеток, таком как биореактор, в условиях, которые поддерживают рост и поддержание организма, предназначенного для экспрессии белка. В ходе получения рекомбинантного белка с целью управления и/или поддержания процесса культивирования клеток выполняется отслеживание параметров процесса, таких как концентрации компонентов питательной среды, в том числе питательных веществ и метаболитов (например, глюкозы, лактата, глутамата, глутамина, аммиака, аминокислот, Na^+ , K^+ и других питательных веществ или метаболитов), состояние питательной среды (pH, pCO_2 , pO_2 , температура, осмоляльность и т.д.), а также параметров клеток и/или белков (например, плотности жизнеспособных клеток (VCD), титра, состояния клеток, критических показателей качества и т.д.).

Для устранения некоторых из вышеописанных ограничений лучших современных промышленных применений описанные в данном документе варианты осуществления относятся к системам и способам, совершенствующим традиционные методики спектроскопического анализа биофармацевтических процессов, такие как рамановская спектроскопия. В частности, для построения и обслуживания калибровочных моделей (например, рамановских калибровочных моделей) в реальном времени для биофармацевтических применений используется платформа "Обучение точно в срок" (JTTL). JTTL представляет собой платформу для нелинейного моделирования на основе технологии локального моделирования и выборки из базы данных. В отличие от других способов машинного обучения в JTTL в целом предполагается, что все доступные наблюдения хранятся в центральной базе данных, а модели динамически строятся в реальном времени на основе запроса с использованием наиболее релевантных данных из базы данных. Это позволяет получить хорошее приближение сложной динамики процесса с использованием относительно простых локальных моделей. В условиях инфраструктуры JTTL библиотека может содержать спектральные данные не только для одного процесса, эксплуатируемого в определенных рабочих условиях, но также данные для разных процессов, разных профилей питательной среды и/или разных рабочих условий. Это может значительно сократить время, требуемое для калибровки и обслуживания моделей, особенно для находящихся в разработке лекарств, которые могут иметь небольшую историю или не иметь истории на данный момент.

Платформа JTTL обслуживает динамическую библиотеку, которая может обновляться каждый раз, когда доступно новое аналитическое измерение. Кроме того, для обеспечения адаптации локальных моделей к новым условиям процесса, последнее доступное аналитическое измерение (например, для продукта, отслеживание которого выполняется в настоящий момент) всегда может быть включено в обучающий набор для локального моделирования. Это позволяет быстрее адаптировать локальную модель к новым условиям или к новым линейкам продуктов, не имеющим истории. С использованием этого подхода можно автоматизировать как калибровку модели, так и обслуживание модели, а также можно значительно сократить время и стоимость (например, материальные затраты и трудозатраты), связанные с повседневными калибровками в традиционных системах. Кроме того, возможность предоставления границ правдоподобия (или других индикаторов достоверности, таких как степени достоверности) по предсказаниям на основе модели может обеспечивать робастные стратегии отслеживания и управления.

В некоторых вариантах осуществления для локального моделирования в инфраструктуре JTTL используются модели на основе гауссовского процесса. Модели на основе гауссовского процесса представляют собой мощные статистические модели машинного обучения, которые могут эффективно улавливать сложную нелинейную динамику процессов и могут легко адаптироваться практически к любым изменениям в процессе. В отличие от PLS регрессии основных компонентов (PCR) и регрессионных моделей других типов, модели на основе гауссовского процесса представляют собой непараметрические методы, и они являются намного более приспособленными к улавливанию сложных корреляций между рамановскими спектрами и аналитическими измерениями на основе ограниченных наборов данных. Кроме того, модели на основе гауссовского процесса в целом не требуют фильтрации путем предварительной

обработки результатов рамановского сканирования. Соответственно в некоторых вариантах осуществления модели на основе гауссовского процесса альтернативно калибруются на свежих результатах рамановского сканирования (в логарифмической шкале), что может сэкономить множество этапов в процессе калибровки/обслуживания модели. Кроме того, модели на основе гауссовского процесса предоставляют границы правдоподобия по предсказаниям, получение которых с использованием моделей PLS или PCR может быть чрезвычайно затруднительным. Границы правдоподобия могут быть особенно полезны для разработки оптимальных стратегий взятия проб для аналитических приборов и/или для реализации управления в замкнутом контуре (например, управления по модели предсказания, или MPC), например, во избежание внесения изменений на основе ненадежных предсказаний.

Хотя JTL представляет собой нелинейную инфраструктуру моделирования, и хотя вышеописанный подход предусматривает некоторую адаптивность за счет обновления динамической библиотеки недавними аналитическими измерениями, JTL само по себе может не являться достаточно адаптивным для учета изменяющихся во времени условий процесса (например, резких изменений до заданного значения или других условий процесса). В частности, локальные модели, откалиброванные с использованием JTL, могут быть неспособны использовать недавние выборки. Например и в частности, если имело место недавнее и резкое изменение в условиях процесса, недавние выборки могут быть неспособны удовлетворять критерию подобия, основанному чисто на "пространственном" подобии (например, подобии результатов рамановского сканирования). В данном документе также описаны модифицированные методики JTL, которые могут лучше усреднять информацию, предоставляемую недавними выборками (независимо от пространственного подобия), и поэтому могут лучше адаптироваться к изменяющимся во времени изменениям в процессе. В частности, в данном документе описаны методики "адаптивного" JTL (A-JTL) и "пространственно-временного" JTL (ST-JTL) для калибровки и обслуживания моделей.

Для успеха методик JTL может быть важно обслуживание моделей в реальном времени, при котором локальные модели могут обучаться на новейших аналитических измерениях и, таким образом, быстро адаптироваться к изменяющимся во времени условиям. Однако частый доступ к аналитическим приборам/измерениям (например, при автономном анализе выборок), как правило, является в высокой степени ресурсоемким. Для сведения к минимуму такого использования ресурсов без чрезмерного ухудшения качества модели может быть реализован основанный на качестве протокол обслуживания модели, в котором система планирует/инициирует аналитическое измерение в ответ на определение недопустимо/ненадежного текущего качества модели.

Краткое описание графических материалов

Специалисту в данной области техники понятно, что фигуры, описанные в данном документе, включены для целей иллюстрации, а не ограничения настоящего изобретения. Графические материалы не обязательно изображены в масштабе, акцент делается на иллюстрацию принципов настоящего изобретения. Следует понимать, что в некоторых случаях различные аспекты описанных вариантов реализации могут быть укрупнены или увеличены для улучшения понимания описанных вариантов реализации. На графических материалах подобные ссылочные позиции на разных фигурах в целом выполняют одинаковые функции и/или являются структурно одинаковыми компонентами.

На фиг. 1 представлена упрощенная блок-схема примерной системы рамановской спектроскопии, которая может использоваться для предсказания аналитических измерений биофармацевтических процессов.

На фиг. 2 представлена упрощенная блок-схема примерной системы рамановской спектроскопии, которая может использоваться для предсказания аналитических измерений биофармацевтических процессов для управления в замкнутом контуре концентрацией глюкозы.

На фиг. 3 изображены экспериментальные результаты управления в замкнутом контуре концентрацией глюкозы с использованием примерной реализации системы рамановской спектроскопии, описанной в данном документе.

На фиг. 4 изображен примерный поток данных, который может иметь место при анализе биофармацевтического процесса с использованием методики "Обучение точно в срок" (JTL).

На фиг. 5 изображен примерный поток данных, который может иметь место при анализе биофармацевтического процесса с использованием методики адаптивного JTL (A-JTL).

На фиг. 6 изображен примерный поток данных, который может иметь место при анализе биофармацевтического процесса с использованием методики пространственно-временного JTL (ST-JTL).

На фиг. 7 представлена схема последовательности операций примерного способа анализа биофармацевтического процесса.

Подробное описание

Различные концепции, представленные выше и более подробно обсужденные далее, могут быть реализованы многими способами, и описанные концепции не ограничены каким-либо определенным способом реализации. Примеры вариантов реализации представлены для иллюстративных целей.

На фиг. 1 представлена упрощенная блок-схема примерной системы 100 рамановской спектроскопии, которая может использоваться для предсказания аналитических измерений биофармацевтических процессов. Хотя на фиг. 1 изображена система 100, в которой реализуются методики рамановской спек-

троскопии, понятно, что в других вариантах осуществления в системе 100 могут быть реализованы другие спектроскопические методики, подходящие для анализа биофармацевтических процессов, такие как, например, спектроскопия в ближней инфракрасной области (NIR).

Система 100 содержит биореактор 102, один или несколько аналитических приборов 104, рамановский анализатор 106 с рамановским зондом 108, компьютер 110 и сервер 112 базы данных, соединенный с компьютером 110 через сеть 114. Биореактор 102 может представлять собой любой подходящий сосуд, устройство или систему, поддерживающую биологически активную среду, которая может содержать живые организмы и/или полученные из них вещества (например, культуру клеток) в питательной среде. Биореактор 102 может содержать рекомбинантные белки, экспрессируемые культурой клеток, например, для исследовательских целей, клинического применения, коммерческой продажи или другого распространения. В зависимости от отслеживаемого биофармацевтического процесса питательная среда может содержать определенную текучую среду (например, "бульон") и определенные питательные вещества и может иметь целевые параметры состояния питательной среды, такие как целевой уровень или диапазон pH, целевая температура или температурный диапазон и т.д. Питательная среда может также содержать организмы и вещества, полученные из организмов, такие как метаболиты и рекомбинантные белки. Содержимое и параметры/характеристики питательной среды совместно называются в данном документе "профилем питательной среды".

Аналитический прибор (приборы) 104 может представлять собой поточный, расположенный у потока и/или автономный прибор или приборы, выполненные с возможностью измерения одной или нескольких характеристик или параметров биологически активного содержимого в биореакторе 102 на основе взятых из него проб. Например, аналитический прибор (приборы) 104 может измерять концентрации одного или нескольких компонентов питательной среды, такие как уровни питательных веществ и/или метаболитов (например, глюкозы, лактата, глутамата, глутамина, аммиака, аминокислот, Na^+ , K^+ и т.д.), и параметры состояния питательной среды (pH, pCO_2 , pO_2 , температура, осмоляльность и т.д.). Дополнительно или альтернативно аналитический прибор (приборы) 104 может измерять осмоляльность, плотность жизнеспособных клеток (VCD), титр, критические показатели качества, состояние клеток (например, клеточный цикл) и/или другие характеристики или параметры, связанные с содержимым биореактора 102. В качестве более конкретного примера пробы могут быть взяты, отцентрифугированы, очищены при помощи нескольких колонок и пропущены через первый из аналитических приборов 104 (например, прибор для высокоэффективной жидкостной хроматографии (HPLC) или ультравысокоэффективной жидкостной хроматографии (UPLC)) и затем через второй из аналитических приборов 104 (например, масс-спектрометр), при этом оба, первый и второй, аналитические приборы 104 предоставляют аналитические измерения. В одном, некоторых или всех аналитических приборах 104 могут использоваться разрушающие методики анализа.

Рамановский анализатор 106 может содержать спектрографическое устройство, соединенное с рамановским зондом 108 (или, в некоторых реализациях, несколькими рамановскими зондами). Рамановский анализатор 106 может содержать источник лазерного излучения, обеспечивающий лазерное излучение для рамановского зонда 108 по оптоволоконному кабелю, и может также содержать устройство с зарядовой связью (CCD) или другую подходящую камеру/регистрирующее устройство для регистрации сигналов, принимаемых из рамановского зонда 108, например, по другому каналу оптоволоконного кабеля. Альтернативно источник лазерного излучения может быть встроен в сам рамановский зонд 108. Рамановский зонд 108 может представлять собой погружной зонд или зонд любого другого подходящего типа (например, зонд отражательной способности и зонд пропускания).

Рамановский анализатор 106 и рамановский зонд 108 совместно выполнены с возможностью неразрушающего сканирования биологически активного содержимого в ходе биофармацевтического процесса в биореакторе 102 путем возбуждения, наблюдения и регистрации молекулярных "отпечатков" биофармацевтического процесса. Молекулярные "отпечатки" соответствуют колебательным, вращательным и/или другим низкочастотным модам молекул в биологически активном содержимом в биофармацевтическом процессе при возбуждении содержимого биореактора лазерным излучением, обеспечиваемым рамановским зондом 108. В результате этого процесса сканирования рамановский анализатор 106 генерирует один или несколько векторов рамановского сканирования, каждый из которых представляет собой интенсивность в зависимости от рамановского сдвига (частоты).

Компьютер 110 соединен с рамановским анализатором 106 и аналитическим прибором (приборами) 104 и в целом выполнен с возможностью анализа векторов рамановского сканирования, генерируемых рамановским анализатором 106, с целью предсказания одного или нескольких аналитических измерений биофармацевтического процесса. Например, компьютер 110 может анализировать векторы рамановского сканирования для предсказания аналитического измерения (измерений) одного и того же типа (типов), выполняемого аналитическим прибором (приборами) 104. В качестве более конкретного примера компьютер 110 может предсказывать концентрации глюкозы, тогда как аналитический прибор (приборы) 104 фактически измеряет концентрации глюкозы. Однако в то время как аналитический прибор (приборы) 104 может выполнять относительно нечастые "автономные" аналитические измерения проб, извлекаемых из биореактора 102 (например, по причине ограниченных количеств питательной среды в био-

фармацевтическом процессе и/или по причине более высокой стоимости выполнения таких измерений и т.д.), компьютер 110 может выполнять относительно частые "оперативные" предсказания аналитических измерений в реальном времени. Компьютер 110 также может быть выполнен с возможностью передачи аналитических измерений, выполненных аналитическим прибором (приборами) 104, на сервер 112 базы данных через сеть 114, как будет более подробно обсуждаться ниже.

В примерном варианте осуществления, показанном на фиг. 1, компьютер 110 содержит блок 120 обработки данных, сетевой интерфейс 122, дисплей 124, устройство 126 пользовательского ввода и запоминающее устройство 128. Блок 120 обработки данных содержит один или несколько процессоров, каждый из которых может представлять собой программируемый микропроцессор, который исполняет программные команды, хранящиеся в запоминающем устройстве 128, с целью исполнения некоторых или всех функций компьютера 110, описанных в данном документе. Альтернативно один, несколько или все процессоры в блоке 120 обработки данных могут представлять собой процессоры других типов (например, специализированные интегральные микросхемы (ASIC), вентиляционные матрицы с эксплуатационным программированием (FPGA) и т.д.), и функциональные возможности компьютера 110, описанные в данном документе, альтернативно могут быть частично или полностью реализованы в аппаратном обеспечении. Запоминающее устройство 128 может содержать одно или несколько физических запоминающих устройств или блоков, включающих энергозависимое и/или энергонезависимое запоминающее устройство. Могут использоваться запоминающие устройства любого подходящего типа или типов, такие как постоянное запоминающее устройство (ROM), твердотельные накопители (SSD), накопители на жестких дисках (HDD) и т.д.

Сетевой интерфейс 122 может содержать любое подходящее аппаратное обеспечение (например, клиентское приемопередающее аппаратное обеспечение), программно-аппаратное обеспечение и/или программное обеспечение, выполненное с возможностью установления связи через сеть 114 с использованием одного или нескольких протоколов связи. Например, сетевой интерфейс 122 может представлять собой или содержать интерфейс Ethernet. Сеть 114 может представлять собой одну сеть связи или может содержать несколько сетей связи одного или нескольких типов (например, одну или несколько проводных и/или беспроводных локальных сетей (LAN) и/или одну или несколько проводных и/или беспроводных глобальных сетей (WAN), таких как, например, интернет или корпоративная сеть).

В дисплее 124 может использоваться любая подходящая технология (например, LED, OLED, LCD и т.д.) для представления информации пользователю, а интерфейс 126 пользовательского ввода может представлять собой клавиатуру или другое подходящее устройство ввода. В некоторых вариантах осуществления дисплей 124 и устройство 126 пользовательского ввода объединены в одном устройстве (например, сенсорном дисплее). В целом дисплей 124 и устройство 126 пользовательского ввода могут объединяться для обеспечения возможности взаимодействия пользователя с графическими интерфейсами пользователя (GUI), предоставляемыми компьютером 110, например, для таких целей, как ручное отслеживание различных процессов, исполняемых в системе 100. Однако в некоторых вариантах осуществления компьютер 110 не содержит дисплей 124 и/или устройство 126 пользовательского ввода, или одно или оба из дисплея 124 и устройства 126 пользовательского ввода включены в другой компьютер или систему, которая соединена с возможностью связи с компьютером 110 (например, в некоторых вариантах осуществления, где предсказания отправляются непосредственно в систему управления, реализующую управление в замкнутом контуре).

В запоминающем устройстве 128 хранятся команды одного или нескольких программных приложений, в том числе приложения 130 предсказателя "Обучение точно в срок" (JTTL). Приложение 130 предсказателя JTTL при исполнении блоком 120 обработки данных в целом выполнено с возможностью предсказания аналитических измерений биофармацевтического процесса в биореакторе 102 путем калибровки локальной модели 132 и использования локальной модели 132 для анализа векторов рамановского сканирования, генерируемых рамановским анализатором 106. В зависимости от частоты, с которой рамановский анализатор 106 генерирует данные векторы сканирования, приложение 130 предсказателя JTTL может предсказывать аналитические измерения на периодической или другой подходящей временной основе. Рамановский анализатор 106 может самостоятельно управлять временем генерирования векторов сканирования, или компьютер 110 может инициировать генерирование векторов сканирования путем отправки команды в рамановский анализатор 106. Приложение 130 предсказателя JTTL может предсказывать аналитическое измерение только одного типа на основе каждого вектора сканирования (например, только концентрацию глюкозы) или может предсказывать на основе каждого вектора сканирования несколько типов аналитических измерений (например, концентрацию глюкозы и плотность жизнеспособных клеток). В других вариантах осуществления каждое из нескольких разных приложений предсказателя JTTL (например, каждое из которых аналогично приложению 130 предсказателя JTTL) генерирует отличную локальную модель для предсказания аналитического измерения другого типа, и все они основаны на одном и том же векторе сканирования. Приложение 130 предсказателя JTTL и локальная модель 132 будут более подробно обсуждены ниже.

Сервер 112 базы данных может являться удаленным от компьютера 110 (например, так, что локальная установка может содержать только биореактор 102, аналитический прибор (приборы) 104, раманов-

ский анализатор 106 с рамановским зондом 108 и компьютер 110) и, как видно на фиг. 1, может содержать или являться соединенным с возможностью связи с базой 136 данных наблюдений, в которой хранятся наборы данных наблюдений, связанные с прошлыми наблюдениями. Каждый набор данных наблюдений в базе 136 данных наблюдений может содержать спектральные данные (например, один или несколько векторов рамановского сканирования в виде, получаемом рамановским анализатором 106) и одно или несколько соответствующих аналитических измерений (например, одно или несколько измерений в виде (видах), получаемом аналитическим прибором (приборами) 104). В зависимости от варианта осуществления и/или сценария прошлые наблюдения могут быть собраны для нескольких разных биофармацевтических процессов в нескольких разных рабочих условиях (например, при разных заданных значениях концентрации метаболитов) и/или с несколькими разными профилями питательной среды (например, разными текучими средами, питательными веществами, уровнями pH, температурами и т.д.). В целом может быть желательно, чтобы база 136 данных наблюдений представляла широкое разнообразие процессов, рабочих условий и профилей питательной среды. База 136 данных наблюдений может содержать или не содержать информацию, указывающую эти процессы, клеточные линии, белки, метаболиты, рабочие условия и/или профили питательной среды, однако, в зависимости от варианта осуществления (как дополнительно обсуждено ниже). В некоторых вариантах осуществления сервер 112 базы данных удаленно соединен с несколькими другими компьютерами, аналогичными компьютеру 110, через сеть 114 и/или другие сети. Это может потребоваться для сбора большого количества наборов данных наблюдений с целью сохранения в базе 136 данных наблюдений. Однако в других вариантах осуществления система 100 не содержит сервер 112 базы данных, и компьютер 110 получает доступ непосредственно к локальной базе 136 данных наблюдений.

Понятно, что вместо показанных на фиг. 1 могут использоваться другие конфигурации и/или компоненты. Например, другой компьютер (не показан на фиг. 1) может передавать измерения, предоставляемые аналитическим прибором (приборами) 104, на сервер 112 базы данных, одно или несколько дополнительных вычислительных устройств или систем могут действовать в качестве промежуточных звеньев между компьютером 110 и сервером 112 базы данных, некоторые или все функциональные возможности компьютера 110, описанные в данном документе, могут альтернативно выполняться удаленно сервером 112 базы данных и/или другим удаленным сервером и т.д.

В ходе режима выполнения системы 100 рамановский анализатор 106 и рамановский зонд 108 используются для сканирования (т.е. генерирования векторов рамановского сканирования для) биофармацевтического процесса в биореакторе 102, и вектор (векторы) рамановского сканирования затем передается (передаются) из рамановского анализатора 106 на компьютер 110. Рамановский анализатор 106 и рамановский зонд 108 могут предоставлять векторы сканирования для поддержки предсказаний (выполняемых приложением 130 предсказателя JTTL) в соответствии с предварительно определенным планом периодов отслеживания, как, например, один раз в минуту или один раз в час и т.д. Альтернативно предсказания могут выполняться через неравномерные промежутки (например, в ответ на определенный иницирующий фактор на основе процесса, такой как изменение в измеренном уровне pH и/или температуре), таким образом каждый период отслеживания имеет переменную или неопределенную длительность. В зависимости от варианта осуществления рамановский анализатор 106 может отправлять на компьютер 110 только один вектор сканирования за один период отслеживания или несколько векторов сканирования на компьютер 110 за один период отслеживания, в зависимости от того, сколько векторов сканирования локальная модель 132 принимает в качестве элемента входных данных для одного предсказания. Например, несколько векторов сканирования могут повышать точность предсказания локальной модели 132.

