



(21)申請案號：108138146 (22)申請日：中華民國 108 (2019) 年 10 月 23 日

(51)Int. Cl. : **G01N21/65 (2006.01)** **G01N21/84 (2006.01)**

(30)優先權：2018/10/23 美國 62/749,359  
2019/04/12 美國 62/833,044  
2019/06/21 美國 62/864,565

(71)申請人：美商安進公司 (美國) AMGEN INC. (US)  
美國

(72)發明人：托斯洋 艾迪亞 TULSYAN, ADITYA (IN)

(74)代理人：陳長文

(56)參考文獻：

|    |                |    |                |
|----|----------------|----|----------------|
| CN | 102527307A     | CN | 104541153B     |
| US | 2018/0238805A1 | US | 2018/0291329A1 |

審查人員：施孝欣

申請專利範圍項數：15 項 圖式數：7 共 78 頁

(54)名稱

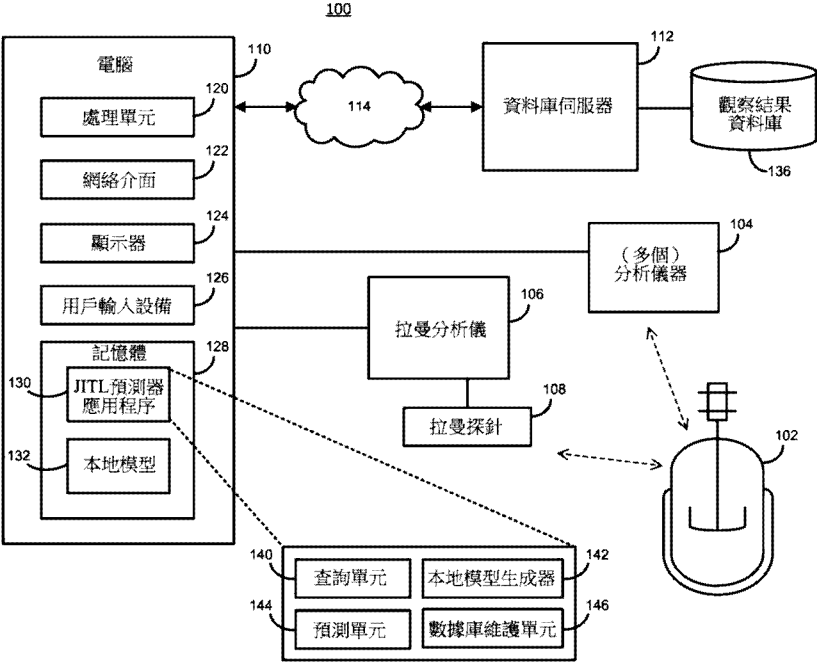
用於即時預測的拉曼光譜模型之自動校準及自動維護

(57)摘要

一種用於監測和/或控制生物製藥過程之方法包括：確定與由光譜系統（例如拉曼光譜系統）對該過程進行的掃描相關聯的查詢點，以及查詢觀察結果資料庫，該觀察結果資料庫包含與生物製藥過程的過去觀察結果相關聯的觀察結果資料集。該等觀察結果資料集中的每一個包括光譜資料和相應的實際分析測量結果。查詢該觀察結果資料庫包括從該等觀察結果資料集中選擇滿足關於該查詢點的一個或多個相關性標準的那些資料集作為訓練資料。該方法還包括使用所選訓練資料來校準特定於該生物製藥過程的本地模型。訓練該本地模型（例如高斯過程模型）以基於光譜資料登錄來預測分析測量結果。該方法還包括使用該本地模型來預測該生物製藥過程的分析測量結果。

A method for monitoring and/or controlling a biopharmaceutical process includes determining a query point associated with scanning of the process by a spectroscopy system (e.g., a Raman spectroscopy system), and querying an observation database containing observation data sets associated with past observations of biopharmaceutical processes. Each of the observation data sets includes spectral data and a corresponding actual analytical measurement. Querying the observation database includes selecting as training data, from among the observation data sets, those data sets that satisfy one or more relevancy criteria with respect to the query point. The method also includes using the selected training data to calibrate a local model specific to the biopharmaceutical process. The local model (e.g., a Gaussian process model) is trained to predict analytical measurements based on spectral data inputs. The method also includes using the local model to predict an analytical measurement of the biopharmaceutical process.

指定代表圖：



【圖1】

符號簡單說明：

- 100:系統
- 102:生物反應器
- 104:一個或多個分析儀器
- 106:拉曼分析儀
- 108:拉曼探針
- 110:電腦
- 112:資料庫伺服器
- 114:網路
- 120:處理單元
- 122:網路介面
- 124:顯示器
- 126:用戶輸入裝置
- 128:記憶體
- 130:JITL 預測器應用程式
- 132:本地模型
- 136:觀察結果資料庫
- 140:查詢單元
- 142:本地模型生成器
- 144:預測單元
- 146:資料庫維護單元

**公告本**

I844570

**【發明摘要】****【中文發明名稱】**

用於即時預測的拉曼光譜模型之自動校準及自動維護

**【英文發明名稱】**

AUTOMATIC CALIBRATION AND AUTOMATIC MAINTENANCE  
OF RAMAN SPECTROSCOPIC MODELS FOR REAL-TIME PREDICTIONS

**【中文】**

一種用於監測和/或控制生物製藥過程之方法包括：確定與由光譜系統（例如拉曼光譜系統）對該過程進行的掃描相關聯的查詢點，以及查詢觀察結果資料庫，該觀察結果資料庫包含與生物製藥過程的過去觀察結果相關聯的觀察結果資料集。該等觀察結果資料集中的每一個包括光譜資料和相應的實際分析測量結果。查詢該觀察結果資料庫包括從該等觀察結果資料集中選擇滿足關於該查詢點的一個或多個相關性標準的那些資料集作為訓練資料。該方法還包括使用所選訓練資料來校準特定於該生物製藥過程的本地模型。訓練該本地模型（例如高斯過程模型）以基於光譜資料登錄來預測分析測量結果。該方法還包括使用該本地模型來預測該生物製藥過程的分析測量結果。

**【英文】**

A method for monitoring and/or controlling a biopharmaceutical process includes determining a query point associated with scanning of the process by a spectroscopy system (e.g., a Raman spectroscopy system), and querying an observation database containing observation data sets associated with past observations of biopharmaceutical

processes. Each of the observation data sets includes spectral data and a corresponding actual analytical measurement. Querying the observation database includes selecting as training data, from among the observation data sets, those data sets that satisfy one or more relevancy criteria with respect to the query point. The method also includes using the selected training data to calibrate a local model specific to the biopharmaceutical process. The local model (e.g., a Gaussian process model) is trained to predict analytical measurements based on spectral data inputs. The method also includes using the local model to predict an analytical measurement of the biopharmaceutical process.