Блок 140 запроса приложения 130 предсказателя JTTL использует вектор (векторы) сканирования, принятые для одного периода отслеживания, с целью генерирования точки запроса, которая будет использоваться для запроса базы 136 данных наблюдений. В некоторых вариантах осуществления точка запроса (т.е. данные, задающие точку запроса) содержит только данные, представляющие вектор (векторы) рамановского сканирования, принятый (принятые) из рамановского анализатора 106 (например, строки интенсивностей/частот, которые содержит каждый вектор сканирования). В других вариантах осуществления точка запроса также содержит информацию одного или нескольких других типов. Например, точка запроса может также содержать данные, представляющие рабочие условия, связанные с процессом (например, заданное значение концентрации метаболита в системе управления или длина волны и/или интенсивность лазерного излучения, связанная с рамановским анализатором 106 или рамановским зондом 108, и т.д.), данные, представляющие профиль питательной среды для питательной среды биофармацевтического процесса (например, тип текучей среды, типы или концентрации питательных веществ, уровень pH и т.д.) и/или другие данные (например, индикаторы клеточных линий, белков или метаболитов, связанных с биофармацевтическим процессом).

В целом точка запроса может содержать данные, представляющие те же векторы, параметры и/или классификации, которые локальная модель 132 использует в качестве входных данных (т.е. в качестве набора признаков локальной модели 132). Использование для набора признаков нескольких разных типов данных может повышать точность предсказаний аналитических измерений, выполняемых локальной

моделью 132. Однако, поскольку в целом требуется, чтобы каждый набор данных наблюдений в базе 136 данных наблюдений содержал одни и те же векторы, параметры и/или характеристики в качестве набора признаков, может являться предпочтительным ограничение точки запроса, и включение в набор признаков/входные данные локальной модели 132 только одного или нескольких векторов рамановского сканирования. Это может предоставлять различные преимущества, такие как возможность сбора большего количества информации для хранения в базе 136 данных наблюдений и/или упрощение сбора этой информации. Например, если используются только векторы рамановского сканирования, наборы данных наблюдений могут быть включены в базу 136 данных наблюдений, даже если о процессах, клеточных линиях, белках, метаболитах, рабочих условиях и/или профилях питательной среды, которые существовали на момент сбора наборов данных, известно мало или ничего не известно.

Блок 140 запроса затем запрашивает базу 136 данных наблюдений с использованием сгенерированной точки запроса. В примерном варианте осуществления, представленном на фиг. 1, блок 140 запроса выполняет запрос, обеспечивая передачу сетевым интерфейсом 122 точки запроса (например, в сообщении запроса) на сервер 112 базы данных через сеть 114, что, в свою очередь, обеспечивает извлечение сервером 112 базы данных соответствующих данных из базы 136 данных наблюдений. Однако в вариантах осуществления, где база 136 данных наблюдений альтернативно включена в компьютер 110 (или в запоминающее устройство, соединенное с ним с возможностью связи), блок 140 запроса может альтернативно запрашивать базу 136 данных наблюдений более непосредственно. Для простоты объяснения остальное описание фиг. 1 будет предполагать, что база 136 данных наблюдений соединена с сервером 112 базы данных, как изображено на фиг. 1. Однако специалисту в данной области техники будет легко понятно, как могут отличаться каналы связи, если база 136 данных наблюдений альтернативно расположена локально по отношению к компьютеру 110 или в другом подходящем местоположении в архитектуре системы.

После приема точки запроса сервер 112 базы данных использует точку запроса для выбора из базы 136 данных наблюдений релевантных наборов данных наблюдений, которые будут полезны в качестве обучающих данных для локальной модели 132. Для идентификации того, какие наборы данных наблюдений являются "релевантными", сервер 112 базы данных может применять любые подходящие критерии релевантности, в зависимости от варианта осуществления. В одном варианте осуществления, например, точка запроса содержит один вектор рамановского сканирования, и сервер 112 базы данных определяет, является ли данный набор данных наблюдений релевантным, путем вычисления евклидова расстояния между вектором рамановского сканирования этого набора данных наблюдений и вектором рамановского сканирования точки запроса. Если евклидово расстояние находится ниже некоторого предварительно определенного порогового значения (или ниже переменного порогового значения, такого как пороговое значение, вычисленное на основе среднего евклидова расстояния между вектором сканирования точки запроса и векторами сканирования всех наборов данных наблюдений и т.д.), набор данных наблюдений идентифицирован как релевантный набор данных наблюдений. Специалисту в данной области техники будет понятно, каким образом можно легко распространить подход на варианты осуществления, в которых точка запроса (и каждый набор данных наблюдений) содержит несколько векторов рамановского сканирования. В некоторых ситуациях использование евклидова расстояния для выбора релевантных наборов данных наблюдений может являться субоптимальной методикой. Однако, если локальная модель 132 представляет собой модель на основе гауссовского процесса (как обсуждено ниже), использование евклидова расстояния в качестве критерия релевантности может являться особенно предпочтительным. Причиной этого является то, что модели на основе гауссовского процесса с радиальными базисными функциями или квадратичными экспоненциальными ядрами сами основаны на евклидовом расстоянии. Тем не менее, в других вариантах осуществления могут применяться другие критерии релевантности (например, критерии на основе углов или на основе коэффициентов корреляции и т.д.). Понятно, что в вариантах осуществления, где локальная модель 132 также допускает другую информацию в качестве элемента входных данных/набора признаков (например, рабочие условия, профиль питательной среды, данные процесса, информацию о клеточных линиях, информацию о белках и/или информацию о метаболитах и т.д.), для идентификации "релевантных" наборов данных наблюдений могут использоваться более сложные методики. В некоторых вариантах осуществления сервер 112 базы данных выбирает только предварительно определенное количество релевантных наборов данных наблюдений в ответ на один запрос или выбирает не больше некоторого максимально допустимого количества релевантных наборов данных наблюдений с целью обеспечения извлечения лишь относительно небольшого поднабора из всех наборов данных в базе 136 данных наблюдений. Однако в других вариантах осуществления сервер 112 базы данных может выбирать любое количество релевантных наборов данных наблюдений до тех пор, пока для каждого такого набора данных наблюдений удовлетворяются критерии релевантности.

В некоторых вариантах осуществления, как будет более подробно описано ниже (например, со ссылкой на фиг. 5 и 6), релевантные наборы данных наблюдений выбираются не только на основе релевантности точки запроса в "пространственном" смысле (например, подобия векторов рамановского сканирования), но также на основе релевантности во временном смысле (например, того, какие наборы дан-

ных являются самыми недавними, независимо от пространственного подобия). В этих методиках может быть более преимущественно задействован тот факт, что более недавние аналитические измерения могут предоставлять полезную информацию, даже если эти недавние измерения соответствуют другому заданному значению, и т.д.

После идентификации релевантных наборов данных наблюдений (каждый из которых может соответствовать или может не соответствовать таким же условиям процесса, как у биофармацевтического процесса в биореакторе 102, отслеживание которого выполняется в настоящий момент) сервер 112 базы данных извлекает эти наборы данных (например, векторы рамановского сканирования и соответствующее аналитическое измерение (измерения)) и передает извлеченные наборы данных на компьютер 110 через сеть 114. Блок 140 запроса затем может передавать релевантные наборы данных на генератор 142 локальной модели, и генератор 142 локальной модели использует релевантные наборы данных в качестве обучающих данных для калибровки локальной модели 132. То есть генератор 142 локальной модели использует вектор (векторы) рамановского сканирования (и, возможно, другие данные), связанный с каждым набором данных наблюдений, в качестве набора признаков и использует аналитическое измерение (измерения), связанное с тем же набором данных наблюдений, в качестве метки для этого набора признаков.

В некоторых вариантах осуществления, как указано выше, генератор 142 локальной модели строит модель на основе гауссовского процесса с целью эффективного улавливания сложной нелинейной динамики процесса и быстрой адаптации практически к любым изменениям в процессе. В отличие от моделей PLS и PCR, в моделях на основе гауссовского процесса используются непараметрические методы, и они являются намного более приспособленными к улавливанию сложных нелинейных корреляций между векторами рамановского сканирования и аналитическими измерениями даже при использовании весьма ограниченного количества обучающих выборок. Это может являться особенно важным в сценариях, где новые продукты или процессы соответствуют лишь ограниченному количеству наборов данных в базе 136 данных наблюдений. В таких сценариях модель на основе гауссовского процесса в целом способна извлекать наибольшее количество информации из таких ограниченных наборов данных в сочетании с другими релевантными наборами данных, которые сервер 112 базы данных извлекает из базы 136 данных наблюдений. Однако в других вариантах осуществления генератор 142 локальной модели может альтернативно строить модель машинного обучения любого другого подходящего типа (например, рекурсивную нейронную сеть, сверточную нейронную сеть и т.д.) до тех пор, пока время обучения не превысит минимальную желаемую длительность периода отслеживания. Генератор 142 локальной модели может также строить локальную модель 132 так, что локальная модель 132 может выводить границы правдоподобия или любой другой подходящий индикатор достоверности предсказания (например, степень достоверности). По меньшей мере по сравнению с моделями PLS и PCR, модели на основе гауссовского процесса особенно хорошо подходят для представления границ правдоподобия по предсказаниям аналитических измерений. Хотя были описаны различные преимущества моделей на основе гауссовского процесса перед моделями PLS и PCR, понятно, что в некоторых вариантах осуществления генератор 142 локальной модели может использовать методы моделирования PLS или PCR для построения локальной модели 132.

Генератор 142 локальной модели может строить локальную модель 132 оперативным образом в реальном времени так, что блок 144 предсказания может затем использовать обученную локальную модель 132 для предсказания одного или нескольких аналитических измерений биофармацевтического процесса путем обработки того же вектора (векторов) рамановского сканирования, который был использован блоком 140 запроса для генерирования точки запроса. Более того, в некоторых вариантах осуществления блок 140 запроса может выполнять новый запрос и генератор 142 локальной модели может генерировать новую версию локальной модели 132 каждый раз, когда рамановский анализатор 106 предоставляет на компьютер 110 новый вектор рамановского сканирования (или новый набор векторов рамановского сканирования). В других вариантах осуществления, однако, блок 140 запроса выполняет новый запрос (и генератор 142 локальной модели генерирует новую версию локальной модели 132) на менее частой основе, как, например, один раз каждые 10 предсказаний/периодов отслеживания или один раз каждые 100 предсказаний/периодов отслеживания и т.д.

Блок 146 обслуживания базы данных также может обеспечивать периодический сбор аналитическим прибором (приборами) 104 данных одного или нескольких аналитических измерений с частотой, которая значительно меньше периода отслеживания рамановского анализатора 106 (например, лишь один или два раза в день и т.д.). Измерение (измерения) аналитическим прибором (приборами) 104 может являться разрушающим в некоторых вариантах осуществления и требующим полного удаления пробы из процесса в биореакторе 102. В момент или приблизительно в момент, в который блок 146 обслуживания базы данных обеспечивает сбор и предоставление аналитическим прибором (приборами) 104 данных фактического аналитического измерения (измерений), блок 146 обслуживания базы данных также может обеспечивать предоставление рамановским анализатором 106 одного или нескольких векторов рамановского сканирования. Блок 146 обслуживания базы данных затем может обеспечивать отправку сетевым интерфейсом 122 вектора (векторов) рамановского сканирования и соответствующего фактиче-

ского аналитического измерения (измерений) на сервер 112 базы данных через сеть 114 для хранения в качестве нового набора данных наблюдений в базе 136 данных наблюдений. База 132 данных наблюдений может обновляться в соответствии с любым подходящим расчетным временем, которое может изменяться в зависимости от варианта осуществления. Если аналитический прибор (приборы) 104 выводит (выводят) фактические аналитические измерения через несколько секунд после измерения пробы, например, база 132 данных наблюдений может обновляться новыми измерениями почти сразу же после взятия проб. В некоторых других вариантах осуществления фактические аналитические измерения могут происходить в течение минут, часов или даже дней обработки одним или несколькими аналитическими приборами 104, и в этом случае база 132 данных наблюдений не обновляется до тех пор, пока обработка не будет завершена. В еще одних вариантах осуществления новые наборы данных наблюдений могут добавляться в базу 132 данных наблюдений постепенно по мере того, как разные аналитические приборы 104 завершают их соответствующие измерения.

Таким образом, база 136 данных наблюдений предоставляет "динамическую библиотеку" прошлых наблюдений, к которой генератор 142 локальной модели может обращаться для обучения модели. В некоторых вариантах осуществления новейшее аналитическое измерение (измерения) всегда добавляется (добавляются) в базу 136 данных наблюдений, и генератор 142 локальной модели всегда может использовать самый недавний набор (наборы) данных наблюдений в базе 136 данных наблюдений при калибровке локальной модели 132. Это может обеспечивать возможность декодирования локальной моделью 132 информации о процессе из недавнего прошлого и быстро адаптироваться к новым условиям или быстро адаптироваться к условиям нового процесса, не имеющего истории. Кроме того, и калибровка, и обслуживание локальной модели 132 могут быть автоматизированы. В некоторых вариантах осуществления возможность адаптации локальной модели 132 дополнительно увеличивается, например, как обсуждено ниже в связи с методиками A-JITL и ST-JITL.

В некоторых вариантах осуществления блок 146 обслуживания базы данных может обеспечивать сбор и предоставление аналитическим прибором (приборами) 104 данных фактического аналитического измерения (измерений) на какой-либо другой временной основе или условии, таком как, например, текущее качество модели. Например, если локальная модель 132 выводит интервал правдоподобия (например, диапазон значений по предсказанному значению, в пределах которого имеется 95% вероятность или достоверность того, что в него попадет фактическое/измеренное значение) или какой-либо другой индикатор достоверности наряду с предсказанием (например, если локальная модель 132 представляет собой модель на основе гауссовского процесса), и если индикатор достоверности показывает особенно ненадежное предсказание (например, если интервал/диапазон превышает пороговое значение ширины/диапазона, и т.д.), то блок 146 обслуживания базы данных может инициировать сбор данных одного или нескольких фактических аналитических измерений. В качестве более конкретного примера блок 146 обслуживания базы данных может инициировать сбор данных аналитического измерения (измерений) в ответ на определение превышения интервалом с 95% правдоподобием предварительно заданного порогового значения. Оптимальное планирование аналитических измерений более подробно обсуждено ниже. После выполнения измерения (измерений) блок 146 обслуживания базы данных может обеспечивать генерирование рамановским анализатором 106 одного или нескольких векторов рамановского сканирования и обеспечивать предоставление сетевым интерфейсом 122 фактического аналитического измерения (измерений) и соответствующего вектора (векторов) рамановского сканирования на сервер 112 базы данных для хранения в качестве нового набора данных наблюдений в базе 132 данных наблюдений (например, способом, обсужденным выше). Генератор 142 локальной модели может затем использовать этот новейший набор данных наблюдений, если это необходимо (например, в зависимости от релевантности текущему запросу, или от того, всегда ли в варианте осуществления используется самый недавний набор данных наблюдений), при калибровке локальной модели 132.

Некоторые или все вышеописанные процессы могут повторяться несколько раз в течение времени выполнения биофармацевтического процесса в биореакторе с целью непрерывного отслеживания процесса с использованием локальной модели, для которой как калибровка, так и обслуживание являются полностью автоматизированными и выполняемыми в реальном времени. Аналитическое измерение (измерения) может предсказываться с разными целями в зависимости от варианта осуществления и/или сценария. Например, некоторые параметры могут быть отслежены (т.е. предсказаны) как часть процесса управления качеством, для того чтобы убедиться в том, что процесс по-прежнему соответствует действующим нормам. В качестве другого примера один или несколько параметров могут отслеживаться или предсказываться для обеспечения обратной связи в системе управления в замкнутом контуре. Например, на фиг. 2 изображена система 150, подобная системе 100, но в которой предпринята попытка управления концентрацией глюкозы в биофармацевтическом процессе (т.е. попытка обеспечить совпадение предсказанной концентрации глюкозы с требуемым заданным значением в пределах некоторого допустимого отклонения). Понятно, что в других вариантах осуществления система 150 может альтернативно (или дополнительно) использоваться для управления параметрами процесса, отличными от уровня глюкозы, или для управления уровнем глюкозы на основе предсказаний одного или нескольких других параметров процесса (например, уровня лактата). На фиг. 2 для указания компонентов, соответствующих компонен-

там, представленным на фиг. 1, используются такие же ссылочные позиции. Например, приложение 130 предсказателя JTTL, представленное на фиг. 2, может являться таким же, как приложение 130 предсказателя JTTL, представленное на фиг. 1 (при этом различные блоки приложения 130 предсказателя JTTL для ясности не показаны на фиг. 2).

Как видно на фиг. 2, в системе 150 в запоминающем устройстве 128 также хранится блок 152 управления. Блок 152 управления выполнен с возможностью управления насосом 154 глюкозы, т.е. с возможностью обеспечения выборочного введения насосом 154 глюкозы дополнительной глюкозы в биофармацевтический процесс в биореакторе 102. Блок 152 управления может содержать программные команды, исполняемые, например, блоком 120 обработки данных, и/или соответствующим программно-аппаратным обеспечением, и/или аппаратным обеспечением. В некоторых вариантах осуществления в блоке 152 управления реализуется методика управления по модели предсказания (MPC) с использованием концентраций глюкозы в качестве входных данных в архитектуру с замкнутым контуром. В вариантах осуществления, где локальная модель 132 предоставляет границы правдоподобия или другие индикаторы достоверности вместе с каждым предсказанием (например, в некоторых вариантах осуществления, где локальная модель 132 представляет собой модель на основе гауссовского процесса), блок 152 управления также может принимать индикаторы достоверности в качестве входных данных. Например, блок 152 управления может генерировать только управляющие команды для насоса 154 глюкозы на основе предсказаний концентрации глюкозы, имеющих достаточно высокий индикатор достоверности (например, только на основе предсказаний, связанных с границами правдоподобия, которые не превышают некоторую процентную долю или абсолютный диапазон измерений, или только на основе предсказаний, связанных со степенями достоверности выше некоторого минимального порогового значения степени, и т.д.), или может увеличивать и/или уменьшать весовой коэффициент данного предсказания на основе его индикатора достоверности и т.д.

На фиг. 3 изображены экспериментальные результаты 200 для одной примерной реализации, в которой для калибровки и обслуживания локальной модели на основе гауссовского процесса используются методики JTTL. На графике, представленном на фиг. 3, горизонтальная штриховая линия 202 представляет заданное значение концентрации глюкозы, кружки 204 представляют фактические измерения концентрации глюкозы (например, выполненные аналитическим прибором, аналогичным одному из аналитических приборов 104, представленных на фиг. 1), сплошная линия 206 представляет предсказанные измерения концентрации глюкозы (например, предсказанные моделью, аналогичной локальной модели 132), а затененные области 208 представляют границы правдоподобия (для правдоподобия 95%), связанные с предсказанными измерениями. Как видно на фиг. 3, для заданного значения концентрации глюкозы, равного 3 граммам на литр (г/л), предсказания, выполненные с использованием методики JTTL, в целом близко совпадают с аналитическими измерениями.

Процесс проведения запроса и построения/калибровки локальной модели 132 ниже будет более подробно описан математически со ссылкой на один конкретный вариант осуществления JTTL, в котором локальная модель 132 представляет собой модель на основе гауссовского процесса, в которой в качестве элемента входных данных используется один вектор рамановского сканирования, и которая предсказывает одно аналитическое измерение.

Пусть $D = \{b_j, a_j\}_{j=1}^J$ (или $D = \{\bar{b}, \bar{a}\}$ в сокращенной записи) обозначает набор упорядоченных пар входных и выходных данных, так что $\bar{a} \equiv \{a_1, a_2, \dots, a_J\}$ - входные данные, и $\bar{b} \equiv \{b_1, b_2, \dots, b_J\}$ - выходные данные. Кроме того, предполагается, что $a_j \in \mathbb{R}^{n_a}$ - n_a -мерный входной вектор, и $b_j \in \mathbb{R}$ - скалярный вывод. Физически $a_j \in \mathbb{R}^{n_a}$ можно представить себе как спектроскопическое измерение (например, NIR или рамановское), а $b_j \in \mathbb{R}$ - как аналитическое измерение для представляющего интерес состояния (например, концентрации глюкозы или лактата). Для данного набора D обучающих данных целью задачи калибровки спектроскопической модели является определение взаимосвязи между входными данными и выходными данными для модели в форме

$$b_j = f(a_j) + \epsilon_j \text{ Уравнение (1)}$$

где $f \in \mathbb{R}$ - спектроскопическая модель и $\epsilon_j \sim N(0, \sigma^2)$ - нормально распределенный шум измерения с нулевым средним, дисперсия σ^2 которого неизвестна. Стандартной практикой при калибровке моделей является предположение того, что $f(\bullet)$ является линейной, а затем использование для обучения модели таких методов, как PLS. Вместо приписывания $f(\bullet)$ какой-либо ограничивающей или фиксированной формы здесь предполагается, что $f(\bullet)$ представляет собой скрытую функцию, моделируемую как гауссовский процесс так, что $f(\bar{a}) \equiv (f(a_1), f(a_2), \dots, f(a_J)) \sim GP(\mu_\theta(\bar{a}), k_\theta(\bar{a}, \bar{a}))$ представляет случайную выборку из гауссовского процесса, имеющего среднее $\mu_\theta(\bullet) \in \mathbb{R}^J$ и ковариационную функцию $k_\theta(\bullet, \bullet) \in \mathbb{R}^{J \times J}$, которые обычно задаются следующим образом:

$$\mu_{\theta}(\bar{\mathbf{a}}) \equiv [\mu_{\theta}(\mathbf{a}_1), \mu_{\theta}(\mathbf{a}_2), \dots, \mu_{\theta}(\mathbf{a}_j)]^T, \text{ Уравнение (2a)}$$

$$k_{\theta}(\bar{\mathbf{a}}; \bar{\mathbf{a}}) \equiv \begin{bmatrix} k_{\theta}(\mathbf{a}_1, \mathbf{a}_1) & k_{\theta}(\mathbf{a}_1, \mathbf{a}_2) & \dots & k_{\theta}(\mathbf{a}_1, \mathbf{a}_j) \\ k_{\theta}(\mathbf{a}_2, \mathbf{a}_1) & k_{\theta}(\mathbf{a}_2, \mathbf{a}_2) & \dots & k_{\theta}(\mathbf{a}_2, \mathbf{a}_j) \\ \vdots & \vdots & \ddots & \vdots \\ k_{\theta}(\mathbf{a}_j, \mathbf{a}_1) & k_{\theta}(\mathbf{a}_j, \mathbf{a}_2) & \dots & k_{\theta}(\mathbf{a}_j, \mathbf{a}_j) \end{bmatrix}. \text{ Уравнение (2b)}$$

Кроме того, $\theta \in \mathbb{R}^{n_{\theta}}$ обозначает гиперпараметры для модели на основе гауссовского процесса. Гауссовский процесс представляет собой набор случайных переменных, любое конечное количество которых имеет совместное гауссовское распределение, так что для набора конечных входных данных $\bar{\mathbf{a}} \equiv \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_j\}$ можно записать:

$$p(\mathbf{f}|\bar{\mathbf{a}}) = \mathcal{N}(\mu_{\theta}(\bar{\mathbf{a}}), \mathbf{k}_{\theta}(\bar{\mathbf{a}}, \bar{\mathbf{a}})) \text{ Уравнение (3)}$$

Тогда задача калибровки спектроскопической модели сводится к обучению скрытой функции $f \in \mathbb{R}$ гауссовского процесса с использованием D. Для удобства с математической точки зрения и общей лаконичности здесь предполагается, что $\mu_{\theta} = 0_{n_a}$; однако это не обязательно имеет место в общем случае и результаты здесь можно легко распространить на модели при $\mu_{\theta} \neq 0_{n_a}$. Роль ковариационной функции в гауссовских процессах аналогична роли ядер, используемых в машинах опорных векторов (SVM). Общеупотребительной ковариационной функцией является гауссовское ядро, и она имеет вид:

$$k_{\theta}(\mathbf{a}_i, \mathbf{a}_j) = \beta \exp\left(-\frac{1}{2} \sum_{l=1}^{n_a} \frac{(a_i^{(l)} - a_j^{(l)})^2}{\alpha_l}\right), \text{ Уравнение (4)}$$

где $k_{\theta}(\mathbf{a}_i, \mathbf{a}_j) \in \mathbb{R}_+$ - ковариация между парой элементов входных данных, $\{\mathbf{a}_i, \mathbf{a}_j\}$. Гауссовское ядро $k_{\theta}(\mathbf{a}_i, \mathbf{a}_j)$ присваивает более высокий коэффициент корреляции, если входные данные в наборе $\{\mathbf{a}_i, \mathbf{a}_j\}$ "близки" друг к другу, что задается евклидовым расстоянием в уравнении (4).

Для выбранного гауссовского ядра уравнение (4) представляет собой положительно определенную симметричную матрицу, так что $\mathbf{k}_{\theta}(\cdot, \cdot) \in \mathbb{S}_{++}^{J \times J}$. В уравнении (4) набор $\theta \equiv \{\beta, \{\alpha_l\}_{l=1}^{n_a}\}$ представляет собой набор гиперпараметров.

Физически $\alpha_l \in \mathbb{R}_+$ представляет собой параметр линейных размеров и $\beta \in \mathbb{R}_+$ - параметр дисперсии сигнала. Выбор гауссовской ковариационной функции в уравнении (4) соответствует предыдущему предположению о том, что f является гладкой и непрерывной. Таким образом, путем варьирования гиперпараметров ковариационной функции можно варьировать "гладкость" f . Здесь предполагаются гауссовские процессы с гауссовской ковариационной функцией. Однако это не обязательно имеет место в общем случае.

Для данного D целью является обучение гиперпараметрам гауссовского процесса, включая любые другие неизвестные параметры модели. Для гауссовского процесса в уравнении (1) набором неизвестных параметров является $\gamma \equiv \{\theta, \sigma^2\} \in \Gamma \subseteq \mathbb{R}^{n_{\gamma}}$. Этап обучения параметров может выполняться путем максимизации функции предельного правдоподобия (или доказательства) в пространстве неизвестных параметров. Например, для гауссовского процесса в уравнении (1) функция предельного правдоподобия имеет следующий вид:

$$p(\bar{\mathbf{b}}|\bar{\mathbf{a}}) = \int p(\bar{\mathbf{b}}|\mathbf{f}, \bar{\mathbf{a}}) p(\bar{\mathbf{f}}|\bar{\mathbf{a}}) d\mathbf{f}, \text{ Уравнение (5)}$$

где $p(\bar{\mathbf{b}}|\bar{\mathbf{a}})$ - функция предельного правдоподобия, $p(\bar{\mathbf{b}}|\mathbf{f}, \bar{\mathbf{a}})$ - функция правдоподобия, имеющая вид:

$$p(\bar{\mathbf{b}}|\mathbf{f}, \bar{\mathbf{a}}) = \mathcal{N}(\mathbf{f}|\bar{\mathbf{a}}, \sigma^2 \mathbf{I}_{J \times J}), \text{ Уравнение (6)}$$

и $p(\bar{\mathbf{f}}|\bar{\mathbf{a}})$ - предыдущая функция плотности, данная в уравнении (3). Для гауссовского правдоподобия и предыдущих плотностей в уравнениях (6) и (3), соответственно, интеграл в уравнении (5) имеет решение в замкнутой форме, так что функция предельного правдоподобия имеет вид:

$$p(\bar{\mathbf{b}}|\bar{\mathbf{a}}) = \mathcal{N}(\mathbf{0}_J, \mathbf{k}_{\theta}(\bar{\mathbf{a}}, \bar{\mathbf{a}}) + \sigma^2 \mathbf{I}_{J \times J}). \text{ Уравнение (7)}$$

Теперь для данного уравнения (7) $\gamma \equiv \{\theta, \sigma^2\} \in \Gamma \subseteq \mathbb{R}^{n_{\gamma}}$ можно оценить, решив следующую задачу оптимизации:

$$\gamma^* \in \arg \max \log p(\bar{\mathbf{b}}|\bar{\mathbf{a}}), \text{ Уравнение (8)}$$

где $\gamma^* \in \Gamma$ - оптимальная оценка. Из уравнения (7) имеем

$$\log p(\bar{\mathbf{b}}|\bar{\mathbf{a}}) = -\frac{1}{2} \bar{\mathbf{b}}^T \mathbf{k}_{\gamma}^{-1} \bar{\mathbf{b}} - \frac{1}{2} \log |\mathbf{k}_{\gamma}| - \frac{1}{2} \log 2\pi, \text{ Уравнение (9)}$$

где $\mathbf{k}_{\gamma} \equiv \mathbf{k}_{\theta}(\bar{\mathbf{a}}|\bar{\mathbf{a}}) + \sigma^2 \mathbf{I}_{J \times J}$. Для решения задачи оптимизации в уравнении (8) частные производные в уравнении (9) определяют относительно γ , так что для всех $r=1, 2, \dots, n_{\gamma}$,

$$\frac{\partial}{\partial \gamma_r} \log p(\bar{\mathbf{b}}|\bar{\mathbf{a}}) = \frac{1}{2} \bar{\mathbf{b}}^T \mathbf{k}_{\gamma}^{-1} \frac{\partial \mathbf{k}_{\gamma}}{\partial \gamma_r} \mathbf{k}_{\gamma}^{-1} \bar{\mathbf{b}} - \frac{1}{2} \text{Tr} \left[\mathbf{k}_{\gamma}^{-1} \frac{\partial \mathbf{k}_{\gamma}}{\partial \gamma_r} \right], \text{ Уравнение (10a)}$$

$$= \frac{1}{2} \text{Tr} \left((\alpha \alpha^T - \mathbf{k}_{\gamma}^{-1} \frac{\partial \mathbf{k}_{\gamma}}{\partial \gamma_r}) \right), \text{ Уравнение (10b)}$$

где $\alpha = \mathbf{k}_{\gamma}^{-1} \bar{\mathbf{b}}$. Для данной функции предельного правдоподобия в уравнении (7) и ее производных в уравнении (10b) при решении уравнения (8) можно использовать метод градиентного спуска. Поскольку уравнение (8) в целом представляет собой задачу невыпуклой оптимизации с множественными локальными оптимумами, при решении этой задачи оптимизации необходимо проявлять осмотрительность. Здесь предполагается, что γ^* известна и может быть вычислена путем решения уравнения (8). Кроме того, для уменьшения сложности записи здесь будет предполагаться, что γ - оптимальная оценка γ^* , если не

указано иное.

После обучения спектроскопической калибровочной модели на основе гауссовского процесса в уравнении (1) ее можно использовать для применений при предсказаниях в реальном времени. Как и ранее, пусть D будет набором обучающих данных, используемым для обучения модели на основе гауссовского процесса, и пусть $\mathbf{a}^* \in \mathbb{R}^{n_a}$ будет новым испытательным спектроскопическим сигналом. Тогда целью является предсказание элемента выходных данных $\mathbf{b}^* \in \mathbb{R}$, соответствующего испытательному элементу входных данных $\mathbf{a}^*$. Первым этапом при вычислении $\mathbf{b}^*$ является построение совместной плотности распределения всего из обучающего выходного набора $\bar{\mathbf{b}}$ и испытательного элемента выходных данных гауссовского процесса, $f(\mathbf{a}^*)$, которые обусловлены в отношении обучающего входного набора $\bar{\mathbf{a}}$ и испытательного элемента входных данных $\mathbf{a}^*$. Указанная совместная плотность распределения имеет следующий вид:

$$p(\bar{\mathbf{b}}|f(\mathbf{a}^*))|\bar{\mathbf{a}}, \mathbf{a}^* = \mathcal{N}\left(0, \begin{bmatrix} \mathbf{k}_\gamma(\bar{\mathbf{a}}, \bar{\mathbf{a}}) & \mathbf{k}_\theta(\bar{\mathbf{a}}, \mathbf{a}^*) \\ \mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}) & k_\theta(\mathbf{a}^*, \mathbf{a}^*) \end{bmatrix}\right), \text{ Уравнение (11)}$$

где $\mathbf{k}_\gamma \equiv \mathbf{k}_\theta(\bar{\mathbf{a}}, \bar{\mathbf{a}}) + \sigma^2 \mathbf{I}_{J \times J}$. Для данного уравнения (11) в байесовской инфраструктуре элемент выходных данных гауссовского процесса, $f(\mathbf{a}^*)$, вычисляется путем построения распределения по всем выходным данным гауссовского процесса. Иначе говоря, отыскивается апостериорное распределение для элемента выходных данных гауссовского процесса, $f(\mathbf{a}^*)$. Разумеется, апостериорное распределение по $f(\mathbf{a}^*)$ требует включения только тех функций, которые согласуются с обучающим набором D . При вероятностных установках апостериорное распределение по $f(\mathbf{a}^*)$ можно вычислить путем согласования совместного распределения в уравнении (11) в отношении обучающего набора D , что дает

$$p(f(\mathbf{a}^*)|D, \mathbf{a}^*) = \mathcal{N}(\mu_\theta^*, k_\theta^*), \text{ Уравнение (12)}$$

где $p(f(\mathbf{a}^*)|D, \mathbf{a}^*)$ - апостериорное распределение для элемента выходных данных гауссовского процесса, и $\mu_\theta^* \equiv \mathbb{E}[f(\mathbf{a}^*)|D, \mathbf{a}^*]$ имеет вид

$$\mu_\theta^* = \mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}})[\mathbf{k}_\gamma(\bar{\mathbf{a}}, \bar{\mathbf{a}})]^{-1}\bar{\mathbf{b}}, \text{ Уравнение (13)}$$

и $k_\theta^* \equiv \mathbb{V}[f(\mathbf{a}^*)|D, \mathbf{a}^*]$ имеет вид

$$k_\theta^* = k_\theta(\mathbf{a}^*, \mathbf{a}^*) - \mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}})[\mathbf{k}_\gamma(\bar{\mathbf{a}}, \bar{\mathbf{a}})]^{-1}\mathbf{k}_\theta(\bar{\mathbf{a}}, \mathbf{a}^*). \text{ Уравнение (14)}$$

Для данного уравнения (12) предсказывающее апостериорное распределение для элемента выходных данных $\mathbf{b}^*$ можно вычислить следующим образом:

$$p(\mathbf{b}^*|D, \mathbf{a}^*) = \mathcal{N}(\mu_\theta^*, k_\theta^* + \sigma^2), \text{ Уравнение (15)}$$

где μ_θ^* и k_θ^* даны, соответственно, в уравнениях (13) и (14). Для одного испытательного элемента входных данных $\mathbf{a}^* \in \mathbb{R}^{n_a}$ предсказание на основе гауссовского процесса в уравнении (15) дает распределение выходных данных, имеющих ненулевую вероятность реализации. В приложениях в реальном времени, таких как управление и отслеживание, с большей вероятностью представляет интерес точечная оценка, а не все распределение. Точечную оценку можно вычислить с использованием подхода на основе теории принятия решений. Можно показать, что для гауссовского апостериорного распределения в уравнении (15) средняя функция минимизирует и ожидаемую безусловную функцию, и квадратичную функцию риска, при этом $\hat{\mathbf{b}} = \mu_\theta^*$ представляет собой наиболее вероятный элемент выходных данных для элемента входных данных $\mathbf{a}^*$. Кроме того, для выбора $\hat{\mathbf{b}} = \mu_\theta^*$ в качестве предсказания интервал с приблизительно 95% правдоподобием имеет вид:

$$\mathbf{b}^L = (\mu_\theta^* - 2(\sqrt{k_\theta^* + \sigma^2})) \leq \mathbf{b} \leq \mu_\theta^* + 2(\sqrt{k_\theta^* + \sigma^2}) = \mathbf{b}^U. \text{ Уравнение (16)}$$

Интервал в уравнении (16) можно использовать для количественной оценки качества предсказаний на основе гауссовского процесса и/или выполнения управления для предсказания по модели на основе гауссовского процесса или других робастных стратегий отслеживания.

Обращаясь теперь к выбору релевантных выборок (здесь - наборов данных наблюдений) в ответ на запрос, задача заключается в выборе для данной точки запроса, $\mathbf{a}^* \in \mathbb{R}^{n_a}$, и центральной базы данных/библиотеки $\mathcal{L} \equiv \{\mathbf{b}_i, \mathbf{a}_i\}_{i=1}^{L_t}$, содержащей $L \in \mathbb{N}$ пары элементов входных данных и выходных данных (наборы данных наблюдений), локального обучающего набора $\mathcal{D} \equiv \{\mathbf{b}_j, \mathbf{a}_j\}_{j=1}^D$ в момент времени, $t \in \mathbb{N}$ содержащий выборки $D \in \mathbb{N}$, где $D \ll L$. Предполагается, что $\mathcal{L}$ является динамической и может содержать разные элементы в ходе производственного цикла. Существует множество способов построения $\mathcal{D}$ исходя из $\mathcal{L}$. В целях данного анализа $\mathcal{D}$ выбирают на основе евклидова расстояния между спектрами (например, векторами рамановского сканирования) в наборе $\mathcal{L}$. И хотя критерии подобия на евклидовой основе в инфраструктуре JTL, как сообщалось, в некоторых ситуациях являются субоптимальными, они могут представлять собой предпочтительный выбор при использовании модели на основе гауссовского процесса. Причиной этого является то, что модель на основе гауссовского процесса сама основана на евклидовом расстоянии. Гауссовское ядро присваивает более высокий коэффициент корреляции только в том случае, если входные данные в наборе $\{\mathbf{a}_i, \mathbf{a}_j\}$ "близки" друг к другу. Поэтому путем создания локального обучающего набора $\mathcal{D}$, в котором все входные данные являются "близкими" к точке запроса, можно обеспечить улавливание локальной моделью на основе гауссовского максимальной "корреляции" для предсказания элемента выходных данных в точке запроса.

Примерный алгоритм, формально описывающий способ создания локального обучающего набора $\mathcal{D}$ исходя из $\mathcal{L}$, обучения модели на основе гауссовского процесса с использованием этого обучающего набора и выполнения предсказания с использованием обученной модели, представлен ниже в алгоритме 1

1. **Input:** Library $\mathcal{L} = \{(\mathbf{a}_i, b_i)\}_{i=1}^L$, query point $\mathbf{a}^*$
 2. **Output:** Prediction $\hat{b}$ and uncertainty (b^L, b^U)
 3. **for** $t = 1$ to T **do**
 4. Set $I \leftarrow \text{sample_index}(\mathcal{L})$ and $\mathcal{D} \leftarrow \{\emptyset\}$
 5. **for** $d = 1$ to D **do**
 6. $k_* \in \arg \max_{i \in I} \exp(-\|\mathbf{a}_i - \mathbf{a}^*\|)$
 7. $\mathcal{D} \leftarrow \mathcal{D} \cup \{\mathbf{a}_{k_*}, b_{k_*}\}$
 8. $I \leftarrow I \setminus \{i\}$
 9. **end for**
 10. Обучить модель на основе гауссовского процесса по уравнению (1) с использованием $\mathcal{D}$ и оценки γ^*
 11. Вычислить $\hat{b}_i$ (b^L, b^U) с использованием уравнений (13) и (16)
 12. **end for**
- Алгоритм 1**

Обратимся теперь к фиг. 4, на которой показан примерный поток 250 данных, который может иметь место при анализе биофармацевтического процесса с использованием методики JITL, описанной в данном документе. Поток 250 данных может иметь место, например, в системе 100, представленной на фиг. 1, или системе 150, представленной на фиг. 2. В потоке 250 данных спектральные данные 252 предоставляются спектрометром/зондом. Например, спектральные данные 252 могут содержать вектор рамановского сканирования, сгенерированный рамановским анализатором 106, или вектор сканирования NIR, и т.д. Точка 254 запроса генерируется (например, блоком 140 запроса) на основе спектральных данных 252 и используется, например, для запроса глобального набора 256 данных, который может содержать все наборы данных наблюдений в базе 136 данных наблюдений. На основе запроса в глобальном наборе 256 данных идентифицируется локальный набор 258 данных. Локальный набор 258 данных может быть выбран, например, на основе критериев релевантности (например, евклидова расстояния), как описано выше.

Локальный набор 258 данных затем используется в качестве обучающих данных (например, генератором 142 локальной модели) для калибровки локальной модели 260 (например, локальной модели 132). Локальная модель 132 затем используется (например, блоком 144 предсказания) для предсказания элемента 262 выходных данных (аналитического измерения), такого как концентрация компонентов питательной среды, состояние питательной среды (например, глюкоза, лактат, глутамат, глутамин, аммиак, аминокислоты, Na^+ , K^+ , а также другие питательные вещества или метаболиты, pH, pCO_2 , pO_2 , температура, осмоляльность и т.д.), плотность жизнеспособных клеток, титр, критические показатели качества, состояние клеток и т.д., а также возможно вывода границ правдоподобия или другого подходящего индикатора достоверности.

Несмотря на то, что локальная модель на основе JITL (например, такая как алгоритм 1 и поток 250 данных) предоставляет робастную нелинейную инфраструктуру моделирования, такой подход не имеет внутреннего механизма для адаптации к зависящим от времени изменениям в процессе. Для устранения этого недостатка в некоторых вариантах осуществления может использоваться стратегия "адаптивного" JITL (A-JITL). Как отмечено выше, новые выборки могут включаться в $\mathcal{L}$ по мере того, как эти выборки становятся доступными. В таких вариантах осуществления (т.е., где $\mathcal{L}$ является динамической) $\mathcal{L}$ можно обозначить как $\mathcal{L}_t$. В одном таком варианте осуществления реализован способ скользящего временного окна, в котором вновь полученная выборка добавляется в $\mathcal{L}_t$, а самая старая выборка удаляется из $\mathcal{L}_t$. Исключение самой старой выборки может являться преимущественным, поскольку в адаптивных стратегиях сохранение размера $\mathcal{L}_t$ может являться критичным для обеспечения вычислительной разрешимости инфраструктуры JITL в целом. Однако одной основной проблемой, связанной с данным подходом, является то, что простое исключение старых выборок может приводить к потере информации, так как старые выборки могут содержать релевантную информацию.