【指定代表圖】

圖1

【代表圖之符號簡單說明】

- 100       系統
- 102       生物反應器
- 104       一個或多個分析儀器
- 106       拉曼分析儀
- 108       拉曼探針
- 110       電腦
- 112       資料庫伺服器
- 114       網路
- 120       處理單元
- 122       網路介面

124      顯示器

126      用戶輸入裝置

128      記憶體

130      JITL預測器應用程式

132      本地模型

136      觀察結果資料庫

140      查詢單元

142      本地模型生成器

144      預測單元

146      資料庫維護單元

# 【發明說明書】

## 【中文發明名稱】

用於即時預測的拉曼光譜模型之自動校準及自動維護

## 【英文發明名稱】

AUTOMATIC CALIBRATION AND AUTOMATIC MAINTENANCE  
OF RAMAN SPECTROSCOPIC MODELS FOR REAL-TIME PREDICTIONS

## 【技術領域】

【0001】 本申請總體上係關於使用諸如拉曼（Raman）光譜法等光譜技術來監測和/或控制生物製藥過程，並且更具體地係關於對預測模型之線上校準及維護。

## 【先前技術】

【0002】 藉由生物製藥過程穩定的生產生物治療蛋白質通常需要生物反應器保持平衡且一致的參數（例如，細胞代謝濃度），這進而需要嚴格的過程監測和控制。為了滿足該等需求，越來越多地採用過程分析技術（PAT）工具。線上監測pH、溶解氧和細胞培養溫度係已經在回饋控制系統中使用的傳統PAT工具的一些示例。近年來，已經研究並部署了其他的過程中探針以連續監測更複雜的種類，諸如活細胞密度（VCD）、葡萄糖、乳酸鹽、以及其他關鍵細胞代謝物、胺基酸、滴定度和關鍵品質屬性。

【0003】 拉曼光譜法係一種流行的PAT工具，其廣泛用於生物製造中的線上監測。拉曼光譜法係一種能夠對化學成分和分子結構進行非破壞性分析之光學方法。在拉曼光譜法中，入射雷射由於分子振動模式而非彈性地散射。入射光子與散射光子之間的頻率差被稱為「拉曼位移」，並且

可以對拉曼位移與強度水平的向量（在本文中稱為「拉曼光譜」、「拉曼掃描」或「拉曼掃描向量」）進行分析以確定樣本的化學成分和分子結構。隨著雷射採樣和檢測器技術的改進，拉曼光譜法在聚合物、藥物、生物製造和生物醫學分析中的應用在過去的三十年中激增。由於該等技術的進步，拉曼光譜法現在已經成為在實驗室內外使用的實用分析技術。自從首次報導原位拉曼測量在生物製造中的應用以來，其已被用來提供對諸如葡萄糖、乳酸鹽、麩胺酸鹽、麩醯胺酸、氨、VCD等幾個關鍵過程狀態的線上即時預測。該等預測通常是基於依據來自分析儀器之分析測量結果在離線環境中構建的校準模型或軟感測器模型。通常使用偏最小平方（PLS）和多重線性回歸建模方法來將拉曼光譜與分析測量結果進行關聯。該等模型通常需要在對分析測量結果進行校準之前對拉曼掃描進行濾波預處理。一旦訓練了校準模型，就可以在即時環境中實施該模型，以提供用於過程監測和/或控制的原位測量結果。

**【0004】** 由於生物製藥過程通常在嚴格的約束和規定下操作，因此用於生物製藥應用的拉曼模型校準係很必要的。生物製藥行業中用於拉曼模型校準的當前最新方法係首先進行多次活動試驗，以生成用於將拉曼光譜與（多個）分析測量結果進行關聯的相關資料。該等試驗既昂貴又費時，因為例如每種活動在實驗室環境中都可能持續二至四周。此外，僅有限的樣本可以用於分析儀器（例如，以確保實驗室規模的生物反應器保持健康的活細胞量）。實際上，線上分析儀器或離線分析儀器中每天僅能進行一次或兩次測量並不罕見。使情況進一步惡化的是，當前的最佳實踐產生與生物反應器培養基的特定過程、特定配方或配置以及特定操作條件相關的校準模型。因此，如果要改變上述變數中的任何一個變數，則可能需

要基於新資料對模型進行重新校準。實際上，拉曼模型校準和模型維護都需要大量的資源配置，並且通常是在離線環境中執行的。雖然已經提出了使模型適應新操作條件的方法（例如，遞迴方法、移動窗口方法和時差方法），但是該等方法可能不足以處理突然的過程變化。

**【0005】** 許多出版物描述了基於傳統化學計量學方法（例如PLS建模）的用於多個分子的通用拉曼模型。然而，該等通用模型假設該等過程使用相似的（如果不是相同的話）培養基配方和/或運行過程條件。培養基和過程通常是被平台化（**platformed**）的，具有很小的差異或幾乎沒有差異。這種類型的通用模型的缺點係：一旦過程偏離規範，或者如果訓練資料集包含的過程範圍過大以試圖考慮不同分子之間的差異（例如，培養基添加劑、過程持續時間和/或其他過程變化），通用模型就失去準確性和精度。因此，該等「通用」模型僅在所描述的嚴格界限內是通用的。參見 Mehdizaheh 等人的 *Biotechnol. Prog.* 31(4): 1004-1013, 2015 和 Webster等人的*Biotechnol. Prog.* 34(3): 730-737, 2018。