Во избежание такой потери информации в одном варианте осуществления новые выборки добавляются в $\mathcal{L}_t$ без удаления каких-либо старых/существующих выборок. Таким образом, центральная база данных $\mathcal{L}_t$ расширяется на возрастающее количество выборок по мере того, как становятся доступными новые аналитические измерения. В применении процессов культивирования клеток расширение базы данных может не приводить к каким-либо значительным вычислительным проблемам вследствие того, что такие процессы обычно эксплуатируются как периодические процессы с временем цикла от двух до трех недель. Это естественным образом ограничивает количество новых выборок, которые необходимо

включать в $\mathcal{L}_t$. Кроме того, в ходе процесса культивирования клеток обычно выборке подвергается лишь ограниченное количество аналитических измерений (в отличие, например, от химической промышленности, в которой аналитические измерения подвергаются выборке часто). Поэтому обычно будет иметь место лишь умеренное увеличение размера базы данных $\mathcal{L}_t$ без какой-либо значительной нагрузки на вычислительную устойчивость инфраструктуры JITL в целом.

Несмотря на то, что включение новых выборок в $\mathcal{L}_t$ важно для непрерывной адаптации алгоритма 1 (см. выше), успех данного подхода полагается на выбор указанных новых выборок в локальной базе данных $\mathcal{D}$ для калибровки локальной модели. Алгоритм 1, в котором выборки для $\mathcal{D}$ из $\mathcal{L}$ выбираются на основе евклидова расстояния (например, строка 6 алгоритма 1), может называться подходом "релевантности в пространстве", так как он отдает предпочтение только тем выборкам, которые являются релевантными (близкими) в пространстве. Если новые выборки не являются близкими к запрашиваемой выборке, что, вероятно, имеет место в случае, когда происходит резкое изменение заданного значения (или другое резкое изменение условий процесса), алгоритм 1 может быть неспособен включить эти выборки в $\mathcal{D}$. С другой стороны, рекурсивные методы (например, регуляризованный метод частных наименьших квадратов (RLPS), рекурсивный метод наименьших квадратов (RLS) и рекурсивный метод N-ходовых частных наименьших квадратов (RNPLS)) являются "релевантными по времени", так как они отдают предпочтение новейшим измерениям независимо от релевантности в пространстве. Обновление локальной модели с использованием новейших выборок может обеспечить возможность успешной адаптации рекурсивных методов к текущим условиям процесса.

Один такой вариант осуществления, называемый в данном документе "адаптивным" JITL (A-JITL), отдает предпочтение выборкам, которые являются релевантными как по времени, так и в пространстве. Если $\mathcal{G}^- = \{\{\mathbf{a}_i^-, \mathbf{b}_i^-\}\}_{i=1}^L$ представляет набор из L исторических измерений, доступных перед началом текущего эксперимента (т.е. эксперимента/процесса, в котором имеет место запрос $\mathbf{a}^*$), и если $\mathcal{G}^+ = \{\{\mathbf{a}_j^+, \mathbf{b}_j^+\}\}_{j=1}^n$ представляет собой набор из n измерений, доступных из текущего эксперимента, выборки могут быть распределены следующим образом:

$$\mathcal{L}_t = \mathcal{G}^- \cup \mathcal{G}^+ \setminus \{\{\mathbf{a}_j^+, \mathbf{b}_j^+\}\}_{j=n-k+1}^n, \text{ Уравнение (17a)}$$

$$\mathcal{K} = \{\{\mathbf{a}_j^+, \mathbf{b}_j^+\}\}_{j=n-k+1}^n, \text{ Уравнение (17b)}$$

где $\mathcal{L}_t$ представляет центральную базу данных, и $\mathcal{K}$ представляет набор последних (самых недавних) k измерений. В некоторых вариантах осуществления $\mathcal{K}$ содержит последние k выборки из текущего эксперимента/процесса, и $\mathcal{L}_t$ содержит выборки из предыдущих экспериментов/процессов, а также (потенциально) выборки из текущего эксперимента/процесса, старше последних k выборок. Приведенные выше уравнения (17a) и (17b) заданы для данного запроса $\mathbf{a}^*$. Для запроса, поступающего в другой момент времени, наборы данных $\mathcal{L}_t$ и $\mathcal{K}$ могут содержать разные выборки в зависимости от количества измерений, доступных в этот момент времени. Например, как только становится доступной выборка $(\mathbf{a}_{n+1}^+, \mathbf{b}_{n+1}^+)$, $(\mathbf{a}_{n-k+1}^+, \mathbf{b}_{n-k+1}^+)$ удаляется из $\mathcal{K}$, а $(\mathbf{a}_{n+1}^+, \mathbf{b}_{n+1}^+)$ включается в $\mathcal{K}$. Исключенная выборка $(\mathbf{a}_{n-k+1}^+, \mathbf{b}_{n-k+1}^+)$ затем включается в $\mathcal{L}_t$ для предотвращения какой-либо потери информации. Обновление $\mathcal{K}$ новейшими измерениями обеспечивает то, что $\mathcal{K}$ отражает по меньшей мере некоторые текущие условия.

Для данных $\mathcal{L}_t$ и $\mathcal{K}$ целью является выбор $\mathcal{D}$. Как указано выше, для A-JITL в $\mathcal{D}$ включаются выборки, релевантные как по времени, так и в пространстве. Если предположить, что $\mathcal{D}$ можно разложить как

$$\mathcal{D} \equiv \mathcal{D}_S \cup \mathcal{D}_T, \text{ Уравнение (18)}$$

где $\mathcal{D}_S$ и $\mathcal{D}_T$ представляют собой наборы, релевантные в пространстве и по времени, соответственно, то целью является выбор $\mathcal{D}_S$ и $\mathcal{D}_T$. В первую очередь, предполагается, что $\mathcal{D}_S \cap \mathcal{D}_T = \emptyset$, так что $\mathcal{D}$ содержит только уникальные выборки. Для конструирования $\mathcal{D}_S$ D-k выборок выбирают из $\mathcal{L}_t$ на основе основанной на расстоянии (пространственной) метрики, такой как "индекс подобия" или "s-значение":

$$s_i = \text{sim}(\mathbf{a}_i, \mathbf{a}^*) = \exp(-\|\mathbf{a}_i - \mathbf{a}^*\|). \text{ Уравнение (19)}$$

Уравнение (19) можно использовать, например, в качестве метрики подобия в вышеописанной методике (неадаптивного) JITL. Так, например, из $\mathcal{L}_t$ для включения в $\mathcal{D}_S$ можно выбрать D-k выборок с наибольшими s-значениями. Для конструирования $\mathcal{D}_T$, если предполагается, что последние k выборок из текущего эксперимента/процесса релевантны по времени, $\mathcal{D}_T$ в некоторых вариантах осуществления может быть задан как равный $\mathcal{K}$. Следует отметить, что, в отличие от s-значений, которые определяют членство выборок в $\mathcal{D}_S$, решение о членстве в $\mathcal{D}_T$ принимается на основе времен выборок.

Разумеется, в зависимости от сценария, выборки в $\mathcal{D}_T$ могут характеризоваться большими s-значениями. Независимо от s-значения, предполагается, что $\mathcal{D}_T$ является релевантным только по времени. Аналогично, $\mathcal{D}_S$ является релевантным только в пространстве, поскольку вследствие построения $\mathcal{L}_t$ не обладает релевантностью по времени. Следует отметить, что $\mathcal{D}_S$ и $\mathcal{D}_T$ задают для данного запроса $\mathbf{a}^*$, выборки в $\mathcal{D}_S$ выбирают на основе их s-значений, вычисленных относительно $\mathbf{a}^*$, и выборки в $\mathcal{D}_T$ выбирают на основе их времен выборок, вычисленных относительно времени выборки $\mathbf{a}^*$. Для удобства $\mathcal{D}_S$ и $\mathcal{D}_T$ обобщенно задают следующим образом:

$$\mathcal{D}_S \equiv \{\bar{\mathbf{a}}_S, \bar{\mathbf{b}}_S\}, \text{ Уравнение (20a)}$$

$$\mathcal{D}_T \equiv \{\bar{\mathbf{a}}_T, \bar{\mathbf{b}}_T\}, \text{ Уравнение (20b)}$$

где $\bar{\mathbf{a}}_S$ и $\bar{\mathbf{a}}_T$ - соответственно релевантные в пространстве и по времени выборки из рамановского спектрометра, и $\bar{\mathbf{b}}_S$ и $\bar{\mathbf{b}}_T$ - соответственно релевантные в пространстве и по времени выборки из аналитического прибора, так что

$$\bar{\mathbf{a}}_S \equiv [\mathbf{a}_1, \dots, \mathbf{a}_{D-k}]^T; \bar{\mathbf{a}}_T \equiv [\mathbf{a}_{D-k+1}, \dots, \mathbf{a}_D]^T, \text{ Уравнение (21a)}$$

$$\bar{\mathbf{b}}_S \equiv [b_1, \dots, b_{D-k}]^T; \bar{\mathbf{b}}_T \equiv [b_{D-k+1}, \dots, b_D]^T. \text{ Уравнение (21b)}$$

Подстановка уравнений (20a) и (20b) в уравнение (18) дает набор $\mathcal{D}$, обобщенно обозначенный как $\mathcal{D} \equiv \{\bar{\mathbf{a}}, \bar{\mathbf{b}}\}$, где $\bar{\mathbf{a}} \equiv [\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T]^T$, и $\bar{\mathbf{b}} \equiv [\bar{\mathbf{b}}_S, \bar{\mathbf{b}}_T]^T$. В отличие от обсужденной выше методики (неадаптивного) JTL, в локальной библиотеке/наборе данных $\mathcal{D}$ отдается предпочтение выборкам, релевантным в пространстве и по времени.

Для данных $\mathcal{D}_T$ и запроса $\mathbf{a}^*$ можно откалибровать модель на основе гауссовского процесса по уравнению (1) (например, локальную модель 132). Точечную оценку и интервал правдоподобия для $\mathbf{a}^*$ можно вычислить с использованием, соответственно, уравнении (13) и (16), где $\mathbf{k}_\gamma(\bar{\mathbf{a}}, \bar{\mathbf{a}})$ и $\mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}})$ имеют вид

$$\mathbf{k}_\gamma(\bar{\mathbf{a}}, \bar{\mathbf{a}}) \equiv \begin{bmatrix} \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) & \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) \\ \mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_S) & \mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_T) \end{bmatrix} + \sigma^2 \mathbf{I}_{D \times D}, \text{ Уравнение (22a)}$$

$$\mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}) \equiv [\mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_S) \quad \mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_T)], \text{ Уравнение (22b)}$$

где $\mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) \in S_+^{(D-k)}$ и $\mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_T) \in S_+^k$ - ковариационные функции, связанные с $\mathcal{D}_S$ и $\mathcal{D}_T$ соответственно, и где $\mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) \in \mathbb{R}^{(D-k)k}$ - ковариация между $\mathcal{D}_S$ и $\mathcal{D}_T$.

Примерный алгоритм, формально описывающий методику А-JTL, представлен ниже в алгоритме 2

```

1. Input: Library  $\mathcal{L}_t = \{(\mathbf{a}_i, \mathbf{b}_i)\}_{i=1}^L$ , query point  $\mathbf{a}^*$ 
2. Output: Prediction  $\hat{\mathbf{b}}$  and uncertainty  $(b^L, b^U)$ 
3. Set  $\mathcal{K} \leftarrow \{\emptyset\}$ 
4. for  $t = 1$  to  $T$  do
5. Set  $I \leftarrow \text{sample\_index}(\mathcal{L}_t)$ ,  $\mathcal{D}_S \leftarrow \{\emptyset\}$ ,  $\mathcal{D}_T \leftarrow \{\emptyset\}$ 
6. for  $d = 1$  to  $D - \text{set\_cardinality}(\mathcal{K})$  do
7.  $i_* \in \arg \max_{i \in I} \text{sim}(\mathbf{a}_i, \mathbf{a}^*)$ 
8.  $\mathcal{D}_S \leftarrow \mathcal{D}_S \cup \{\mathbf{a}_{i_*}, \mathbf{b}_{i_*}\}$ 
9.  $I \leftarrow I \setminus \{i_*\}$ 
10. end for
11. if  $\text{set\_cardinality}(\mathcal{K}) \geq 1$  then
12.  $\mathcal{D}_T \leftarrow \mathcal{K}$ 
13. end if
14.  $\mathcal{D} \leftarrow \mathcal{D}_S \cup \mathcal{D}_T$ 
15. Обучить модель на основе гауссовского процесса по уравнению (1) с использованием  $\mathcal{D}$  и оценки  $\gamma^*$ 
16. Вычислить  $\hat{\mathbf{b}}$  и  $(b^L, b^U)$  с использованием уравнений (13) и (16)
17. if  $b^*$  is available then
18. if  $\text{size}(\mathcal{K}) = k$  then
19.  $\mathcal{L}_t \leftarrow \mathcal{L}_t \cup \text{select\_oldest}(\mathcal{K})$ 
20.  $\mathcal{K} \leftarrow \text{delete\_oldest}(\mathcal{K})$ 
21.  $\mathcal{K} \leftarrow \mathcal{K} \cup \{\mathbf{a}^*, b^*\}$ 
22. end if
23.  $\mathcal{K} \leftarrow \mathcal{K} \cup \{\mathbf{a}^*, b^*\}$ 
24. end if
25. end for

```

Алгоритм 2

Таким образом, в алгоритме 2 (релевантное в пространстве) JTL скомбинировано с рекурсивным обучением (релевантным по времени). Например, для $|\mathcal{D}_T| = 0$ калибровка локальной модели 132 с использованием алгоритма 2 подобна релевантному в пространстве JTL, тогда как для $|\mathcal{D}_S| = 0$ калибровка локальной модели 132 с использованием алгоритма 2 подобна рекурсивному обучению. Таким образом, путем подгонки $|\mathcal{D}_S|$ и $|\mathcal{D}_T|$ можно надлежащим образом сбалансировать (не являющееся рекурсивным) JTL и рекурсивное обучение.

Обратимся теперь к фиг. 5, на которой показан примерный поток 300 данных, который может иметь

место при анализе биофармацевтического процесса с использованием методики А-ИТЛ, описанной в данном документе. Поток 300 данных может иметь место, например, в системе 100, представленной на фиг. 1, или системе 150, представленной на фиг. 2. В потоке 300 данных спектральные данные 302 предоставляются спектрометром/зондом. Например, спектральные данные 302 могут содержать вектор рамановского сканирования, сгенерированный рамановским анализатором 106, или вектор сканирования NIR и т.д. Точка 304 запроса генерируется (например, блоком 140 запроса) на основе спектральных данных 302 и используется, например, для запроса глобального набора 306 данных, который может содержать все наборы данных наблюдений в базе 136 данных наблюдений. Глобальный набор 306 данных логически разделен на последние k элементов 307А (например, все из которых получены из текущего эксперимента/процесса) и на все элементы 307В перед последними k элементами 307А (например, из предыдущих экспериментов/процессов, а также, возможно, текущего эксперимента/процесса). Значение k можно определить на основе номера выборки точки 304 запроса. В рамках данного документа термин "номер выборки" может в широком смысле относиться к любому индикатору времени или относительно времени, связанного с данной выборкой/наблюдением. Некоторые элементы из элементов 307В добавлены в локальный набор 308 данных на основе пространственного подобия (например, евклидова расстояния) точке 304 запроса, тогда как все элементы 307А могут быть добавлены в локальный набор 308 данных независимо от пространственного подобия. Локальный набор 308 данных можно сгенерировать из элементов 307А и элементов 307В, например, по алгоритму 2.

Локальный набор 308 данных затем используется в качестве обучающих данных (например, генератором 142 локальной модели) для калибровки локальной модели 310 (например, локальной модели 132). Локальная модель 310 затем используется (например, блоком 144 предсказания) для предсказания элемента 312 выходных данных (аналитического измерения), такого как концентрация компонентов питательной среды, состояние питательной среды (например, глюкоза, лактат, глутамат, глутамин, аммиак, аминокислоты, Na^+ , K^+ , а также другие питательные вещества или метаболиты, pH, pCO_2 , pO_2 , температура, осмоляльность и т.д.), плотность жизнеспособных клеток, титр, критические показатели качества, состояние клеток и т.д., а также, возможно, вывода границ правдоподобия или другого подходящего индикатора достоверности.

Если фактическое аналитическое измерение (например, измерение, выполненное таким аналитическим прибором, как один из аналитических приборов 104) является доступным, создается новый элемент 314, который добавляется в глобальный набор 306 данных. Такие измерения могут быть доступны, например, на основе периодической выборки (например, один или два раза в сутки) и/или их можно сделать доступными в ответ на инициирующий фактор с переменным расчетным временем (например, если некоторое количество предсказаний в строке имеет недопустимо широкие границы правдоподобия, и т.д.), как более подробно обсуждено ниже.

И хотя включение в $\mathcal{D}$ выборок, релевантных в пространстве и по времени, необходимо для непрерывной адаптации обсужденного выше подхода А-ИТЛ, общая степень адаптации, достигаемая А-ИТЛ, зависит от того, насколько эффективно $\mathcal{D}$ используется для калибровки локальной модели. Для выборки/точки запроса, $\mathbf{a}^*$, релевантная по времени выборка $(\mathbf{a}_i, b_i) \in \mathcal{D}_S$ обеспечивает удовлетворительную корреляцию между функциями $(f(\mathbf{a}^*), f(\mathbf{a}_i))$. Причиной этого является то, что для запроса $\mathbf{a}^*$ как пространственная релевантность $(\mathbf{a}_i, b_i)$, так и коэффициент корреляции между $(f(\mathbf{a}^*), f(\mathbf{a}_i))$ вычисляются на основе евклидова расстояния между $(\mathbf{a}_i, \mathbf{a}^*)$.

Таким образом, для выбора критерия подобия на евклидовой основе в уравнении (19) и ядра на евклидовой основе в уравнении (4) ожидается, что выборки в $\mathcal{D}_S$ обеспечивают высокие функциональные корреляции. И наоборот, релевантная по времени выборка $(\mathbf{a}_j, b_j) \in \mathcal{D}_T$ может не обеспечивать сильную корреляцию между функциями $(f(\mathbf{a}^*), f(\mathbf{a}_j))$. Причиной этого является то, что, как отмечено выше, выборки в $\mathcal{D}_T$ необязательно релевантны в пространстве. Как следствие, коэффициент корреляции между $(f(\mathbf{a}^*), f(\mathbf{a}_j))$, приписываемый гауссовским ядром в уравнении (4), будет небольшим, если небольшой является пространственная релевантность $(\mathbf{a}_j, b_j)$. С точки зрения моделирования, обучение модели на основе гауссовского процесса в уравнении (1) выборками, несущими небольшие коэффициенты корреляции, является нежелательным, так как это ведет к неудовлетворительному качеству модели. Математически это можно продемонстрировать следующим образом.

Для запроса $\mathbf{a}^*$ и откалиброванной модели на основе гауссовского процесса по алгоритму 2 предсказание на основе модели, $\hat{b}$, можно вычислить с использованием уравнения (13). Без потери общности, если $\sigma^2 = 0$ (случай отсутствия шума), уравнение (13) можно записать следующим образом:

$$\hat{b} = \begin{bmatrix} \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \mathbf{a}^*) \\ \mathbf{k}_\theta(\bar{\mathbf{a}}_T, \mathbf{a}^*) \end{bmatrix}^T \begin{bmatrix} \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) & \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) \\ \mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_S) & \mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_T) \end{bmatrix}^{-1} \begin{bmatrix} \bar{\mathbf{b}}_S \\ \bar{\mathbf{b}}_T \end{bmatrix}. \text{ Уравнение (23)}$$

Если $(\bar{\mathbf{a}}_T, \bar{\mathbf{b}}_T)$ имеет пренебрежимо несущественную пространственную релевантность (т.е. s -значение между $\bar{\mathbf{a}}_T$ и $\mathbf{a}^*$ является бесконечно большим), то уравнение (4) приводит к $\mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_T) \approx 0_{1 \times k}$. Кроме того, по построению, так как $\bar{\mathbf{a}}_S$ ближе к $\mathbf{a}^*$, чем к $\bar{\mathbf{a}}_T$, результатом являются $\mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) \approx 0_{(D-k) \times k}$ и $\mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_S) \approx 0_{k \times (D-k)}$. Их подстановка в уравнение (23) дает

$$\hat{b} \approx \begin{bmatrix} \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \mathbf{a}^*)^T \\ 0_{k \times 1} \end{bmatrix}^T \begin{bmatrix} \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) & 0_{(D-k) \times k} \\ 0_{k \times (D-k)} & \mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_T) \end{bmatrix}^{-1} \begin{bmatrix} \bar{\mathbf{b}}_S \\ \bar{\mathbf{b}}_T \end{bmatrix}, \text{ Уравнение (24a)}$$

$$= \begin{bmatrix} \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \mathbf{a}^*)^T \\ 0_{k \times 1} \end{bmatrix}^T \begin{bmatrix} \mathbf{k}_\theta^{-1}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) & 0_{(D-k) \times k} \\ 0_{k \times (D-k)} & \mathbf{k}_\theta^{-1}(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_T) \end{bmatrix} \begin{bmatrix} \bar{\mathbf{b}}_S \\ \bar{\mathbf{b}}_T \end{bmatrix} \text{ уравнение (24b)}$$

$$= \mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_S) \mathbf{k}_\theta^{-1}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) \bar{\mathbf{b}}_S. \text{ Уравнение (24c)}$$

Из уравнения (24c) ясно, что точечная оценка не зависит от $\mathcal{D}_T$. Аналогично, можно показать, что уравнение (16) также не зависит от $\mathcal{D}_T$. Например, k_θ^* в уравнении (16) можно вычислить следующим образом:

$$-k_\theta^* + k_\theta(\mathbf{a}^*, \mathbf{a}^*) = \mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}) [\mathbf{k}_\theta(\bar{\mathbf{a}}, \bar{\mathbf{a}})]^{-1} \mathbf{k}_\theta(\bar{\mathbf{a}}, \mathbf{a}^*), \text{ Уравнение (25a)}$$

$$\approx \begin{bmatrix} \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \mathbf{a}^*)^T \\ \mathbf{k}_\theta(\bar{\mathbf{a}}_T, \mathbf{a}^*)^T \end{bmatrix}^T \begin{bmatrix} \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) & \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) \\ \mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_S) & \mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_T) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \mathbf{a}^*) \\ \mathbf{k}_\theta(\bar{\mathbf{a}}_T, \mathbf{a}^*) \end{bmatrix}, \text{ уравнение (25b)}$$

$$= \begin{bmatrix} \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \mathbf{a}^*)^T \\ 0_{k \times 1} \end{bmatrix}^T \begin{bmatrix} \mathbf{k}_\theta^{-1}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) & 0_{(D-k) \times k} \\ 0_{k \times (D-k)} & \mathbf{k}_\theta^{-1}(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_T) \end{bmatrix} \begin{bmatrix} \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \mathbf{a}^*) \\ 0_{k \times 1} \end{bmatrix}, \text{ Уравнение (25c)}$$

$$k_\theta^* \approx k_\theta(\mathbf{a}^*, \mathbf{a}^*) - \mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_S) \mathbf{k}_\theta^{-1}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) \mathbf{k}_\theta(\bar{\mathbf{a}}_S, \mathbf{a}^*). \text{ Уравнение (25d)}$$

Из уравнений (25b) и (25c) можно видеть, что используется несколько приближений, в том числе $\mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_T) \approx 0_{k \times 1}$, $\mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) \approx 0_{(D-k) \times k}$ и $\mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_S) \approx 0_{k \times (D-k)}$. Тогда из уравнений (20a) и (20b) очевидно, что алгоритм 2 не способен удовлетворительно использовать $\mathcal{D}_T$, если набор имеет ограниченную пространственную релевантность.

В некоторых вариантах осуществления для обеспечения возможности внесения вклада в $\mathcal{D}$ выборками, релевантными и в пространстве, и по времени, используется подход "пространственно-временного" JTL (ST-JTL) со следующей пространственно-временной рамановской моделью (например, в качестве локальной модели 132):

$$b_i = g(\mathbf{a}_i, t_i) + \epsilon_i, \text{ Уравнение (26)}$$

где $g: \mathbb{R}^{n_a \times n_t} \rightarrow \mathbb{R}$ - пространственно-временная рамановская модель, и t_i - номер выборки в $\mathbf{a}_i$, и $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ - последовательность независимых гауссовских случайных переменных с нулевым средним и неизвестной дисперсией $\sigma^2 \in \mathbb{R}_+$. В отличие от уравнения (1), пространственно-временная модель в уравнении (26) зависит как от спектрального сигнала, так и от времени его выборки. Как и выше, предполагается, что g - скрытая функция, моделируемая как гауссовский процесс так, что для любого элемента входных данных $(\mathbf{a}, t)$,

$$g(\mathbf{a}, t) \sim \text{GP}(0, r_\theta(\mathbf{a}, \mathbf{a}, t, t)), \text{ Уравнение (27)}$$

является случайной функцией. Для удобства средняя функция в уравнении (27) предполагается равной нулю, но это не обязательно имеет место в общем случае. Кроме того, для любых произвольных входных данных $(\mathbf{a}_i, t_i)$ и $(\mathbf{a}_j, t_j)$, ковариационную функцию $r_\theta(\mathbf{a}_i, \mathbf{a}_j, t_i, t_j)$ можно задать следующим образом:

$$r_\theta(\mathbf{a}_i, \mathbf{a}_j, t_i, t_j) = k_{\text{space}}(\mathbf{a}_i, \mathbf{a}_j) + k_{\text{time}}(t_i, t_j), \text{ Уравнение (28)}$$

где $k_{\text{space}}(\mathbf{a}_i, \mathbf{a}_j) \in \mathbb{R}_+$ и $k_{\text{time}}(t_i, t_j) \in \mathbb{R}_+$ - соответственно пространственная ковариация и временная ковариация между $(g(\mathbf{a}_i, t_i), g(\mathbf{a}_j, t_j))$. Следует отметить, что для Запроса $(\mathbf{a}^*, t^*)$, если выборка $(\mathbf{a}_j, t_j) \in \mathcal{D}_T$ имеет пренебрежимо малую пространственную релевантность, то $k_{\text{space}}(\mathbf{a}_j, \mathbf{a}^*) \approx 0$, но $k_{\text{time}}(t_j, t^*) > 0$, так что уравнение (28) задает ненулевой коэффициент корреляции между $(g(\mathbf{a}^*, t^*), g(\mathbf{a}_j, t_j))$. Наконец, следует отметить, что уравнение (28) является истинной ковариационной функцией, поскольку сумма двух независимых ядер также является ядром. Предположим, что k_{space} и k_{time} являются гауссовскими ядрами, таким образом предназначены для любой пары элементов входных данных $(\mathbf{a}_i, t_i)$ и $(\mathbf{a}_j, t_j)$,

$$k_{\text{space}}(\mathbf{a}_i, \mathbf{a}_j) = \alpha_1 \exp\left(-\frac{\|\mathbf{a}_i - \mathbf{a}_j\|^2}{2\alpha_2}\right), \text{ Уравнение (29a)}$$

$$k_{\text{time}}(a_i, t_j) = \beta_1 \exp\left(-\frac{\|t_i - t_j\|^2}{2\beta_2}\right), \text{ Уравнение (29b)}$$

где $\theta \equiv [\alpha_1, \alpha_2, \beta_1, \beta_2]^T \in \Theta \in \mathbb{R}^4$ - параметр ядра. Для данных уравнений (29a) и (29b) уравнение (28) приписывает высокий коэффициент корреляции между $(g(\mathbf{a}_i, t_i), g(\mathbf{a}_j, t_j))$, если $(\mathbf{a}_i, t_i)$, $(\mathbf{a}_j, t_j)$ близки друг к другу. Если $\bar{\mathbf{t}}_S = [t_1, \dots, t_{D-k}]^T$ и $\bar{\mathbf{t}}_T = [t_{D-k+1}, \dots, t_D]^T$ обозначают номера выборок для релевантных по состоянию и времени выборок в $\mathcal{D}$ соответственно так, что $\bar{\mathbf{t}} = [\bar{\mathbf{t}}_S; \bar{\mathbf{t}}_T]$, то для запроса $(\mathbf{a}^*, t^*)$ ковариационную функцию r_θ уравнении (28) можно записать как

$$\mathbf{r}_\theta(\bar{\mathbf{a}}, \bar{\mathbf{a}}, \bar{\mathbf{t}}, \bar{\mathbf{t}}) \equiv \begin{bmatrix} r_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S, \bar{\mathbf{t}}_S, \bar{\mathbf{t}}_S) & r_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T, \bar{\mathbf{t}}_S, \bar{\mathbf{t}}_T) \\ r_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_S, \bar{\mathbf{t}}_T, \bar{\mathbf{t}}_S) & r_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_T, \bar{\mathbf{t}}_T, \bar{\mathbf{t}}_T) \end{bmatrix}, \text{ Уравнение (30a)}$$

$$\mathbf{r}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}, t^*, \bar{\mathbf{t}}) \equiv [r_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_S, t^*, \bar{\mathbf{t}}_S) \quad r_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_T, t^*, \bar{\mathbf{t}}_T)]. \text{ Уравнение (30b)}$$

Следует отметить, что, в отличие от переменных $\mathbf{a}$ и $\mathbf{b}$ роль t в уравнениях (30a) и (30b) заключается в простом увеличении вклада $\mathcal{D}_T$. Физически для данного $\mathbf{a}$ переменная t не оказывает влияния на $\mathbf{b}$. Поэтому, если $\bar{\mathbf{t}}_T = [t_{D-k+1}, \dots, t_D]^T$ задан как номер выборки, соответствующий выборкам в $\mathcal{D}_T$,

$\bar{\mathbf{t}}_S = [t_1, \dots, t_{D-k}]^T$ можно задать так, что он удовлетворяет следующему:

$$|t_i - t_j| \gg M, \text{ Уравнение (31a)}$$

$$|t_i - t^*| \gg N, \text{ Уравнение (31b)}$$

$$|t_i - t_k| \gg P, \text{ Уравнение (31c)}$$

для всех $i, j \in \{1, \dots, D-k\}$ и $k \in \{D-k+1, \dots, D\}$, где $M, N, P \in \mathbb{R}_+$ - произвольные большие положительные константы. Кроме того, если предположить, что $\bar{\mathbf{t}}_T$ и $\mathbf{t}^*$ таковы, что $k_{\text{time}}(\bar{\mathbf{t}}_T, \bar{\mathbf{t}}_T > 0)$ и $k_{\text{time}}(\mathbf{t}^*, \bar{\mathbf{t}}_T > 0)$, то для $\bar{\mathbf{t}}_T$ и $\bar{\mathbf{t}}_S$, как описано выше, $\mathbf{r}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S, \bar{\mathbf{t}}_S, \bar{\mathbf{t}}_S)$ можно записать следующим образом:

$$\mathbf{r}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S, \bar{\mathbf{t}}_S, \bar{\mathbf{t}}_S) = \mathbf{k}_{\text{space}}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) + \mathbf{k}_{\text{time}}(\bar{\mathbf{t}}_S, \bar{\mathbf{t}}_S), \text{ Уравнение (32a)}$$

$$\approx \mathbf{k}_{\text{space}}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) + \beta_1 \mathbf{I}_{(D-k)}, \text{ Уравнение (32b)}$$

где уравнение (32b) получено из уравнения (31a), что приводит к нулевым недиагональным элементам в $k_{\text{time}}(\bar{\mathbf{t}}_S, \bar{\mathbf{t}}_S)$. Аналогично, ковариацию $\mathbf{r}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_S, \mathbf{t}^*, \bar{\mathbf{t}}_S)$ и $\mathbf{r}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T, \bar{\mathbf{t}}_S, \bar{\mathbf{t}}_T)$ можно вычислить следующим образом:

$$\mathbf{r}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_S, \mathbf{t}^*, \bar{\mathbf{t}}_S) = \mathbf{k}_{\text{space}}(\mathbf{a}^*, \bar{\mathbf{a}}_S) + \mathbf{k}_{\text{time}}(\mathbf{t}^*, \bar{\mathbf{t}}_S), \text{ Уравнение (33a)}$$

$$\approx \mathbf{k}_{\text{space}}(\mathbf{a}^*, \bar{\mathbf{a}}_S), \text{ Уравнение (33b)}$$

$$\mathbf{r}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T, \bar{\mathbf{t}}_S, \bar{\mathbf{t}}_T) = \mathbf{k}_{\text{space}}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) + \mathbf{k}_{\text{time}}(\bar{\mathbf{t}}_S, \bar{\mathbf{t}}_T), \text{ Уравнение (33c)}$$

$$\approx \mathbf{k}_{\text{space}}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T), \text{ Уравнение (33d)}$$

где уравнение (33b) основано на уравнении (31b), и уравнение (33d) основано на уравнении (31c). Подстановка уравнений (32b), (33b) и (33d) в уравнения (30a) и (30b) дает

$$\mathbf{r}_\theta(\bar{\mathbf{a}}, \bar{\mathbf{a}}, \bar{\mathbf{t}}, \bar{\mathbf{t}}) = \begin{bmatrix} \mathbf{k}_{\text{space}}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) + \beta_1 \mathbf{I}_{(D-k)} & \mathbf{k}_{\text{space}}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) \\ \mathbf{k}_{\text{space}}(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_S) & \mathbf{r}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_T, \bar{\mathbf{t}}_T, \bar{\mathbf{t}}_T) \end{bmatrix}, \text{ Уравнение (34a)}$$

$$\mathbf{r}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}, \mathbf{t}^*, \bar{\mathbf{t}}) = [\mathbf{k}_{\text{space}}(\mathbf{a}^*, \bar{\mathbf{a}}_S) \quad \mathbf{r}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_T, \mathbf{t}^*, \bar{\mathbf{t}}_T)]. \text{ Уравнение (34b)}$$

Исходя из уравнений (30a) и (30b), легко подтвердить, что ковариация $\mathbf{r}_\theta$ включает вклады как от $\mathbf{k}_{\text{space}}$, так и от $\mathbf{k}_{\text{time}}$. Для данных в уравнениях (30a) и (30b) ковариационных функций для пространственно-временной рамановской модели параметр ядра, θ , и дисперсию σ^2 шума можно оценить путем максимизации

$$\log p(\bar{\mathbf{b}}|\bar{\mathbf{a}}, \bar{\mathbf{t}}) = -\frac{1}{2} \bar{\mathbf{b}}^T \mathbf{r}_\gamma^{-1} \bar{\mathbf{b}} - \frac{1}{2} \log |\mathbf{r}_\gamma| - \frac{D}{2} \log 2\pi, \text{ Уравнение (35)}$$

где $\gamma = [\theta, \sigma^2]^T \in \Gamma \subset \mathbb{R}^5$, $\log p(\bar{\mathbf{b}}|\bar{\mathbf{a}}, \bar{\mathbf{t}})$ - логарифмическая функция предельного правдоподобия, и $\mathbf{r}_\gamma = \mathbf{r}_\theta + \mathbf{I}_{D \times D}$. Максимизация уравнения (35) по Γ дает оптимальную оценку, γ^* . Для оптимизаторов на основе градиента градиенты для уравнения (35) относительно γ можно вычислить способом, аналогичным уравнению (10b). Для данного γ^* точечную оценку и апостериорную дисперсию для запроса $(\mathbf{a}^*, \mathbf{t}^*)$ можно вычислить как

$$\hat{\mathbf{b}} = \mathbf{r}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}, \mathbf{t}^*, \bar{\mathbf{t}}) [\mathbf{r}_\gamma(\bar{\mathbf{a}}, \bar{\mathbf{a}}, \bar{\mathbf{t}}, \bar{\mathbf{t}})]^{-1} \bar{\mathbf{b}}, \text{ Уравнение (36a)}$$

$$\mathbf{r}_\theta^* = \mathbf{r}_\theta(\mathbf{a}^*, \mathbf{a}^*, \mathbf{t}^*, \mathbf{t}^*) - \mathbf{r}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}, \mathbf{t}^*, \bar{\mathbf{t}}) [\mathbf{r}_\gamma(\bar{\mathbf{a}}, \bar{\mathbf{a}}, \bar{\mathbf{t}}, \bar{\mathbf{t}})]^{-1} \mathbf{r}_\theta(\bar{\mathbf{a}}, \bar{\mathbf{a}}, \bar{\mathbf{t}}, \bar{\mathbf{t}}), \text{ Уравнение (36b)}$$

где ковариационные функции даны в уравнениях (34a) и (34b). Аналогично, границы правдоподобия ($b^L \leq \hat{\mathbf{b}} \leq b^U$) в отношении точечной оценки в уравнении (36a) можно вычислить следующим образом:

$$b^L = \hat{\mathbf{b}} - 2\sqrt{\mathbf{r}_\gamma^*}, \text{ Уравнение (37a)}$$

$$b^U = \hat{\mathbf{b}} + 2\sqrt{\mathbf{r}_\gamma^*}, \text{ Уравнение (37b)}$$

где $\mathbf{r}_\gamma^* = \mathbf{r}_\theta^* + \sigma^2 \mathbf{I}$. Из уравнений (36a), (37a) и (37b) легко видеть, что вклад в вычисления предсказания на основе модели и границ правдоподобия вносят как релевантные в пространстве, так и релевантные по времени выборки. Наконец, подстановка уравнений (34a) и (34b) в уравнения (36a) и (36b) дает, соответственно, апостериорное среднее и дисперсию. Следует отметить, что в отличие от случая алгоритма 2, предсказание на основе модели в уравнении (36a) и интервалы правдоподобия в уравнениях (37a) и (37b) зависят от $\mathcal{D}_T$, даже когда $\mathcal{D}_T$ не имеет пространственной релевантности. Например, если $\mathcal{D}_T$ не имеет пространственной релевантности (т.е. $\mathbf{k}_{\text{space}}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) \approx 0_{(D-k) \times k}$, и $\mathbf{k}_{\text{space}}(\mathbf{a}^*, \bar{\mathbf{a}}_T) \approx 0_{1 \times k}$), то уравнения (36a) и (36b) можно записать как

$$\mathbf{r}_\theta(\bar{\mathbf{a}}, \bar{\mathbf{a}}, \bar{\mathbf{t}}, \bar{\mathbf{t}}) = \begin{bmatrix} \mathbf{k}_{\text{space}}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) + \beta_1 \mathbf{I}_{(D-k)} & 0_{(D-k) \times k} \\ 0_{k \times (D-k)} & \mathbf{r}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_T, \bar{\mathbf{t}}_T, \bar{\mathbf{t}}_T) \end{bmatrix}, \text{ Уравнение (38a)}$$

$$\mathbf{r}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}, \mathbf{t}^*, \bar{\mathbf{t}}) = [\mathbf{k}_{\text{space}}(\mathbf{a}^*, \bar{\mathbf{a}}_S) \quad \mathbf{k}_{\text{time}}(\mathbf{t}^*, \bar{\mathbf{t}}_T)]. \text{ Уравнение (38b)}$$

Из вышесказанного можно видеть, что уравнения (38a) и (38b) по-прежнему содержат вклады как от $\mathbf{k}_{\text{space}}$, так и от $\mathbf{k}_{\text{time}}$. Примерный алгоритм, формально описывающий методику ST-JITL, представлен ниже в алгоритме 3:

1. **Input:** Library $\mathcal{L}_t = \{(\mathbf{a}_i, \mathbf{b}_i)\}_{i=1}^L$, query point $\mathbf{a}^*$
2. **Output:** Prediction $\hat{\mathbf{b}}$ and uncertainty (b^L, b^U)
3. Set $\mathcal{K} \leftarrow \{\emptyset\}$ and $\bar{\mathbf{t}}_T \leftarrow \{\emptyset\}$
4. **for** $t = 1$ to T **do**
5. Set $I \leftarrow \text{sample_index}(\mathcal{L}_t)$, $\mathcal{D}_S \leftarrow \{\emptyset\}$, $\mathcal{D}_T \leftarrow \{\emptyset\}$
6. **for** $d = 1$ to $D - \text{set_cardinality}(\mathcal{K})$ **do**
7. $i_* \in \arg \max_{i \in I} \text{sim}(\mathbf{a}_i, \mathbf{a}^*)$
8. $\mathcal{D}_S \leftarrow \mathcal{D}_S \cup \{\mathbf{a}_{i_*}, \mathbf{b}_{i_*}\}$
9. $I \leftarrow I \setminus \{i_*\}$
10. **end for**
11. **if** $\text{set_cardinality}(\mathcal{K}) \geq 1$ **then**
12. $\mathcal{D}_T \leftarrow \mathcal{K}$
13. **end if**
14. $\mathcal{D} \leftarrow \mathcal{D}_S \cup \mathcal{D}_T$
15. Задать $\bar{\mathbf{t}}_S$ согласно уравнениям (31a)-(31c)
16. Задать $\bar{\mathbf{t}} \leftarrow [\bar{\mathbf{t}}_S; \bar{\mathbf{t}}_T]$
17. Обучить модель на основе гауссовского процесса в уравнении (28) с использованием $\mathcal{D}$ и $\bar{\mathbf{t}}$, и оценить γ^*
18. Вычислить $\hat{\mathbf{b}}$ с использованием уравнения (36a) и вычислить (b^L, b^U) с использованием уравнений (37a) и (37b)
19. **if** b^* is available **then**
20. **if** $\text{size}(\mathcal{K}) = k$ **then**
21. $\mathcal{L}_t \leftarrow \mathcal{L}_t \cup \text{select_oldest}(\mathcal{K})$
22. $\mathcal{K} \leftarrow \text{delete_oldest}(\mathcal{K})$
23. $\mathcal{K} \leftarrow \mathcal{K} \cup \{\mathbf{a}^*, b^*\}$
24. **end if**
25. $\mathcal{K} \leftarrow \mathcal{K} \cup \{\mathbf{a}^*, b^*\}$
26. **end if**
27. **end for**

Алгоритм 3

Следует отметить, что A-JITL и ST-JITL (в алгоритмах 2 и 3 соответственно) могут быть идентичны в случае, когда $\beta_1 = 0$. Причиной этого является то, что для $\beta_1 = 0$, $k_{\text{time}} = 0$, так что $\mathbf{r}_0 = \mathbf{k}_{\text{space}} = \mathbf{k}_0$ (как видно из уравнений (28) и (29b)).