**【發明內容】**

**【0006】** 術語「生物製藥過程（**biopharmaceutical process**）」係指在生物製藥製造中使用的過程，諸如用於生產期望的重組蛋白質之細胞培養過程。細胞培養在諸如生物反應器等細胞培養容器中、在支持經工程改造以表現蛋白質的生物體的生長和維持的條件下進行。在重組蛋白質生產過程中，對過程參數（諸如培養基成分濃度（包括營養素和代謝物（例如，葡萄糖、乳酸鹽、麩胺酸鹽、麩醯胺酸、氨、胺基酸， $\text{Na}^+$ 、 $\text{K}^+$ 以及其他營養素或代謝物））、培養基狀態（ $\text{pH}$ 、 $\text{pCO}_2$ 、 $\text{pO}_2$ 、溫度、滲透壓等））、以及細胞和/或蛋白質參數（例如，活細胞密度（**VCD**）、

滴定度、細胞狀態、關鍵品質屬性等）進行監測以控制和/或維持細胞培養過程。

**【0007】** 為了解決當前最佳工業實踐的上述限制中的一些限制，本文描述的實施方案係關於對用於生物製藥過程的光譜分析（諸如拉曼光譜法）的傳統技術進行改進的系統和方法。具體地，使用「即時學習（Just-In-Time Learning）」（JITL）平台來為生物製藥應用即時構建和維護校準模型（例如，拉曼校準模型）。JITL是基於本地建模和資料庫採樣技術的非線性建模平台。與其他機器學習方法不同，JITL通常假設所有可用的觀察結果都存儲在中央資料庫中，並且模型係使用來自資料庫的最相關資料基於查詢而動態地即時構建的。這允許使用相對簡單的本地模型來很好地近似複雜的過程動力學。在JITL框架下，庫可以包含不止用於在特定操作條件下操作的單個過程的光譜資料，還可以包含用於不同過程、不同培養基配置和/或不同操作條件的資料。這可以顯著減少校準和維護模型所需的時間，尤其是對於可以具有很少過去生產歷史或沒有過去生產歷史的研發中的藥品而言。

**【0008】** JITL平台維護著可以在每次新的分析測量結果可用時進行更新的動態庫。此外，為了確保本地模型適應新的過程條件，最後一個可用的分析測量結果（例如，對於當前正在監測的產品）可以始終包括在用於本地建模的訓練集中。這允許本地模型更快地適應新條件或適應無歷史的新產品線。使用這種方法，模型校準和模型維護都可以是自動的，並且可以大大減少與常規系統中的常規校準相關聯的時間和費用（例如，材料和人工成本）。此外，提供圍繞模型預測的可信度邊界（或其他置信度指標，諸如置信度評分）的能力可以允許穩健的監測和控制策略。

【0009】 在一些實施方式中，將高斯過程模型用於JITL框架內的本地建模。高斯過程模型係功能強大的統計機器學習模型，其可以高效地捕獲複雜的非線性過程動力學，並且可以輕鬆地適應幾乎任何過程變化。與PLS、主成分回歸（PCR）和其他類型的回歸模型相比，高斯（Gaussian）過程模型係非參數化的方法，並且遠更能從有限的資料集中捕獲拉曼光譜與分析測量結果之間的複雜相關性。此外，高斯過程模型通常不需要對拉曼掃描進行濾波預處理。因此，在一些實施方式中，高斯過程模型反而係在原始拉曼掃描（以對數尺度）上校準的，這可以省去模型校準/維護過程中的許多步驟。此外，高斯過程模型提供了圍繞預測的可信度邊界，該等可信度邊界可能是使用PLS模型或PCR模型極難獲得的。可信度邊界對於設計分析儀器的最佳採樣策略和/或實施閉環控制（例如模型預測控制或MPC）例如以便避免基於不可靠的預測進行改變可能特別有用。

【0010】 儘管JITL係非線性建模框架，並且儘管上述方法藉由利用最新的分析測量結果更新動態庫提供了一定的適應性，但僅JITL可能不能充分自我調整以考慮時變過程條件（例如，設定點或其他過程條件的突變）。具體地，使用JITL校準的本地模型可能無法利用最新樣本。例如，並且特別是如果存在過程條件的最新且突然的變化，則最新的樣本可能無法滿足完全基於「空間」相似性（例如，拉曼掃描的相似性）的相似性準則。本文還描述了改良的JITL技術，其可以更好地利用由最新樣本提供的資訊（與空間相似性無關），並且因此可以更好地適應時變過程變化。具體地，本文描述了用於模型校準和維護的「自我調整」JITL（A-JITL）技術和「時空」JITL（ST-JITL）技術。

【0011】 即時模型維護對JITL技術的成功可能是很重要的，在即時模型維護中，本地模型可以從最近的分析測量結果中學習、並由此快速適應時變條件。但是，頻繁訪問分析儀器/測量結果（例如，分析離線樣本）往往佔用大量資源。為了在不過度降低模型性能的情況下最小化這種資源使用，可以實施基於性能的模型維護協定，在該基於性能的模型維護協定中，系統響應於確定當前模型性能不可接受/不可靠而調度/觸發分析測量。

【圖式簡單說明】

【0012】 熟悉該項技術者將理解，本文描述之圖式係出於說明之目的而包括的，而非限制本揭露。圖式不一定是按比例繪製，而是將重點放在說明本揭露之原理上。應理解，在一些情況下，所描述之實施方式之不同方面可以被擴大或放大，以有助於理解所描述之實施方式。在圖式中，在整個不同的圖式中，相似的附圖標記通常指代功能相似和/或結構相似的部件。

【0013】 [圖1]係可以用於預測生物製藥過程的分析測量結果的示例拉曼光譜系統之簡化框圖。

【0014】 [圖2]係可以用於預測生物製藥過程的分析測量結果以用於葡萄糖濃度的閉環控制的示例拉曼光譜系統之簡化框圖。

【0015】 [圖3]描繪了使用本文所描述的拉曼光譜系統的示例實施方式對葡萄糖濃度的閉環控制的實驗結果。

【0016】 [圖4]描繪了當使用即時學習（JITL）技術來分析生物製藥過程時可能發生之示例資料流程。

【0017】 [圖5]描繪了當使用自我調整JITL（A-JITL）技術來分析

生物製藥過程時可能發生之示例資料流程。

【0018】 [圖6]描繪了當使用時空JITL（ST-JITL）技術來分析生物製藥過程時可能發生之示例資料流程。

【0019】 [圖7]係用於分析生物製藥過程的示例方法之流程圖。

【實施方式】

## 相關申請的交叉引用

【0020】 要求2018年10月23日提交的美國臨時專利申請案號62/749,359、2019年4月12日提交的美國臨時專利申請案號62/833,044和2019年6月21日提交的美國臨時專利申請案號62/864,565之優先權，該等申請中的每一者特此藉由引用以其全文併入本文。