Обратимся теперь к фиг. 6, на которой показан примерный поток 350 данных, который может иметь место при анализе биофармацевтического процесса с использованием методики ST-JITL, описанной в данном документе. Поток 350 данных может иметь место, например, в системе 100, представленной на фиг. 1, или системе 150, представленной на фиг. 2. В потоке 350 данных спектральные данные 352 предоставляются спектрометром/зондом. Например, спектральные данные 352 могут содержать вектор рамановского сканирования, сгенерированный рамановским анализатором 106, или вектор сканирования NIR и т.д. Точка 354 запроса генерируется (например, блоком 140 запроса) на основе спектральных данных 352 и используется, например, для запроса глобального набора 356 данных, который может содержать все наборы данных наблюдений в базе 136 данных наблюдений. Глобальный набор 356 данных логически разделен на последние k элементов 357A (например, все из текущего эксперимента/процесса) и на все элементы 357B перед последними k элементами 357A (например, из предыдущих, а также, возможно, текущего эксперимента/процесса). Значение k можно определить на основе номера выборки точки 354 запроса. Локальный набор 358 данных можно сгенерировать из элементов 357A и элементов 357B, например, в соответствии с алгоритмом 3.

Локальный набор 358 данных затем используется в качестве обучающих данных (например, генератором 142 локальной модели) для калибровки локальной модели 360 (например, локальной модели 132). Локальная модель 360 затем используется (например, блоком 144 предсказания) для предсказания элемента 362 выходных данных (аналитического измерения), такого как концентрация компонентов питательной среды, состояние питательной среды (например, глюкоза, лактат, глутамат, глутамин, аммиак, аминокислоты, Na^+ , K^+ , а также другие питательные вещества или метаболиты, pH, pCO_2 , pO_2 , температура, осмоляльность и т.д.), плотность жизнеспособных клеток, титр, критические показатели качества,

состояние клеток и т.д., а также, возможно, вывода границ правдоподобия или другого подходящего индикатора достоверности.

Если фактическое аналитическое измерение (например, измерение, выполненное таким аналитическим прибором, как один из аналитических приборов 104) является доступным, создается новый элемент 364 (содержащий номер его выборки) и добавляется в глобальный набор 356 данных. Такие измерения могут быть доступны, например, на основе периодической выборки (например, один или два раза в сутки) и/или их можно сделать доступными в ответ на иницирующий фактор с переменным расчетным временем (например, если некоторое количество предсказаний в строке имеет недопустимо широкие границы правдоподобия и т.д.).

Как указано выше, аналитические измерения могут планироваться/иницироваться на основе текущего и/или недавнего выполнения одной или нескольких локальных моделей (например, локальной модели 132, 260, 310 или 360) с целью сохранения или повышения точности предсказания с одновременным уменьшением использования ресурсов (например, использования аналитических приборов). Данную методику можно использовать, например, совместно с A-JITL, ST-JITL или обычным JITL.

В одном варианте осуществления для инициации обслуживания модели используются интервалы правдоподобия. В частности, если ширина интервала правдоподобия (например, расстояние между границами правдоподобия, вычисленными с использованием уравнения (16) или уравнений (37a), (37b)) по данному предсказанию на основе модели (например, по самому недавнему предсказанию, выполненному локальной моделью 132, 260, 310 или 360) больше предварительно заданного порогового значения, блок 146 обслуживания базы данных может генерировать сообщение с запросом и обеспечивать отправку компьютером 110 сообщения на аналитический прибор (приборы) 104 с целью запроса измерения. В примерных результатах, представленных на фиг. 3, например, блок 146 обслуживания базы данных может инициировать новые аналитические измерения ближе к концу суток 08.12.17, 09.12.17 и 14.12.17, где затененные области 208 указывают широкий интервал правдоподобия (т.е. большое значение $b^U - b^L$).

В ответ на сообщение с запросом аналитический прибор (приборы) 104 выполняет (выполняют) аналитическое измерение (измерения) и предоставляет измерение (измерения) на компьютер 110. Блок 146 обслуживания базы данных может затем отправлять измерение (измерения) и соответствующий вектор (векторы) рамановского сканирования, принятые из рамановского анализатора 106, на сервер 112 базы данных для хранения в базе 136 данных наблюдений. Например, измерение (измерения) и вектор (векторы) сканирования могут быть добавлены в библиотеку $\mathcal{L}$ (для обычного JITL) или библиотеку $\mathcal{J}$ (для A-JITL или ST-JITL), как обсуждено выше.

И, наоборот, если ширина интервала правдоподобия по данному предсказанию на основе модели не превышает предварительно заданное пороговое значение, блок 146 обслуживания базы данных может не запрашивать новое аналитическое измерение, и в этом случае библиотека в базе 136 данных наблюдений остается неизменной. В вариантах осуществления, где аналитический прибор (приборы) 104 содержит несколько приборов, измеряющих разные свойства, такие как концентрация компонентов питательной среды, состояние питательной среды (например, глюкоза, лактат, глутамат, глутамин, аммиак, аминокислоты, Na^+ , K^+ , а также другие питательные вещества или метаболиты, pH, pCO_2 , pO_2 , температура, осмоляльность и т.д.), плотность жизнеспособных клеток, титр, критические показатели качества, состояние клеток и т.д., и для предсказания различных значений свойств используются отдельные локальные модели, процесс планирования может быть реализован по отдельности для каждого предсказываемого свойства и аналитического прибора, измеряющего это свойство, возможно, с отличающимися пороговыми значениями ширины интервала правдоподобия для каждого свойства.

Математически, блок 146 обслуживания базы данных может планировать/инициировать новое аналитическое измерение (измерения) в точке запроса, $\mathbf{a}^*$, при условии:

$$b^U - b^L \geq \text{THR}, \text{ Уравнение (39)}$$

где THR - задаваемое пользователем пороговое значение. В некоторых вариантах осуществления THR может регулироваться пользователем для удовлетворения требованиям определенного применения или случая использования. Например, пользователь может устанавливать относительно небольшое значение THR (используемое блоком 146 обслуживания базы данных) для применения, в котором надежность модели является критичной, что, таким образом, вызывает более частое выполнение операций обслуживания модели/библиотеки. В целом THR может быть установлено в виде разных значений на основе критичности процесса, на основе предсказываемого параметра, такого как концентрация компонентов питательной среды, состояние питательной среды (например, глюкоза, лактат, глутамат, глутамин, аммиак, аминокислоты, Na^+ , K^+ , а также другие питательные вещества или метаболиты, pH, pCO_2 , pO_2 , температура, осмоляльность и т.д.), плотность жизнеспособных клеток, титр, критические показатели качества, состояние клеток и т.д., и/или на основе текущего периода времени (например, использование меньшего THR для более поздних суток культивации по сравнению с начальными сутками). Выбор THR представляет компромисс между точностью модели и использованием ресурсов (аналитического прибора), при этом менее высокие пороговые значения стремятся к повышению точности модели за счет увеличения использования ресурсов.

Также возможны изменения данного протокола планирования. В одном варианте осуществления блок 146 обслуживания базы данных, например, может применять один или несколько критериев качества модели, не только к текущему (самому недавнему) предсказанию, но и к одному или нескольким другим недавним предсказаниям (например, самым недавним N предсказаний, где $N > 1$). В качестве примера такого варианта осуществления блок 146 обслуживания базы данных может вычислять среднюю ширину интервалов правдоподобия для самых недавних N предсказаний ($N \geq 1$), а затем сравнивать эту среднюю ширину с пороговым значением THR. В качестве другого примера блок 146 обслуживания базы данных может идентифицировать X наибольших значений ширины интервала правдоподобия среди последних Y предсказаний ($X < Y$) и планировать/инициировать новое аналитическое измерение только тогда, когда каждое из этих X значений ширины больше порогового значения THR.

На фиг. 7 представлена схема последовательности операций примерного способа 400 анализа биофармацевтического процесса (например, в целях отслеживания и/или управления). Способ 400 может быть реализован таким компьютером, как компьютер 110, представленный на фиг. 1 (например, блоком 120 обработки, исполняющим команды приложения 130 предсказателя JTTL) или фиг. 2, и/или таким сервером, как сервер 112 базы данных, представленный на фиг. 1 или 2, например.

В блоке 402 определяется точка запроса, связанная со сканированием биофармацевтического процесса системой спектроскопии (например, рамановским анализатором 104 и рамановским зондом 106 системы 100 или системы 150). Точку запроса можно определить на основе по меньшей мере частично вектора спектрального сканирования (например, вектора рамановского сканирования или сканирования NIR), который был сгенерирован системой спектроскопии, например, при сканировании биофармацевтического процесса. В зависимости от варианта осуществления точка запроса может быть определена на основе необработанного вектора спектрального сканирования или после подходящей фильтрации путем предварительной обработки необработанного вектора спектрального сканирования. В некоторых вариантах осуществления точка запроса также определяется на основе другой информации, такой как профиль питательной среды, связанный с биофармацевтическим процессом (например, тип текучей среды, конкретные питательные вещества, уровень pH и т.д.), и/или, например, одно или несколько рабочих условий, в которых подвергают анализу биофармацевтический процесс (например, заданное значение концентрации метаболита и т.д.).

В блоке 404 запрашивается база данных наблюдений (например, база 136 данных наблюдений). База данных наблюдений может содержать наборы данных наблюдений, связанные с прошлыми наблюдениями нескольких биофармацевтических процессов. Каждый из наборов данных наблюдений может содержать спектральные данные (например, вектор рамановского сканирования или сканирования NIR) и соответствующее аналитическое измерение (или, в некоторых вариантах осуществления, два или более аналитических измерений). Аналитическое измерение может представлять собой, например, концентрацию компонентов питательной среды, состояние питательной среды (например, глюкоза, лактат, глутамат, глутамин, аммиак, аминокислоты, Na^+ , K^+ , а также другие питательные вещества или метаболиты, pH, pCO_2 , pO_2 , температура, осмоляльность и т.д.), плотность жизнеспособных клеток, титр, критические показатели качества и/или состояние клеток.

Блок 404 может включать выбор в качестве обучающих данных из наборов данных наблюдений тех наборов данных наблюдений, которые удовлетворяют одному или нескольким критериям релевантности относительно точки запроса. Если точка запроса включена в вектор спектрального сканирования, например, блок 404 может включать сравнение этого вектора спектрального сканирования с векторами спектрального сканирования, связанными с каждым из прошлых наблюдений, представленных в базе данных наблюдений (например, путем вычисления евклидова или других расстояний между (1) вектором спектрального сканирования, на котором было основано определение точки запроса, и (2) каждым из векторов спектрального сканирования, связанных с прошлыми наблюдениями, а затем выбора в качестве обучающих данных любого из векторов спектрального сканирования, связанных с прошлыми наблюдениями, для которых определено нахождение в пределах порогового расстояния от вектора спектрального сканирования, на котором было основано определение точки запроса).

В блоке 406 выбранные обучающие данные используются для калибровки локальной модели, которая характерна для отслеживаемого биофармацевтического процесса. Локальная модель (например, локальная модель 132) обучается в блоке 406 с целью предсказания аналитических измерений на основе входных спектральных данных (например, векторов рамановского сканирования или сканирования NIR). В некоторых вариантах осуществления локальная модель представляет собой модель машинного обучения на основе гауссовского процесса.

В блоке 408 аналитическое измерение биофармацевтического процесса предсказывается с использованием локальной модели. Блок 408 может включать использование локальной модели для анализа спектральных данных (например, вектора рамановского сканирования или сканирования NIR), генерируемых системой спектроскопии при сканировании биофармацевтического процесса. Например, блок 408 может включать предсказывание аналитического измерения путем использования локальной модели для обработки того же вектора спектрального сканирования или других спектральных данных, на кото-

рых была основана точка запроса. В зависимости от варианта осуществления локальная модель может использоваться для анализа необработанных спектральных данных (например, необработанного вектора рамановского сканирования) или для анализа спектральных данных после подходящей фильтрации путем предварительной обработки необработанных спектральных данных. В некоторых вариантах осуществления блок 408 также включает определение индикатора достоверности (например, границ правдоподобия, степени достоверности и т.д.), связанного с предсказанным аналитическим измерением биофармацевтического процесса. В некоторых вариантах осуществления локальная модель также предсказывает в блоке 408 одно или несколько дополнительных аналитических измерений.

В некоторых вариантах осуществления способ 400 включает один или несколько дополнительных блоков, не показанных на фиг. 5. Например, способ 400 может включать дополнительный блок, в котором управление по меньшей мере одним параметром биофармацевтического процесса выполняется на основе по меньшей мере частично аналитического измерения, предсказанного в блоке 408. В зависимости от варианта осуществления параметр может относиться к тому же типу, что и предсказанное аналитическое измерение (например, управление концентрацией глюкозы на основе предсказанной концентрации глюкозы), или к другому типу. Для управления параметром (или параметрами) могут использоваться, например, методики управления по модели предсказания (MPC).

В качестве другого примера способ 400 может включать первый дополнительный блок, в котором получается фактическое аналитическое измерение биофармацевтического процесса (например, одним из аналитических приборов 104 или из него в ответ на определение того, что предсказанное аналитическое измерение и, возможно, одно или несколько ранних/недавних измерений не удовлетворяют одному или нескольким критериям качества модели, как обсуждено выше), и второй дополнительный блок, в котором обеспечивается добавление (1) спектральных данных, которые система спектроскопии генерирует при получении фактического аналитического измерения, и (2) фактического аналитического измерения биофармацевтического процесса в базу данных наблюдений (например, путем отправки спектральных данных и аналитического измерения в такой сервер базы данных, как сервер 112 базы данных, или путем непосредственного добавления спектральных данных и аналитического измерения в локальную базу данных наблюдений и т.д.). В тех вариантах осуществления, где предсказывается несколько типов аналитических измерений, несколько фактических аналитических измерений могут быть получены или добавлены в базу данных наблюдений.

В качестве еще одного примера способ 400 может включать один или несколько дополнительных наборов блоков, каждый из которых подобен блокам 402-408. В каждом из этих дополнительных наборов блоков локальная модель может калиброваться путем запроса базы данных наблюдений (или другой базы данных наблюдений) и использоваться для предсказания аналитического измерения другого типа.

Ниже будут рассмотрены дополнительные соображения в отношении настоящего изобретения.

Термины "полипептид" или "белок" везде используются взаимозаменяемо и относятся к молекуле, содержащей два или более аминокислотных остатков, соединенных друг с другом пептидными связями. Полипептиды и белки также включают макромолекулы, имеющие одну или несколько делеций, вставок и/или замен аминокислотных остатков нативной последовательности, то есть полипептида или белка, полученного встречающейся в природе и нерекомбинантной клеткой; или полученного генетически модифицированной или рекомбинантной клеткой, и включают молекулы, имеющие одну или несколько делеций, вставок и/или замен аминокислотных остатков аминокислотной последовательности нативного белка. Полипептиды и белки также включают полимеры из аминокислот, в которых одна или несколько аминокислот являются химическими аналогами соответствующих встречающихся в природе аминокислот и полимеров. Полипептиды и белки также включают модификации, в том числе, но без ограничения, гликозилирование, присоединение липида, сульфатирование, гамма-карбоксилирование остатков глутаминовой кислоты, гидроксильное и АДФ-рибозилирование.

Полипептиды и белки могут представлять научный и коммерческий интерес, в том числе для терапии на основе белков. Белки включают, помимо прочего, секреторируемые белки, несекреторируемые белки, внутриклеточные белки или мембраносвязанные белки. Полипептиды и белки могут быть получены при помощи рекомбинантных клеточных линий животных с использованием методов культивирования клеток и могут называться "рекомбинантными белками". Экспрессируемый белок (белки) может быть получен внутри клетки или секретирован в культуральную среду, из которой он может быть извлечен и/или собран. Белки включают белки, оказывающие терапевтическое воздействие путем связывания с мишенью, в частности с мишенью из тех, которые перечислены ниже, в том числе с полученными из них мишенями, относящимися к ним мишенями и их модификациями.

Белки "антигенсвязывающие белки"

Термин "антигенсвязывающий белок" относится к белкам или полипептидам, содержащим антигенсвязывающую область или антигенсвязывающую часть, которая имеет сильное сродство к другой молекуле (антигену), с которой она связывается. Антигенсвязывающие белки охватывают антитела, пептиды, фрагменты антител, производные антител, аналоги антител, белки слияния (в том числе одноцепочечные вариабельные фрагменты (scFv) и двухцепочечные (двухвалентные) scFv, мутеины, xMAb и химические антигенные рецепторы (CAR)).

ScFv представляет собой фрагмент одноцепочечного антитела, содержащий вариабельные области тяжелой и легкой цепей антитела, связанные вместе. См. патенты США №№ 7741465 и 6319494, а также Eshhar и др., *Cancer Immunol Immunotherapy* (1997) 45: 131-136. ScFv сохраняет способность исходного антитела специфично взаимодействовать с антигеном-мишенью.

Термин "антитело" включает отсылку как к гликозилированным, так и негликозилированным иммуноглобулинам любого изотипа или подкласса или к их антигенсвязывающей области, конкурирующей с интактным антителом за специфичное связывание. Если не указано иное, антитела включают человеческие, гуманизированные, химерные, мультиспецифичные, моноклональные, поликлональные, гетеро-IgG, XmAb, биспецифичные и олигомерные антитела или их антигенсвязывающие фрагменты. Антитела включают типы IgG1-, IgG2-, IgG3- или IgG4. Также включены белки, имеющие антигенсвязывающий фрагмент или область, такую как Fab, Fab', F(ab')₂, Fv, диатела, Fd, dAb, макситела, молекулы одноцепочечных антител, однодоменные VHH, фрагменты определяющей комплементарности области (CDR), scFv, диатела, триатела, тетраатела и полипептиды, содержащие по меньшей мере часть иммуноглобулина, которой достаточно для придания специфичности связыванию антигена с полипептидом-мишенью.

Также включены человеческие, гуманизированные и другие антигенсвязывающие белки, такие как человеческие и гуманизированные антитела, не вызывающие значительные вредные иммунные ответы при введении человеку.

Также включены пептитела, полипептиды, содержащие один или несколько соединенных друг с другом, необязательно посредством линкеров, биологически активных пептидов с Fc-доменом (См. патент США № 6660843, патент США № 7138370 и патент США № 7511012).

Белки также включают генетически модифицированные рецепторы, такие как химерные антигенные рецепторы (CAR или CAR-T) и Т-клеточные рецепторы (TCR). CAR обычно включают антигенсвязывающий домен (такой как scFv) в тандеме с одним или несколькими костимулирующими ("сигнальными") доменами и одним или несколькими активирующими доменами.

Также включены конструкции антител с биспецифичными агентами Т-клеток (BiTE®), которые представляют собой конструкции рекомбинантных белков, полученные из двух гибко связанных связывающих доменов, полученных из антител (см. WO 99/54440 и WO 2005/040220). Один связывающий домен конструкции является специфичным для выбранного опухолеассоциированного поверхностного антигена на клетках-мишенях; второй связывающий домен является специфичным для CD3, субъединицы Т-клеточного рецепторного комплекса на Т-клетках. Конструкции BiTE® также могут включать способность к связыванию с независимым от окружения эпитопом на N-конце цепи CD3 (WO 2008/119567) для более специфичной активации Т-клеток. Конструкции BiTE® с увеличенным периодом полужизни включают слияние малого биспецифичного антитела с белками большего размера, которые предпочтительно не препятствуют терапевтическому воздействию конструкции антитела BiTE®. Примеры таких дополнительных разработок биспецифичных агентов Т-клеток включают биспецифичные молекулы, содержащие Fc, например, описанные в US 2014/0302037, US 2014/0308285, WO 2014/151910 и WO 2015/048272. Альтернативная стратегия заключается в использовании человеческого сывороточного альбумина (HAS), слитого с биспецифичной молекулой, или простое слияние человеческих альбуминсвязывающих пептидов (см., например, WO 2013/128027, WO 2014/140358). Другая стратегия, HLE BiTE®, включает слияние первого домена, связывающегося с поверхностным антигеном клетки-мишени, второго домена, связывающегося с внеклеточным эпитопом цепи CD3ε человека и/или макаки, и третьего домена, представляющего собой домен, обладающий Fc-специфическим способом действия (WO 2017/134140).

Также включены модифицированные белки, такие как белки, модифицированные химически при помощи нековалентного связывания, ковалентного связывания или как ковалентного, так и нековалентного связывания. Также включены белки, дополнительно содержащие одну или несколько посттрансляционных модификаций, которые могут быть выполнены при помощи систем клеточной модификации, или модификаций, вносимых ex vivo при помощи ферментативных и/или химических методов или вносимых другими способами.

Белки также могут включать рекомбинантные белки слияния, содержащие, например, домен мультимеризации, такой как лейциновая застежка, суперспираль, Fc-часть иммуноглобулина и т.п. Также включены белки, содержащие все или часть аминокислотных последовательностей дифференцировочных антигенов (называемых белками CD) или их лигандов, или белки, по существу аналогичные любому из них.