【0021】 以上介紹的以及在下文更詳細地討論的不同構思可以以多種方式中的任一種實施，並且所描述的構思不限於任何特定的實施方式。出於說明之目的，提供了實施方式之示例。

【0022】 圖1係可以用於預測生物製藥過程的分析測量結果的示例拉曼光譜系統100之簡化框圖。雖然圖1描繪了實施拉曼光譜技術的系統100，但是應當理解，在其他實施方式中，系統100可以實施適合於分析生物製藥過程的其他光譜技術，例如諸如近紅外（NIR）光譜法。

【0023】 系統100包括生物反應器102、一個或多個分析儀器104、具有拉曼探針108的拉曼分析儀106、電腦110以及經由網路114耦合到電腦110的資料庫伺服器112。生物反應器102可以是支援生物活性環境的任何合適的容器、設備或系統，該生物活性環境可以包括培養基內的活生物體和/或從該等活生物體衍生的物質（例如，細胞培養物）。生物反應器102可以包含藉由細胞培養表現的重組蛋白質，例如，諸如用於研究目

的、臨床用途、商業銷售或其他分配。取決於所監測的生物製藥過程，培養基可以包括特定流體（例如，「液體培養基」）和特定營養素，並且可以具有目標培養基狀態參數，諸如目標pH水平或範圍、目標溫度或溫度範圍等。培養基還可以包括生物體和從該等生物體衍生的物質（諸如代謝物和重組蛋白質）。培養基的內容物和參數/特性在本文中統稱為「培養基配置（media profile）」。

【0024】（多個）分析儀器104可以是任何線上的、線上的和/或離線的一個或多個儀器，其被配置成基於從生物反應器102內的生物活性內容獲取的樣本來測量該等生物活性內容的一個或多個特性或參數。例如，（多個）分析儀器104可以測量一種或多種培養基成分濃度，諸如營養素水平和/或代謝物水平（例如，葡萄糖、乳酸鹽、麩胺酸鹽、麩醯胺酸、氨，胺基酸、Na<sup>+</sup>、K<sup>+</sup>等）以及培養基狀態參數（pH、pCO<sub>2</sub>、pO<sub>2</sub>、溫度、滲透壓等）。另外地或可替代地，（多個）分析儀器104可以測量滲透壓、活細胞密度（VCD）、滴定度、關鍵品質屬性、細胞狀態（例如，細胞週期）和/或與生物反應器102的內容物相關聯的其他特性或參數。作為更具體的示例，樣本可以被採集、旋轉沉降、通過多個柱被純化，並且可以穿過第一個分析儀器104（例如，高效液相色譜（HPLC）或超高效液相色譜（UPLC）儀器）、然後穿過第二個分析儀器104（例如，質譜儀），其中第一個和第二個分析儀器104均提供分析測量結果。（多個）分析儀器104中的一個、一些或全部可以使用破壞性分析技術。

【0025】拉曼分析儀106可以包括耦合到拉曼探針108（或者在一些實施方式中為多個拉曼探針）的光譜儀設備。拉曼分析儀106可以包括經由光纖電纜將雷射遞送到拉曼探針108的雷射源，並且還可以包括電荷耦

合器件（CCD）或其他合適的相機/記錄設備以記錄例如經由光纖電纜的另一通道從拉曼探針108接收到的信號。可替代地，雷射源可以集成在拉曼探針108本身內。拉曼探針108可以是浸入式探針或任何其他合適類型的探針（例如，反射探針和透射探針）。

【0026】 拉曼分析儀106和拉曼探針108被共同配置成在生物反應器102內的生物製藥過程期間藉由激發、觀察和記錄生物製藥過程的分子「指紋」來非破壞性地掃描生物活性內容。當生物反應器的內容物被由拉曼探針108遞送的雷射激發時，分子指紋對應於生物製藥過程內生物活性內容內的分子的振動、旋轉和/或其他低頻模式。作為這個掃描過程的結果，拉曼分析儀106生成一個或多個拉曼掃描向量，該等向量各自將強度表示為拉曼位移（頻率）的函數。

【0027】 電腦110耦合到拉曼分析儀106和（多個）分析儀器104，並且通常被配置成分析由拉曼分析儀106生成的拉曼掃描向量以便預測生物製藥過程的一個或多個分析測量結果。例如，電腦110可以分析拉曼掃描向量以便預測由（多個）分析儀器104進行的（多種）相同類型的（多個）分析測量結果。作為更具體的示例，電腦110可以預測葡萄糖濃度，而（多個）分析儀器104實際上測量葡萄糖濃度。然而，儘管（多個）分析儀器104可以對從生物反應器102提取的樣本進行相對不頻繁的「離線」分析測量（例如，由於來自生物製藥過程的培養基的數量有限和/或由於進行此類測量的較高成本等），但電腦110可以即時地對分析測量結果進行相對頻繁的「線上」預測。電腦110還可以被配置成經由網路114將由（多個）分析儀器104進行的分析測量結果傳輸至資料庫伺服器112，如將在下面進一步詳細討論的。

【0028】在圖1所示的示例實施方式中，電腦110包括處理單元120、網路介面122、顯示器124、用戶輸入裝置126和記憶體128。處理單元120包括一個或多個處理器，每個處理器可以是可程式設計微處理器，該可程式設計微處理器執行存儲在記憶體128中的軟體指令以執行如本文所描述的電腦110的一些或全部功能。可替代地，處理單元120中的一個、一些或全部處理器可以是其他類型的處理器（例如，專用積體電路（ASIC）、現場可程式設計閘陣列（FPGA）等），並且如本文所描述的電腦110的功能可以替代地部分或全部地以硬體實施。記憶體128可以包括包含易失性和/或非易失性記憶體的一個或多個物理記憶體設備或單元。可以使用任何合適的一種或多種記憶體類型，諸如唯讀記憶體（ROM）、固態驅動器（SSD）、硬碟驅動器（HDD）等。