В некоторых вариантах осуществления белки могут включать колониестимулирующие факторы, такие как колониестимулирующий фактор гранулоцитов (G-CSF). Такие средства на основе G-CSF включают без ограничения Neupogen® (филграстим) и Neulasta® (пэгфилграстим). Также включены стимуляторы эритропоэза (ESA), такие как Erogen® (эпоэтин-альфа), Aranesp® (дарбэпоэтин-альфа), Dynepo® (эпоэтин-дельта), Mircera® (метоксиполиэтиленгликоль-эпоэтин-бета), Hematide®, MRK-2578, INS-22, Retacrit® (эпоэтин-дзета), Neorecormon® (эпоэтин-бета), Silapo® (эпоэтин-дзета), Binocrit® (эпоэтин-альфа), эпоэтин-альфа Hexal, Abseamed® (эпоэтин-альфа), Ratioepo® (эпоэтин-тета), Eporatio®

(эпоэтин-тета), Biopoin® (эпоэтин-тета), эпоэтин-альфа, эпоэтин-бета, эпоэтин-дзета, эпоэтин-тета и эпоэтин-дельта, эпоэтин-омега, эпоэтин-йота, тканевой активатор плазминогена, агонисты рецептора GLP-1, а также молекулы или их варианты или аналоги и биосимиляры любого из вышеперечисленного.

В некоторых вариантах осуществления белки могут включать белки, которые специфически связываются с одним или несколькими белками CD, белками семейства рецепторов HER, молекулами клеточной адгезии, факторами роста, факторами роста нервов, факторами роста фибробластов, трансформирующими факторами роста (TGF), инсулиноподобными факторами роста, остеиндуцирующими факторами, инсулином и относящимися к инсулину белками, коагулирующими и относящимися к коагуляции белками, колониестимулирующими факторами (CSF), другими белками крови и сыворотки, антигенами групп крови, рецепторами, рецептор-ассоциированными белками, гормонами роста, рецепторами гормона роста, рецепторами Т-клеток; нейротрофическими факторами, нейротрофинами, релаксинами, интерферонами, интерлейкинами, вирусными антигенами, липопротеинами, интегринами, ревматоидными факторами, иммунотоксинами, поверхностными мембранными белками, транспортными белками, рецепторами самонаведения, адресинами, регуляторными белками и иммуноадгезинами. В некоторых вариантах осуществления белки могут включать белки, связывающиеся с одним или несколькими из следующего, по отдельности или в любой комбинации: CD белками, в том числе, но без ограничения, CD3, CD4, CD5, CD7, CD8, CD19, CD20, CD22, CD25, CD30, CD33, CD34, CD38, CD40, CD70, CD123, CD133, CD138, CD171 и CD174, белками семейства рецепторов HER, в том числе, например, HER2, HER3, HER4 и рецептором EGF, EGFRvIII, молекулами клеточной адгезии, например LFA-1, Mol, p150,95, VLA-4, ICAM-1, VCAM и интегрин альфа v/бета 3, факторами роста, в том числе, но без ограничения, например, фактором роста эндотелия сосудов ("VEGF"); VEGFR2, гормоном роста, тиреостимулирующим гормоном, фолликулостимулирующим гормоном, лютеинизирующим гормоном, рилизинг-фактором гормона роста, паратиреоидным гормоном, мюллеровым ингибирующим фактором, воспалительным белком макрофагов человека (MIP-1-альфа), эритропоэтином (EPO), фактором роста нервов, таким как NGF-бета, фактором роста тромбоцитов (PDGF), фактором роста фибробластов, в том числе, например, aFGF и bFGF, эпидермальным фактором роста (EGF), Cripto, трансформирующими факторами роста (TGF), в том числе, помимо прочего, TGF- α и TGF- β , в том числе TGF- β 1, TGF- β 2, TGF- β 3, TGF- β 4 или TGF- β 5, инсулиноподобными факторами роста-I и -II (IGF-I и IGF-II), des(1-3)-IGF-I (мозговой IGF-I) и остеиндуцирующими факторами, инсулинами и относящимися к инсулину белками, в том числе, но без ограничения инсулином, А-цепью инсулина, В-цепью инсулина, проинсулином и инсулиноподобными белками, связывающими фактор роста; (белками, относящимися к коагуляции, такими как, среди прочего, фактор VIII, тканевой фактор, фактор фон Виллебранда, протеин С, альфа-1-антитрипсин, активаторами плазминогена, такими как урокиназа и тканевый активатор плазминогена ("t-PA"), бомбазином, тромбином, тромбопоэтином и рецептором тромбопоэтина, колониестимулирующими факторами (CSF), в том числе следующими, среди прочего, М-CSF, GM-CSF и G-CSF, другими белками крови и сыворотки, в том числе, но без ограничения, альбумин, IgE и антигены групп крови, рецепторами и ассоциированными с рецептором белками, в том числе, например, рецептором flk2/flt3, рецептором ожирения (OB), рецепторами гормона роста и рецепторами Т-клеток; (x) нейротрофическими факторами, в том числе, но без ограничения, нейротропным фактором костной ткани (BDNF) и нейротрофином-3, -4, -5 или -6 (NT-3, NT-4, NT-5 или NT-6); (xi) А-цепью релаксина, В-цепью релаксина и прорелаксином, интерферонами, в том числе, например, интерферонами-альфа, -бета и -гамма, интерлейкинами (IL), например, IL-1-IL-10, IL-12, IL-15, IL-17, IL-23, IL-12/IL-23, IL-2Ra, IL1-R1, рецептором IL-6, рецептором IL-4 и/или рецепторами IL-13-IL-13RA2, или рецептором IL-17, IL-1RAP; (xiv) вирусными антигенами, в том числе, но без ограничения, антигеном оболочки вируса СПИДа, липопротеинами, кальцитонином, глюкагоном, предсердным натрийуретическим фактором, сурфактантом легких, альфа- и бета-факторами некроза опухоли, энкефалиназой, BCMA, IgKappa, ROR-1, ERBB2, мезотелином, RANTES (регулируется при активации, обычно экспрессируемой и секретируемой Т-клетками), мышинным гонадотропин-ассоциированным пептидом, ДНКазой, FR-альфа, ингибином и активином, интегрином, белком А или D, ревматоидными факторами, иммунотоксинами, костным морфогенетическим белком (BMP), супероксиддисмутазой, поверхностными мембранными белками, фактором ускорения распада (DAF), оболочкой вируса СПИДа, транспортными белками, хоминг-рецепторами, MIC (MIC-a, MIC-B), ULBP 1-6, EPCAM, адресинами, регуляторными белками, иммуноадгезинами, антигенсвязывающими белками, соматропином, CTGF, CTLA4, эотаксином-1, MUC1, CEA, с-MET, Claudin-18, GPC-3, EPHA2, FPA, LMP1, MG7, NY-ESO-1, PSCA, ганглиозидом GD2, ганглиозидом GM2, BAFF, OPGL (RANKL), миостатином, Dickkopf-1 (DKK-1), Ang2, NGF, рецептором IGF-1, фактором роста гепатоцитов (HGF), TRAIL-R2, c-Kit, B7RP-1, PSMA, NKG2D-1, белком 1 запрограммированной гибели клеток и лигандом, PD1 и PDL1, рецептором маннозы/hCG β , вирусом гепатита С, конъюгатом мезотелина dsFv[PE38, *Legionella pneumophila* (Ily), IFN гамма, интерферон-гамма-индуцированным белком 10 (IP10), IFNAR, TALL-1, тимусным стромальным лимфопоэтином (TSLP), пропротеинконвертазой субтилизином/кексином типа 9 (PCSK9), факторами стволовых клеток, Flt-3, пептидом, связанным с геном кальцитонина (CGRP), OX40L, α 4 β 7, специфичным к тромбоцитам (гликопротеином тромбоцитов Iib/IIIb (PAC-1), трансформирующим фактором роста бета (TFG β), бел-

ком 3 Zona pellucida, связывающим сперматозоиды (ZP-3), TWEAK, рецептором фактора роста тромбоцитов альфа (PDGFR α), склеростином и биологически активными фрагментами или вариантами любого из вышеперечисленного.

В другом варианте осуществления белки включают абциксимаб, адалимумаб, адекатумумаб, афлиберцепт, алемтузумаб, алирокумаб, анакинру, атасцепт, базиликсимаб, белимумаб, бевацизумаб, биосозумаб, блинатумомаб, брентуксимаб ведотин, бродалумаб, кантузумаб мертанзин, канакинумаб, цетуксимаб, цертолизумаб пегол, конатумумаб, даклизумаб, деносумаб, экулизумаб, эдреколомаб, эфализумаб, эспратузумаб, этанерцепт, эволокумаб, галиксимаб, ганитумаб, гемтузумаб, голимумаб, ибритумомаб тиуксетан, инфликсимаб, ипилимумаб, лерделимумаб, люмиликсимаб, lxd-кизумаб, мапатумумаб, мотесаниб дифосфат, муромонаб-CD3, натализумаб, несиритид, нимотузумаб, ниволумаб, окрелизумаб, офатумумаб, омализумаб, опрелвекин, паливизумаб, панитумумаб, пембролизумаб, пертузумаб, пекселизумаб, ранибизумаб, рилотумумаб, ритуксимаб, ромиплостим, ромосозумаб, саргамостим, тоцилизумаб, тозитумомаб, трастузумаб, устекинумаб, ведолизумаб, вилизумаб, волоциксимаб, занолимуаb, залутумумаб и биосимиляры любого из вышеперечисленного.

Белки включают все вышеперечисленное и дополнительно включают антитела, содержащие 1, 2, 3, 4, 5 или 6 определяющих комплементарность областей (CDR) любого из вышеупомянутых антител. Также включены варианты, содержащие область, которая на 70% или более, в частности на 80% или более, в частности на 90% или более, еще более конкретно на 95% или более, в частности на 97% или более, в частности на 98% или более, еще более конкретно на 99% или более идентична по аминокислотной последовательности эталонной аминокислотной последовательности представляющего интерес белка. В данном отношении идентичность может быть определена с использованием разнообразного хорошо известного и легкодоступного программного обеспечения для анализа аминокислотных последовательностей. Предпочтительное программное обеспечение включает то программное обеспечение, которое реализует алгоритмы Смита-Уотермана, которые считаются удовлетворительным решением задачи поиска и выравнивания последовательностей. Могут также использоваться другие алгоритмы, в частности тогда, когда важным критерием является скорость. Обычно используемые программы для выравнивания и гомологического соответствия ДНК, РНК и полипептидов, которые можно использовать в этой связи, включают FASTA, TFASTA, BLASTN, BLASTP, BLASTX, TBLASTN, PROSRCH, BLAZE и MPSRCH, причем последняя является реализацией алгоритма Смита-Уотермана для исполнения на массово-параллельных процессорах, изготавливаемых MasPar.

Некоторые из фигур, описанных в настоящем документе, иллюстрируют примерные блок-схемы, на которых представлены один или несколько функциональных компонентов. Следует понимать, что такие структурные схемы представлены для иллюстративных целей, и описанные и показанные устройства могут иметь дополнительные или альтернативные компоненты, или меньшее количество компонентов, чем проиллюстрировано. Дополнительно в различных вариантах осуществления компоненты (а также функциональные возможности, предоставленные соответствующими компонентами) могут быть связаны с любым подходящим компонентом или иным образом интегрированы в виде его части.

Варианты осуществления настоящего изобретения относятся к энергозависимому машиночитаемому носителю данных, который содержит программный код для выполнения различных действий, осуществляемых компьютером. Термин "машиночитаемый носитель данных" используют в данном документе для описания любого носителя, способного хранить или кодировать последовательность команд или компьютерных кодов для выполнения действий, способов и методов, описанных в данном документе. Носители и компьютерный код могут быть специально созданы и выполнены для достижения целей вариантов осуществления настоящего изобретения, или они могут быть широко известны и доступны для специалистов в области программного обеспечения. Примеры машиночитаемых носителей данных включают, но без ограничения магнитные носители, такие как жесткие диски, гибкие диски и магнитные пленки; оптические носители, такие как CD-ROM и голографические устройства; магнитно-оптические носители, такие как оптические диски; и аппаратные устройства, которые специально выполнены с возможностью хранения и исполнения программного кода, такие как ASIC, программируемые логические устройства ("PLD") и устройства ROM и RAM.

Примеры компьютерного кода включают машинный код, например написанный компилятором, и файлы, содержащие код более высокого уровня, которые исполняются компьютером за счет использования интерпретатора или компилятора. Например, вариант осуществления настоящего изобретения может быть реализован за счет использования Java, C++ или других объектно-ориентированных языков программирования и средств разработки. Дополнительные примеры компьютерного кода включают зашифрованный код и сжатый код. Более того, вариант осуществления настоящего изобретения может быть загружен в виде компьютерного программного продукта, который может быть передан с удаленного компьютера (например, серверного компьютера) на запрашивающий компьютер (например, на компьютер клиента или другой серверный компьютер) посредством канала передачи данных. Другой вариант осуществления настоящего изобретения может быть реализован в виде кабельной схемы вместо программных команд, исполняемых компьютерами, или в сочетании с ними.

В контексте данного документа формы существенного единственного числа также могут обо-

значать формы множественного числа, если в контексте явно не обозначено иное.

В контексте данного документа термины "соединять", "соединенный" и "соединение" обозначают функциональное объединение или сцепление. Соединенные компоненты могут быть непосредственно или опосредовано присоединены друг к другу, например, посредством еще одного набора компонентов.

В контексте данного документа термины "приблизительно", "по существу", "существенный" и "примерно" используются для описания и учета небольших отклонений. Когда термины используются в сочетании с определенным случаем или обстоятельством, они могут относиться к ситуациям, при которых случай и обстоятельство происходят определенным образом, а также ситуациям, при которых случай и обстоятельство происходят с некоторой точностью. Например, когда термины используются в сочетании с числовым значением, они могут обозначать диапазон отклонения, которое меньше или равно $\pm 10\%$ от этого числового значения, например меньше или равно $\pm 5\%$, меньше или равно $\pm 4\%$, меньше или равно $\pm 3\%$, меньше или равно $\pm 2\%$, меньше или равно $\pm 1\%$, меньше или равно $\pm 0,5\%$, меньше или равно $\pm 0,1\%$ или меньше или равно $\pm 0,05\%$. Например, подразумевается, что два числовых значения могут быть "по существу" одинаковыми, если разница между значениями меньше или равна $\pm 10\%$ от среднего из значений, например меньше или равна $\pm 5\%$, меньше или равна $\pm 4\%$, меньше или равна $\pm 3\%$, меньше или равна $\pm 2\%$, меньше или равна $\pm 1\%$, меньше или равна $\pm 0,5\%$, меньше или равна $\pm 0,1\%$ или меньше или равна $\pm 0,05\%$.

Дополнительно количества, соотношения и иные числовые значения иногда представлены в данном документе в формате диапазона. Следует понимать, что такой формат диапазона используется для удобства и краткости, и его следует считать гибким для включения численных значений, явно указанных в качестве границ диапазона, а также для включения всех отдельных численных значений или поддиапазонов, содержащихся в рамках этого диапазона, как если каждое численное значение и поддиапазон указаны явным образом.

Хотя настоящее изобретение было описано и проиллюстрировано со ссылками на свои определенные варианты осуществления, эти описания и иллюстрации не ограничивают настоящее изобретение. Специалистам в данной области техники следует понимать, что могут быть внесены различные изменения, и могут быть использованы эквиваленты без отступления от сущности и объема настоящего изобретения, которые определены прилагаемой формулой изобретения. Графические материалы не обязательно изображены в масштабе. В настоящем изобретении между схематичными представлениями и фактическим устройством могут существовать различия вследствие производственных процессов, допусков и/или по другим причинам. Могут существовать другие варианты осуществления настоящего изобретения, которые не были конкретно проиллюстрированы. Описание (но не формулу изобретения) и графические материалы следует считать иллюстративными, а не ограничительными. Изменения могут быть внесены для приспособления определенной ситуации, материала, состава вещества, метода или процесса к цели, сущности и объему настоящего изобретения. Все подобные изменения находятся в пределах объема согласно приложенной формуле изобретения. Хотя методы, раскрытые в данном документе, описаны со ссылкой на определенные действия, выполняемые в определенном порядке, следует понимать, что эти действия могут быть объединены, подразделены или переупорядочены для создания подобного способа без отступления от идей настоящего изобретения. Соответственно, если конкретно не указано иное, порядок и группирование действий не ограничивают настоящее изобретение.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Реализуемый на компьютере способ отслеживания процесса биофармацевтического производства и/или управления им, при этом способ включает

определение одним или более процессорами точки запроса базы данных наблюдений, связанной со сканированием процесса биофармацевтического производства системой спектроскопии;

запрос одним или более процессорами базы данных наблюдений, содержащей множество наборов данных наблюдений, связанных с прошлыми наблюдениями процессов биофармацевтического производства, при этом каждый из наборов данных наблюдений содержит спектральные данные и соответствующее фактическое аналитическое измерение, и при этом запрос базы данных наблюдений включает выбор в качестве обучающих данных из множества наборов данных наблюдений наборов данных наблюдений, которые удовлетворяют одному или более критериям релевантности относительно точки запроса базы данных наблюдений;

адаптацию одним или более процессорами и с использованием выбранных обучающих данных локальной модели, характерной для процесса биофармацевтического производства, при этом локальную модель обучают предсказывать аналитические измерения на основе входных спектральных данных; и

предсказание одним или более процессорами аналитического измерения процесса биофармацевтического производства, при этом предсказание аналитического измерения процесса биофармацевтического производства включает использование локальной модели для анализа спектральных данных, которые сгенерировала система спектроскопии при сканировании процесса биофармацевтического производства.

2. Реализуемый на компьютере способ по п.1, в котором система спектроскопии представляет собой

систему рамановской спектроскопии.

3. Реализуемый на компьютере способ по п.1 или 2, в котором определение точки запроса базы данных наблюдений включает определение точки запроса базы данных наблюдений на основе, по меньшей мере частично, вектора спектрального сканирования, при этом вектор спектрального сканирования генерируют системой спектроскопии при сканировании процесса биофармацевтического производства и

выбор в качестве обучающих данных наборов данных наблюдений, которые удовлетворяют одному или более критериям релевантности относительно точки запроса базы данных наблюдений, включает сравнение вектора спектрального сканирования, на котором было основано определение точки запроса базы данных наблюдений, с векторами спектрального сканирования, связанными с прошлыми наблюдениями процессов биофармацевтического производства.

4. Реализуемый на компьютере способ по п.3, в котором определение точки запроса базы данных наблюдений дополнительно включает определение точки запроса базы данных наблюдений на основе номера выборки, связанного с вектором спектрального сканирования; и

выбор в качестве обучающих данных наборов данных наблюдений, которые удовлетворяют одному или более критериям релевантности относительно точки запроса базы данных наблюдений, включает (i) сравнение вектора спектрального сканирования, на котором было основано определение точки запроса из базы данных наблюдений, с векторами спектрального сканирования, связанными с прошлыми наблюдениями процессов биофармацевтического производства, и (ii) сравнение номера выборки, связанного с точкой запроса из базы данных наблюдений, с номерами выборок, связанными с прошлыми наблюдениями процессов из биофармацевтического производства.

5. Реализуемый на компьютере способ по п.4, в котором выбор в качестве обучающих данных наборов данных наблюдений, которые удовлетворяют одному или более критериям релевантности относительно точки запроса базы данных наблюдений, включает

выбор самых недавних k наборов данных наблюдений для включения в обучающие данные.

6. Реализуемый на компьютере способ по любому из пп.3-5, в котором предсказание аналитического измерения процесса биофармацевтического производства включает

использование локальной модели для анализа вектора спектрального сканирования, на котором было основано определение точки запроса базы данных наблюдений.

7. Реализуемый на компьютере способ по любому из пп.3-6, в котором выбор в качестве обучающих данных наборов данных наблюдений, которые удовлетворяют одному или более критериям релевантности относительно точки запроса базы данных наблюдений, включает

вычисление расстояний между (i) вектором спектрального сканирования, на котором было основано определение точки запроса базы данных наблюдений, и (ii) векторами спектрального сканирования, связанными с прошлыми наблюдениями процессов биофармацевтического производства; и

выбор в качестве обучающих данных любого из векторов спектрального сканирования, связанных с прошлыми наблюдениями, для которых определено нахождение в пределах порогового расстояния от вектора спектрального сканирования, на котором было основано определение точки запроса базы данных наблюдений.

8. Реализуемый на компьютере способ по любому из пп.1-7, в котором определение точки запроса базы данных наблюдений включает определение точки запроса базы данных наблюдений на основе, по меньшей мере, частично одного или обоих из (i) профиля питательной среды, связанного с процессом биофармацевтического производства, и (ii) одного или более рабочих условий, в которых подвергают анализу процесс биофармацевтического производства.

9. Реализуемый на компьютере способ по любому из пп.1-8, в котором адаптация локальной модели, характерной для процесса биофармацевтического производства, включает

адаптацию модели машинного обучения на основе гауссовского процесса, характерной для этого процесса биофармацевтического производства.