【0029】網路介面122可以包括被配置成使用一種或多種通信協定經由網路114進行通信的任何合適的硬體（例如，前端發射器和接收器硬體）、固件和/或軟體。例如，網路介面122可以是或者包括乙太網介面。網路114可以是單個通信網路，或者可以包括一種或多種類型的多個通信網路（例如，一個或多個有線和/或無線區域網路（LAN）、和/或一個或多個有線和/或無線廣域網路（WAN）（例如，諸如網際網路或內部網路））。

【0030】顯示器124可以使用任何合適的顯示技術（例如，LED、OLED、LCD等）來向用戶呈現資訊，並且用戶輸入裝置126可以是鍵盤或其他合適的輸入裝置。在一些實施方式中，顯示器124和用戶輸入裝置126集成在單個設備（例如，觸控式螢幕顯示器）內。通常，顯示器124和用戶輸入裝置126可以組合以使得用戶能夠與由電腦110提供的圖形化

用戶介面（GUI）進行交互，例如用於諸如手動監測正在系統100內執行的各種過程的目的。但是，在一些實施方式中，電腦110不包括顯示器124和/或用戶輸入裝置126，或者顯示器124和用戶輸入裝置126中的一者或兩者都被包括在與電腦110通信地耦合的另一電腦或系統中（例如，在將預測直接發送到實施閉環控制的控制系統的一些實施方式中）。

【0031】 記憶體128存儲一個或多個軟體應用程式（包括即時學習（JITL）預測器應用程式130）的指令。JITL預測器應用程式130當被處理單元120執行時通常被配置成藉由校準本地模型132並藉由使用本地模型132分析由拉曼分析儀106生成的拉曼掃描向量來預測生物反應器102中生物製藥過程的分析測量結果。取決於拉曼分析儀106生成此類掃描向量的頻率，JITL預測器應用程式130可以基於週期性時間或其他合適的時間來預測分析測量結果。拉曼分析儀106本身可以控制何時生成掃描向量，或者電腦110可以藉由向拉曼分析儀106發送命令來觸發掃描向量的生成。JITL預測器應用程式130可以基於每個掃描向量僅預測單一類型的分析測量結果（例如，僅葡萄糖濃度），或者可以基於每個掃描向量預測多種類型的分析測量結果（例如，葡萄糖濃度和活細胞密度）。在其他實施方式中，多個不同的JITL預測器應用程式（例如，每個都類似於JITL預測器應用程式130）各自生成不同的本地模型，以全部基於同一掃描向量來預測不同類型的分析測量結果。JITL預測器應用程式130和本地模型132將在下面進一步詳細討論。

【0032】 資料庫伺服器112可以遠離電腦110（例如，使得本地設置可以僅包括生物反應器102、（多個）分析儀器104、具有拉曼探針108的拉曼分析儀106以及電腦110），並且如圖1中看出的，可以包含或通信地

耦合到存儲與過去觀察結果相關聯的觀察結果資料集的觀察結果資料庫136。觀察結果資料庫136中的每個觀察結果資料集可以包括光譜資料（例如，由拉曼分析儀106產生的此類一個或多個拉曼掃描向量）以及一個或多個相應的分析測量結果（例如，由（多個）分析儀器104產生的這樣的一個或多個測量結果）。取決於實施方式和/或場景，可能已經針對多種不同的生物製藥過程、在多種不同的操作條件（例如，不同的代謝物濃度設定點）下、和/或在多種不同的培養基配置（例如，不同的流體、營養素、pH水平、溫度等）的情況下收集了過去觀察結果。通常，可能期望使觀察結果資料庫136表示各種各樣的過程、操作條件和培養基配置。然而，取決於實施方式（如下面進一步討論的），觀察結果資料庫136可以存儲或可以不存儲指示那些過程、細胞系、蛋白質、代謝物、操作條件和/或培養基配置的資訊。在一些實施方式中，資料庫伺服器112經由網路114和/或其他網路遠端地耦合到類似於電腦110的多個其他電腦。為了收集更多的觀察結果資料集以存儲在觀察結果資料庫136中，這可能是期望的。然而，在其他實施方式中，系統100不包括資料庫伺服器112，並且電腦110直接訪問本地觀察結果資料庫136。

**【0033】** 應當理解，可以使用其他組態和/或部件來代替圖1中所示的那些。例如，不同的電腦（圖1中未示出）可以將由（多個）分析儀器104提供的測量結果傳輸到資料庫伺服器112，一個或多個附加計算設備或系統可以充當電腦110與資料庫伺服器112之間的仲介，可以替代地由資料庫伺服器112和/或另一遠端伺服器遠端地執行如本文所描述的電腦110的一些或全部功能，等等。

**【0034】** 在系統100的執行時間操作期間，使用拉曼分析儀106和拉

曼探針108來掃描生物反應器102中的生物製藥過程（即，生成該生物製藥過程的拉曼掃描向量），並且然後將該（多個）拉曼掃描向量從拉曼分析儀106傳輸到電腦110。拉曼分析儀106和拉曼探針108可以根據預定的監測週期時間表（諸如每分鐘一次或每小時一次等）來提供掃描向量以支持（由JITL預測器應用程式130進行的）預測。可替代地，可以以不規則的區間進行預測（例如，響應於基於某個過程的觸發，諸如所測量的pH水平和/或溫度的變化），使得每個監測週期具有可變的或不確定的持續時間。取決於實施方式，拉曼分析儀106可以取決於本地模型132接受多少個掃描向量作為單次預測的輸入而在每個監測週期僅向電腦110發送一個掃描向量，或者在每個監測週期向電腦110發送多個掃描向量。例如，多個掃描向量可以提高本地模型132的預測準確性。

**【0035】** JITL預測器應用程式130的查詢單元140使用在單個監測週期內接收到的（多個）掃描向量來生成查詢點，該查詢點將被用於查詢觀察結果資料庫136。在一些實施方式中，查詢點（即，定義查詢點的資料）僅包括表示從拉曼分析儀106接收到的（多個）拉曼掃描向量的資料（例如，包括每個掃描向量的強度/頻率元組）。在其他實施方式中，查詢點還包括一個或多個其他類型資訊。例如，查詢點還可以包括表示與過程相關聯的操作條件的資料（例如，控制系統中的代謝物濃度設定點，或者與拉曼分析儀106或拉曼探針108相關聯的雷射波長和/或強度等）、表示生物製藥過程培養基的培養基配置的資料（例如，流體類型、營養素類型或濃度、pH水平等）、和/或其他資料（例如，與生物製藥過程相關聯的細胞系、蛋白質或代謝物的指標）。

**【0036】** 通常，查詢點可以包括表示被本地模型132用作輸入

(即，作為本地模型132的特徵集)的相同向量、參數和/或分類的資料。將多種不同資料類型用於特徵集可以提高由本地模型132進行的分析測量結果預測的準確性。然而，由於觀察結果資料庫136中的每個觀察結果資料集通常將需要包括相同的向量、參數和/或分類作為特徵集，因此可能較佳的是將查詢點、以及本地模型132的特徵集/輸入限制為僅包括一個或多個拉曼掃描向量。這可以提供各種益處，諸如允許收集更多資訊以存儲在觀察結果資料庫136中、和/或簡化此資訊的收集。例如，如果僅使用拉曼掃描向量，則即使關於在收集觀察結果資料集時存在的過程、細胞系、蛋白質、代謝物、操作條件和/或培養基配置知之甚少或一無所知，該等觀察結果資料集也可以包括在觀察結果資料庫136中。

**【0037】** 然後，查詢單元140使用所生成的查詢點來查詢觀察結果資料庫136。在圖1的示例實施方式中，查詢單元140藉由使網路介面122經由網路114向資料庫伺服器112傳輸查詢點（例如，在查詢消息內）（這進而使資料庫伺服器112從觀察結果資料庫136中檢索適當的資料）來實現這一點。然而，在觀察結果資料庫136替代地包括在電腦110中（或包括在通信地耦合到該電腦的記憶體中）的實施方式中，查詢單元140可以替代地更直接地查詢觀察結果資料庫136。為了便於解釋，圖1的剩餘描述將假設觀察結果資料庫136耦合到資料庫伺服器112，如圖1中所描繪的。然而，熟悉該項技術者將容易地理解，在觀察結果資料庫136替代地對於電腦110而言是本地的、或位於系統架構內的另一合適位置的情況下，通信路徑可以如何不同。

**【0038】** 在接收到查詢點之後，資料庫伺服器112使用查詢點從觀察結果資料庫136中選擇將被用作本地模型132的訓練資料的相關觀察結

果資料集。取決於實施方式，資料庫伺服器112可以應用任何合適的相關性標準來識別哪些觀察結果資料集係「相關的」。在一個實施方式中，例如，查詢點包括單個拉曼掃描向量，並且資料庫伺服器112藉由計算給定的觀察結果資料集的拉曼掃描向量與該查詢點的拉曼掃描向量之間的歐氏距離來確定此觀察結果資料集是否相關。如果歐氏距離低於某個預定閾值（或者低於可變閾值，諸如基於查詢點掃描向量與所有觀察結果資料集掃描向量之間的平均歐氏距離所計算出的閾值等），則觀察結果資料集被識別為相關的觀察結果資料集。熟悉該項技術者將理解可以如何將這種方法容易地擴展到查詢點（以及每個觀察結果資料集）包括多個拉曼掃描向量的實施方式。在一些情況下，使用歐氏距離來選擇相關的觀察結果資料集可能是次優技術。然而，如果本地模型132係高斯過程模型（如下面所討論的），則使用歐氏距離作為相關性準則可能是特別有利的。這係因為具有徑向基函數或平方指數核的高斯過程模型本身基於歐氏距離。然而，在其他實施方式中，可以應用其他相關性標準（例如，基於角度的標準或基於相關性的標準等）。應當理解，在本地模型132還接受其他資訊（例如，操作條件、培養基配置、過程資料、細胞系資訊、蛋白質資訊和/或代謝物資訊等）作為輸入/特徵集的實施方式中，可以使用更複雜的技術來識別「相關的」觀察結果資料集。在一些實施方式中，資料庫伺服器112響應於單次查詢而僅選擇預定數量的相關觀察結果資料集，或者選擇不多於某個最大允許數量的相關觀察結果資料集，以確保僅檢索到觀察結果資料庫136內所有資料集的相對較小子集。然而，在其他實施方式中，只要每個這樣的資料集都滿足相關性標準，資料庫伺服器112就可以選擇任何數量的相關觀察結果資料集。

【0039】 在一些實施方式中，如下面將更詳細描述的（例如，參考圖5和圖6），不僅基於在「空間」意義上與查詢點的相關性（例如，拉曼掃描向量的相似性）、而且還基於在時間意義上的相關性（例如，哪些資料集係最新的，而與空間相似性無關）來選擇相關觀察結果資料集。該等技術可以更好地利用以下事實：即使最新的分析測量結果對應於不同的設定點等，該等最新的測量結果也可以提供有用的資訊。

【0040】 在識別出相關觀察結果資料集（其中每個觀察結果資料集可能對應於或可能不對應於與生物反應器102中的當前正在監測的生物製藥過程相同的過程條件）之後，資料庫伺服器112檢索該等資料集（例如拉曼掃描向量和（多個）相應的分析測量結果），並經由網路114將檢索到的資料集傳輸到電腦110。查詢單元140然後可以將該等相關資料集傳遞給本地模型生成器142，並且本地模型生成器142將該等相關資料集用作訓練資料以校準本地模型132。即，本地模型生成器142使用與每個觀察結果資料集相關聯的（多個）拉曼掃描向量（以及可能的其他資料）作為特徵集，並使用與同一觀察結果資料集相關聯的（多個）分析測量結果作為此特徵集的標籤。

【0041】 在一些實施方式中，如以上所指出的，本地模型生成器142構建高斯過程模型，以便高效地捕獲複雜的、非線性的過程動力學，並且容易地適應幾乎任何過程變化。與PLS模型和PCR模型不同，高斯過程模型使用非參數化的方法，並且即使在使用數量非常有限的訓練樣本的情況下，也能夠捕獲拉曼掃描向量與分析測量結果之間的複雜非線性相關性。這在新產品或新過程僅對應於觀察結果資料庫136中的有限數量的資料集的情況下可能尤其重要。在這種情況下，高斯過程模型通常能夠結合