10. Реализуемый на компьютере способ по любому из пп.1-9, в котором адаптация локальной модели, характерной для процесса биофармацевтического производства, включает

адаптацию модели, которая зависит как от спектральных данных, так и от номера выборки данного набора данных наблюдений.

11. Реализуемый на компьютере способ по любому из пп.1-10, в котором предсказание аналитического измерения процесса биофармацевтического производства включает

использование локальной модели для определения индикатора достоверности, связанного с предсказанным аналитическим измерением процесса биофармацевтического производства.

12. Реализуемый на компьютере способ по любому из пп.1-11, который дополнительно включает управление одним или более процессорами и на основе, по меньшей мере, частично предсказанного аналитического измерения процесса биофармацевтического производства по меньшей мере одним параметром процесса биофармацевтического производства.

13. Реализуемый на компьютере способ по любому из пп.1-12, в котором предсказанное аналитическое

ское измерение процесса биофармацевтического производства представляет собой концентрацию компонентов питательной среды, состояние питательной среды, плотность жизнеспособных клеток, титр, критический показатель качества или состояние клеток.

14. Реализуемый на компьютере способ по любому из пп.1-13, в котором предсказанное аналитическое измерение процесса биофармацевтического производства представляет собой концентрацию глюкозы, лактата, глутамата, глутамина, аммиака, аминокислот, Na^+ или K^+ .

15. Реализуемый на компьютере способ по любому из пп.1-13, в котором предсказанное аналитическое измерение процесса биофармацевтического производства представляет собой pH, pCO_2 , pO_2 , температуру или осмоляльность.

16. Реализуемый на компьютере способ по любому из пп.1-15, который дополнительно включает получение при помощи аналитического прибора фактического аналитического измерения процесса биофармацевтического производства и

обеспечение одним или более процессорами добавления (i) спектральных данных, которые сгенерировала система спектроскопии при получении фактического аналитического измерения, и (ii) фактического аналитического измерения процесса биофармацевтического производства в базу данных наблюдений.

17. Реализуемый на компьютере способ по п.16, который дополнительно включает определение одним или более процессорами того, что, по меньшей мере, предсказанное аналитическое измерение не удовлетворяет одному или более критериям качества модели,

при этом получение фактического аналитического измерения выполняют в ответ на определение того, что, по меньшей мере, предсказанное аналитическое измерение не удовлетворяет одному или более критериям качества модели.

18. Реализуемый на компьютере способ по п.17, в котором определение того, что, по меньшей мере, предсказанное аналитическое измерение не удовлетворяет одному или более критериям качества модели, включает

генерирование интервала правдоподобия, связанного с предсказанным аналитическим измерением; и

сравнение интервала правдоподобия с предварительно заданным пороговым значением.

19. Реализуемый на компьютере способ по любому из пп.1-18, в котором процесс биофармацевтического производства представляет собой процесс культивирования клеток.

20. Система спектроскопии для отслеживания процесса биофармацевтического производства и/или управления им, при этом система спектроскопии содержит

один или более спектроскопических зондов, совместно выполненных с возможностью (i) обеспечения электромагнитного излучения источника для процесса биофармацевтического производства и (ii) сбора электромагнитного излучения во время обеспечения электромагнитного излучения источника для процесса биофармацевтического производства;

одно или более запоминающих устройств, совместно хранящих базу данных наблюдений, содержащую множество наборов данных наблюдений, связанных с прошлыми наблюдениями процессов биофармацевтического производства, при этом каждый из наборов данных наблюдений содержит спектральные данные и соответствующее фактическое аналитическое измерение и

один или более процессоров, выполненных с возможностью определения точки запроса базы данных наблюдений, связанной со сканированием процесса биофармацевтического производства системой спектроскопии,

запроса базы данных наблюдений, по меньшей мере, путем выбора в качестве обучающих данных из множества наборов данных наблюдений наборов данных наблюдений, которые удовлетворяют одному или более критериям релевантности относительно точки запроса базы данных наблюдений,

адаптации с использованием выбранных обучающих данных локальной модели, характерной для процесса биофармацевтического производства, при этом локальная модель обучена предсказывать аналитические измерения на основе входных спектральных данных, и

предсказания аналитического измерения процесса биофармацевтического производства, по меньшей мере, путем использования локальной модели для анализа спектральных данных, которые сгенерировала система спектроскопии при сканировании процесса биофармацевтического производства при помощи одного или более спектроскопических зондов.

21. Система спектроскопии по п.20, в которой система спектроскопии представляет собой систему рамановской спектроскопии.

22. Система спектроскопии по п.20 или 21, в которой один или более процессоров выполнены с возможностью

определения точки запроса базы данных наблюдений на основе, по меньшей мере частично, вектора спектрального сканирования, при этом вектор спектрального сканирования сгенерирован системой спектроскопии при сканировании процесса биофармацевтического производства; и

выбора обучающих данных, по меньшей мере, путем сравнения вектора спектрального сканирования, на котором было основано определение точки запроса базы данных наблюдений, с векторами спек-

трального сканирования, связанными с прошлыми наблюдениями процессов биофармацевтического производства.

23. Система спектроскопии по п.22, в которой один или более процессоров выполнены с возможностью

определения точки запроса базы данных наблюдений на основе частично номера выборки, связанного с вектором спектрального сканирования; и

выбора в качестве обучающих данных наборов данных наблюдений, которые удовлетворяют одному или более критериям релевантности относительно точки запроса базы данных наблюдений, частично путем (i) сравнения вектора спектрального сканирования, на котором было основано определение точки запроса базы данных наблюдений, с векторами спектрального сканирования, связанными с прошлыми наблюдениями процессов биофармацевтического производства, и (ii) сравнения номера выборки, связанного с точкой запроса базы данных наблюдений, с номерами выборок, связанными с прошлыми наблюдениями процессов биофармацевтического производства.

24. Система спектроскопии по п.23, в которой один или более процессоров выполнены с возможностью выбора в качестве обучающих данных наборов данных наблюдений, которые удовлетворяют одному или более критериям релевантности относительно точки запроса базы данных наблюдений, частично путем

выбора самых недавних k наборов данных наблюдений для включения в обучающие данные.

25. Система спектроскопии по любому из пп.20-24, в которой локальная модель представляет собой модель машинного обучения на основе гауссовского процесса.

26. Система спектроскопии по любому из пп.20-25, в которой локальная модель зависит как от спектральных данных, так и от номера выборки данного набора данных наблюдений.

27. Система спектроскопии по любому из пп.20-26, в которой один или более процессоров дополнительно выполнены с возможностью

использования локальной модели для определения индикатора достоверности, связанного с предсказанным аналитическим измерением процесса биофармацевтического производства.

28. Система спектроскопии по любому из пп.20-27, в которой один или более процессоров дополнительно выполнены с возможностью

управления на основе, по меньшей мере, частично предсказанного аналитического измерения процесса биофармацевтического производства по меньшей мере одним параметром процесса биофармацевтического производства.

29. Система спектроскопии по любому из пп.20-28, в которой предсказанное аналитическое измерение процесса биофармацевтического производства представляет собой концентрацию компонентов питательной среды, состояние питательной среды, плотность жизнеспособных клеток, титр, критический показатель качества или состояние клеток.

30. Система спектроскопии по любому из пп.20-29, в которой предсказанное аналитическое измерение процесса биофармацевтического производства представляет собой концентрацию глюкозы, лактата, глутамата, глутамина, аммиака, аминокислот, Na^+ или K^+ .

31. Система спектроскопии по любому из пп.20-29, в которой предсказанное аналитическое измерение процесса биофармацевтического производства представляет собой pH, pCO_2 , pO_2 , температуру или осмоляльность.

32. Система спектроскопии по любому из пп.20-31, которая дополнительно содержит аналитический прибор, выполненный с возможностью получения фактического аналитического измерения процесса биофармацевтического производства,

при этом один или более процессоров дополнительно выполнены с возможностью обеспечения добавления (i) спектральных данных, которые сгенерировала система спектроскопии при получении фактического аналитического измерения, и (ii) фактического аналитического измерения процесса биофармацевтического производства в базу данных наблюдений.

33. Система спектроскопии по п.32, в которой один или более процессоров дополнительно выполнены с возможностью

определения того, что, по меньшей мере, предсказанное аналитическое измерение не удовлетворяет одному или более критериям качества модели; и

получения фактического аналитического измерения из аналитического прибора в ответ на определение того, что, по меньшей мере, предсказанное аналитическое измерение не удовлетворяет одному или более критериям качества модели.

34. Система спектроскопии по п.33, в которой один или более процессоров выполнены с возможностью определения того, что, по меньшей мере, предсказанное аналитическое измерение не удовлетворяет одному или более критериям качества модели, по меньшей мере, путем

генерирования интервала правдоподобия, связанного с предсказанным аналитическим измерением; и сравнения интервала правдоподобия с предварительно заданным пороговым значением.

35. Система спектроскопии по любому из пп.20-34, в которой процесс биофармацевтического производства представляет собой процесс культивирования клеток.

36. Энергонезависимый машиночитаемый носитель данных, на котором хранятся команды для отслеживания процесса биофармацевтического производства и/или управления им, при этом команды при исполнении одним или более процессорами обеспечивают выполнение одним или более процессорами реализуемого на компьютере способа отслеживания процесса биофармацевтического производства и/или управления им по любому одному из пп.1-19.

37. Биореакторная система, содержащая камеру биореактора, выполненную для вмещения процесса биофармацевтического производства; один или более спектроскопических зондов, совместно выполненных с возможностью (i) обеспечения электромагнитного излучения источника для процесса биофармацевтического производства и (ii) сбора электромагнитного излучения во время обеспечения электромагнитного излучения источника для процесса биофармацевтического производства;

одно или более запоминающих устройств, совместно хранящих базу данных наблюдений, содержащую множество наборов данных наблюдений, связанных с прошлыми наблюдениями процессов биофармацевтического производства, при этом каждый из наборов данных наблюдений содержит спектральные данные и соответствующее фактическое аналитическое измерение; и

один или более процессоров, выполненных с возможностью определения точки запроса базы данных наблюдений, связанной со сканированием процесса биофармацевтического производства системой спектроскопии,

запроса базы данных, по меньшей мере, путем выбора в качестве обучающих данных из множества наборов данных наблюдений наборов данных наблюдений, которые удовлетворяют одному или более критериям релевантности относительно точки запроса базы данных наблюдений,

адаптации с использованием выбранных обучающих данных локальной модели, характерной для процесса биофармацевтического производства, при этом локальная модель обучена предсказывать аналитические измерения на основе входных спектральных данных, и

предсказания аналитического измерения процесса биофармацевтического производства, по меньшей мере, путем использования локальной модели для анализа спектральных данных, которые сгенерировала система спектроскопии при сканировании процесса биофармацевтического производства при помощи одного или более спектроскопических зондов.

38. Биореакторная система по п.37, в которой один или более процессоров выполнены с возможностью

определения точки запроса базы данных наблюдений на основе, по меньшей мере частично, вектора спектрального сканирования, при этом вектор спектрального сканирования сгенерирован при сканировании процесса биофармацевтического производства при помощи одного или более спектроскопических зондов; и

выбора обучающих данных, по меньшей мере, путем сравнения вектора спектрального сканирования, на котором было основано определение точки запроса базы данных наблюдений, с векторами спектрального сканирования, связанными с прошлыми наблюдениями процессов биофармацевтического производства.

39. Биореакторная система по п.38, в которой один или более процессоров выполнены с возможностью

определения точки запроса базы данных наблюдений на основе частично номера выборки, связанного с вектором спектрального сканирования; и

выбора в качестве обучающих данных наборов данных наблюдений, которые удовлетворяют одному или более критериям релевантности относительно точки запроса базы данных наблюдений, частично путем (i) сравнения вектора спектрального сканирования, на котором было основано определение точки запроса базы данных наблюдений, с векторами спектрального сканирования, связанными с прошлыми наблюдениями процессов биофармацевтического производства, и (ii) сравнения номера выборки, связанного с точкой запроса базы данных наблюдений, с номерами выборок, связанными с прошлыми наблюдениями процессов биофармацевтического производства.

40. Биореакторная система по п.39, в которой один или более процессоров выполнены с возможностью выбора в качестве обучающих данных наборов данных наблюдений, которые удовлетворяют одному или более критериям релевантности относительно точки запроса базы данных наблюдений, частично путем

выбора самых недавних k наборов данных наблюдений для включения в обучающие данные.

41. Биореакторная система по любому из пп.37-40, в которой локальная модель представляет собой модель машинного обучения на основе гауссовского процесса.

42. Биореакторная система по любому из пп.37-41, в которой локальная модель зависит как от спектральных данных, так и от номера выборки данного набора данных наблюдений.

43. Биореакторная система по любому из пп.37-42, в которой один или более процессоров дополнительно выполнены с возможностью

использования локальной модели для определения индикатора достоверности, связанного с предсказанным аналитическим измерением процесса биофармацевтического производства.

44. Биореакторная система по любому из пп.37-43, в которой один или более процессоров дополнительно выполнены с возможностью

управления на основе, по меньшей мере, частично предсказанного аналитического измерения процесса биофармацевтического производства по меньшей мере одним параметром процесса биофармацевтического производства.

45. Биореакторная система по любому из пп.37-44, в которой предсказанное аналитическое измерение процесса биофармацевтического производства представляет собой концентрацию компонентов питательной среды, состояние питательной среды, плотность жизнеспособных клеток, титр, критический показатель качества или состояние клеток.

46. Биореакторная система по любому из пп.37-45, в которой предсказанное аналитическое измерение процесса биофармацевтического производства представляет собой концентрацию глюкозы, лактата, глутамата, глутамина, аммиака, аминокислот, Na^+ или K^+ .

47. Биореакторная система по любому из пп.37-45, в которой предсказанное аналитическое измерение процесса биофармацевтического производства представляет собой pH, pCO_2 , pO_2 , температуру или осмоляльность.

48. Биореакторная система по любому из пп.37-47, которая дополнительно содержит аналитический прибор, выполненный с возможностью получения фактического аналитического измерения процесса биофармацевтического производства,

при этом один или более процессоров дополнительно выполнены с возможностью обеспечения добавления (i) спектральных данных, которые сгенерировала система спектроскопии при получении фактического аналитического измерения, и (ii) фактического аналитического измерения процесса биофармацевтического производства в базу данных наблюдений.

49. Биореакторная система по п.48, в которой один или более процессоров дополнительно выполнены с возможностью

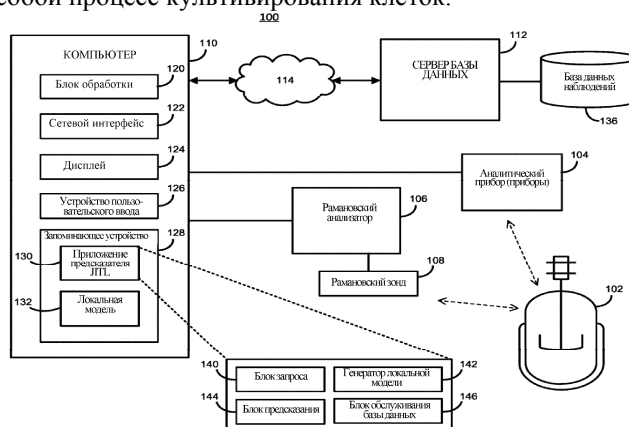
определения того, что, по меньшей мере, предсказанное аналитическое измерение не удовлетворяет одному или более критериям качества модели; и

получения фактического аналитического измерения из аналитического прибора в ответ на определение того, что, по меньшей мере, предсказанное аналитическое измерение не удовлетворяет одному или более критериям качества модели.

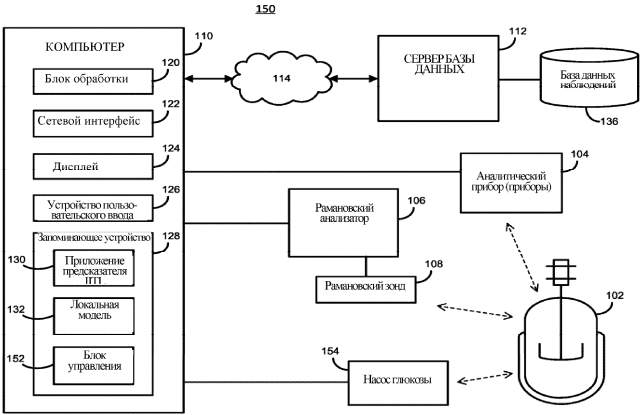
50. Биореакторная система по п.49, в которой один или более процессоров выполнены с возможностью определения того, что, по меньшей мере, предсказанное аналитическое измерение не удовлетворяет одному или более критериям качества модели, по меньшей мере, путем

генерирования интервала правдоподобия, связанного с предсказанным аналитическим измерением; и сравнения интервала правдоподобия с предварительно заданным пороговым значением.

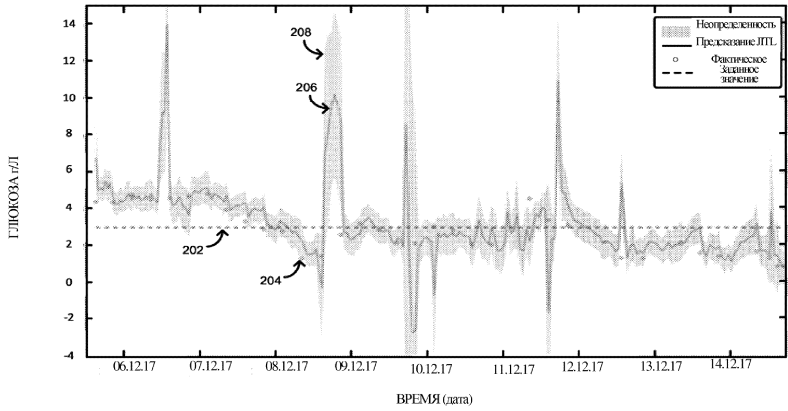
51. Биореакторная система по любому из пп.37-50, в которой процесс биофармацевтического производства представляет собой процесс культивирования клеток.



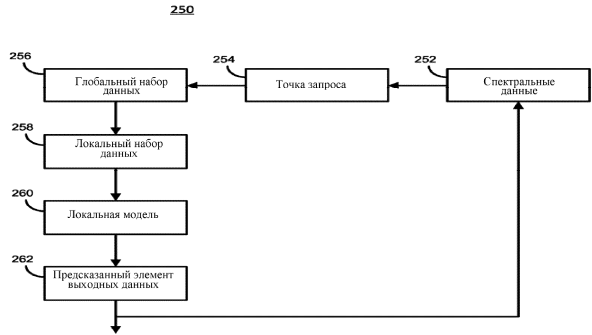
Фиг. 1



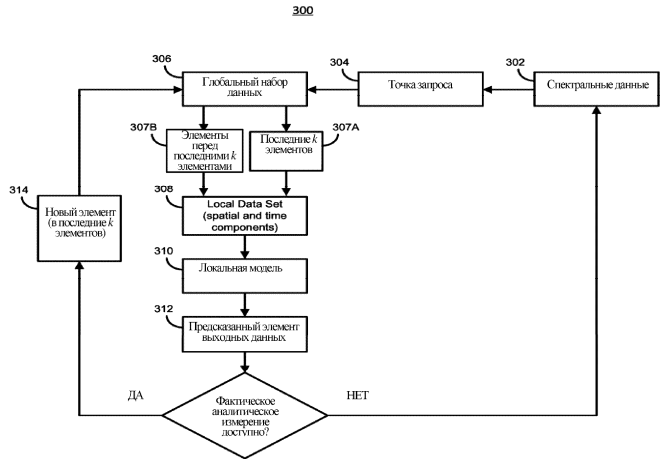
Фиг. 2



Фиг. 3

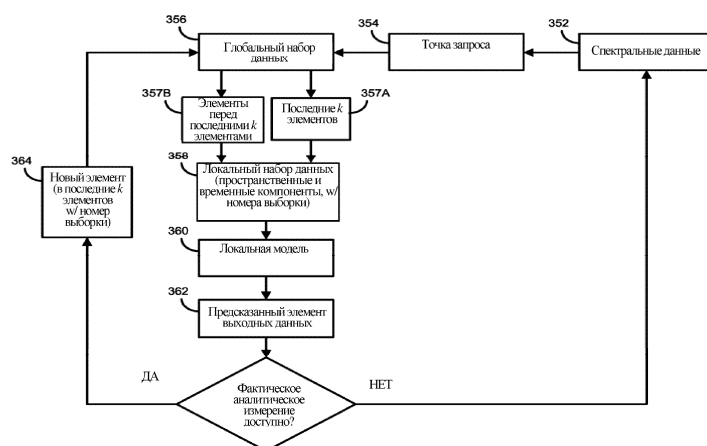


Фиг. 4



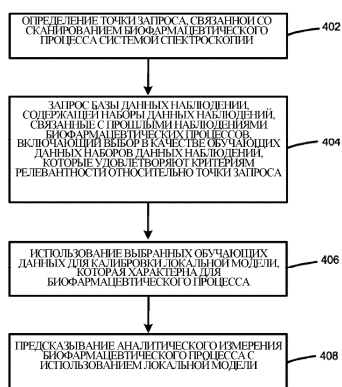
Фиг. 5

350



Фиг. 6

400



Фиг. 7



Евразийская патентная организация, ЕАПВ

Россия, 109012, Москва, Малый Черкасский пер., 2