資料庫伺服器112從觀察結果資料庫136選擇的其他相關資料集來從那些有限的資料集中提取最多的資訊。然而，在其他實施方式中，只要訓練時間不超過監測週期的最小期望持續時間，本地模型生成器142就可以替代地構建任何其他合適類型的機器學習模型（例如，遞迴神經網路、卷積神經網路等）。本地模型生成器142還可以構建本地模型132，使得本地模型132可以輸出可信度邊界、或某種其他合適的預測置信度指標（例如，置信度評分）。至少與PLS模型和PCR模型相比，高斯過程模型特別適合提供圍繞分析測量結果預測的可信度邊界。儘管已經描述了高斯過程模型相對於PLS模型和PCR模型的各種優點，但是應當理解，在一些實施方式中，本地模型生成器142可以使用PLS或PCR建模方法來構建本地模型132。

【0042】本地模型生成器142可以以線上即時的方式構建本地模型132，使得預測單元144然後可以使用經訓練的本地模型132藉由處理已被查詢單元140用來生成查詢點的（多個）相同拉曼掃描向量來預測生物製藥過程的一個或多個分析測量結果。實際上，在一些實施方式中，每一次拉曼分析儀106向電腦110提供新的拉曼掃描向量（或一組新的拉曼掃描向量）時，查詢單元140就可以執行新查詢，並且本地模型生成器142就可以生成本地模型132的新版本。然而，在其他實施方式中，查詢單元140較不頻繁地執行新查詢（並且本地模型生成器142較不頻繁地生成本地模型132的新版本），諸如每10個預測/監測週期一次、或每100個預測/監測週期一次等。

【0043】資料庫維護單元146還可以使（多個）分析儀器104以比拉曼分析儀106的監測週期顯著更低的頻率（例如，每天僅一次或兩次等）

週期性地收集一個或多個實際分析測量結果。在一些實施方式中，（多個）分析儀器104的（多個）測量結果可能是破壞性的，並且需要從生物反應器102中的過程中永久去除樣本。在資料庫維護單元146使（多個）分析儀器104收集並提供（多個）實際分析測量結果的時間或附近，資料庫維護單元146還可以使拉曼分析儀106提供一個或多個拉曼掃描向量。資料庫維護單元146然後可以使網路介面122經由網路114將（多個）拉曼掃描向量和（多個）相應的實際分析測量結果發送到資料庫伺服器112，以作為新的觀察結果資料集存儲在觀察結果資料庫136中。觀察結果資料庫132可以根據任何合適的定時來更新，該定時可以根據實施方式而變化。例如，如果（多個）分析儀器104在測量樣本的幾秒內輸出實際分析測量結果，則觀察結果資料庫132可以在採樣時幾乎立即被新的測量結果更新。然而，在某些其他實施方式中，實際分析測量結果可以是由（多個）分析儀器104中的一個或多個分析儀器進行的幾分鐘、幾小時或甚至幾天的處理的結果，在這種情況下，觀察結果資料庫132直到這種處理已經完成之後才被更新。在仍其他實施方式中，隨著分析儀器104中的不同分析儀器完成其各自的測量結果，可以以增量方式將新的觀察結果資料集添加到觀察結果資料庫132。

【0044】 因此，觀察結果資料庫136提供了可以由本地模型生成器142用來進行模型訓練的過去觀察結果的「動態庫」。在一些實施方式中，總是將（多個）最近的分析測量結果添加到觀察結果資料庫136，並且本地模型生成器142在校準本地模型132時可以總是使用觀察結果資料庫136中的（多個）最新的觀察結果資料集。這可以允許本地模型132對來自最新過去的過程資訊進行編碼，並快速適應新條件，或者快速適應沒

有歷史的新過程條件。此外，對本地模型 132 的校準和維護可以是自動的。在一些實施方式中，例如如下面結合 A-JITL 技術和 ST-JITL 技術所討論的，進一步增強了本地模型 132 的適應性。

**【0045】** 在一些實施方式中，資料庫維護單元 146 可以使（多個）分析儀器 104 基於某個其他時間或者在諸如當前模型性能等條件下收集並提供（多個）實際分析測量結果。例如，如果本地模型 132 將可信度區間（例如，在預測值附近的值範圍，在該範圍記憶體在 95% 的概率或置信度表明實際值/測量值將下降）、或某個其他置信度指標與預測一起輸出（例如，如果本地模型 132 係高斯過程模型），並且如果置信度指標顯示出特別不可靠的預測（例如，如果區間/範圍超過閾值寬度/範圍等），則資料庫維護單元 146 可以觸發對一個或多個實際分析測量結果的收集。作為更具體的示例，資料庫維護單元 146 可以響應於確定 95% 的可信度區間超過預定義閾值而觸發對（多個）分析測量結果的收集。對分析測量結果的最佳調度將在下面進一步詳細討論。在進行了（多次）測量之後，資料庫維護單元 146 可以使拉曼分析儀 106 生成一個或多個拉曼掃描向量，並且使網路介面 122 向資料庫伺服器 112 提供（多個）實際分析測量結果和（多個）相應的拉曼掃描向量，以作為新的觀察結果資料集存儲在觀察結果資料庫 132 中（例如，以上面討論的方式）。然後，在校準本地模型 132 時，本地模型生成器 142 可以在適當的情況下（例如，取決於與當前查詢的相關性，或者該實施方式是否始終利用最新的觀察結果資料集）利用這個最近的觀察結果資料集。

**【0046】** 在生物反應器中的生物製藥過程的整個生命週期中，上述一些或全部過程可以重複多次，以便使用校準和維護均是完全自動且即時

的本地模型來連續監測該過程。取決於實施方式和/或場景，可以出於各種目的來預測（多個）分析測量結果。例如，作為品質控制過程的一部分，可以監測（即，預測）某些參數以確保該過程仍然符合相關規範。作為另一示例，可以監測/預測一個或多個參數以在閉環控制系統中提供回饋。例如，圖2描繪了系統150，該系統類似於系統100，但是試圖控制生物製藥過程中的葡萄糖濃度（即，試圖使預測的葡萄糖濃度在某個可接受的容差內與期望的設定點相匹配）。應當理解，在其他實施方式中，系統150可以替代地（或者也可以）用於控制除葡萄糖水平以外的過程參數，或者基於對一個或多個其他過程參數（例如，乳酸鹽水平）的預測來控制葡萄糖水平。在圖2中，使用相同的附圖標記來指示圖1的相應部件。例如，圖2的JITL預測器應用程式130可以與圖1的JITL預測器應用程式130相同（其中，為了清楚起見，JITL預測器應用程式130的各個單元未在圖2中示出）。

**【0047】** 如圖2中看出的，在系統150內，記憶體128還存儲控制單元152。控制單元152被配置成控制葡萄糖泵154，即，使葡萄糖泵154選擇性地將附加的葡萄糖引入生物反應器102內的生物製藥過程中。控制單元152可以包括例如由處理單元120執行的軟體指令、和/或適當的固件和/或硬體。在一些實施方式中，控制單元152使用葡萄糖濃度作為閉環架構中的輸入來實施模型預測控制（MPC）技術。在本地模型132為每個預測提供可信度邊界或其他置信度指標的實施方式中（例如，在本地模型132為高斯過程模型的某些實施方式中），控制單元152還可以接受置信度指標作為輸入。例如，控制單元152可以基於葡萄糖濃度預測具有足夠高的置信度指標（例如，僅基於與不超過某個百分比或絕對測量結果範圍的可

信度邊界相關聯的預測，或者僅基於與超過某個最小閾值評分的置信度評分相關聯的預測等）而僅生成針對葡萄糖泵154的控制指令，或者可以基於該葡萄糖濃度預測的置信度指標等而增加和/或減少給定預測的權重。

【0048】圖3描繪了針對使用JITL技術來校準和維護本地高斯過程模型一個示例實施方式的實驗結果200。在圖3的繪圖中，水平虛線202表示葡萄糖濃度設定點，圓圈204表示葡萄糖濃度的實際測量結果（例如，由類似於圖1的（多個）分析儀器104之一的分析儀器進行的），實線206表示葡萄糖濃度的預測測量結果（例如，如由類似於本地模型132的模型所預測的），並且陰影區域208表示與預測測量結果相關聯的可信度邊界（95%的可信度）。如圖3中看出的，對於葡萄糖濃度設定點為3克每升（g/L）的情況，使用JITL技術進行的預測通常與分析測量結果很好地吻合。

【0049】現在將參考一個具體的JITL實施方式在數學上更詳細地描述執行查詢和構建/校準本地模型132的過程，在該實施方式中，本地模型132係使用單個拉曼掃描向量作為輸入並預測單個分析測量結果的高斯過程模型：

【0050】令 $D = \{b_j, \mathbf{a}_j\}_{j=1}^J$ （或者簡寫為 $D = \{\bar{\mathbf{b}}, \bar{\mathbf{a}}\}$ ）表示一組有序的輸入和輸出資料對，使得 $\bar{\mathbf{a}} \equiv \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_J\}$ 係輸入並且 $\bar{\mathbf{b}} \equiv \{b_1, b_2, \dots, b_J\}$ 係輸出。此外，假設 $\mathbf{a}_j \in \mathbb{R}^{n_a}$ 係 $n_a$ 維輸入向量，並且 $b_j \in \mathbb{R}$ 係標量輸出。實際上， $\mathbf{a}_j \in \mathbb{R}^{n_a}$ 可以被認為是光譜測量結果（例如NIR或拉曼），並且 $b_j \in \mathbb{R}$ 被認為是感興趣狀態（例如葡萄糖濃度或乳酸鹽濃度）的分析測量結果。給定訓練資料集 $D$ ，光譜模型校準問題的目的是識別以下形式的模型的輸入與輸出之間的關係：

$$b_j = f(\mathbf{a}_j) + \epsilon_j \tag{等式 (1)}$$

其中， $f \in \mathbb{R}$ 係光譜模型，並且 $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$ 係零均值正態分佈的測量結果雜訊，其中方差 $\sigma^2$ 未知。模型校準的標準慣例係假設 $f(\bullet)$ 係線性的，並且然後使用諸如PLS等方法來訓練模型。不是將任何限制或固定形式歸因於 $f(\bullet)$ ，而係在此假設 $f(\bullet)$ 係被建模為高斯過程的隱性函數，使得

$$f(\bar{\mathbf{a}}) \equiv \left(f(\mathbf{a}_1), f(\mathbf{a}_2), \dots, f(\mathbf{a}_J)\right) \sim \text{GP}(\boldsymbol{\mu}_\theta(\bar{\mathbf{a}}), \mathbf{k}_\theta(\bar{\mathbf{a}}, \bar{\mathbf{a}}))$$

表示來自高斯過程的隨機樣本，其中，均值 $\boldsymbol{\mu}_\theta(\bullet) \in \mathbb{R}^J$ 並且協方差函數 $\mathbf{k}_\theta(\bullet, \bullet) \in \mathbb{R}^{J \times J}$ ，其通常定義如下：

$$\boldsymbol{\mu}_\theta(\bar{\mathbf{a}}) \equiv [\mu_\theta(\mathbf{a}_1), \mu_\theta(\mathbf{a}_2), \dots, \mu_\theta(\mathbf{a}_J)]^T, \tag{等式 (2a)}$$

$$k_\theta(\bar{\mathbf{a}}; \bar{\mathbf{a}}) \equiv \begin{bmatrix} k_\theta(a_1, a_1) & k_\theta(a_1, a_2) & \cdots & k_\theta(a_1, a_J) \\ k_\theta(a_2, a_1) & k_\theta(a_2, a_2) & \cdots & k_\theta(a_2, a_J) \\ \vdots & \vdots & \ddots & \vdots \\ k_\theta(a_J, a_1) & k_\theta(a_J, a_2) & \cdots & k_\theta(a_J, a_J) \end{bmatrix}. \tag{等式 (2b)}$$

【0051】 而且， $\theta \in \mathbb{R}^{n_\theta}$ 表示高斯過程模型的超參數。高斯過程係隨機變數的集合，其中任何有限數量的隨機變數都具有聯合高斯分佈，使得對於一組有限輸入 $\bar{\mathbf{a}} \equiv \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_J\}$ ，可以寫出：

$$p(\mathbf{f}|\bar{\mathbf{a}}) = \mathcal{N}(\boldsymbol{\mu}_\theta(\bar{\mathbf{a}}), \mathbf{k}_\theta(\bar{\mathbf{a}}, \bar{\mathbf{a}})) \tag{等式 (3)}$$

【0052】 然後，光譜模型校準問題簡化為使用D來學習隱性高斯過程函數 $f \in \mathbb{R}$ 。為了數學上的方便和簡潔起見，此處假設 $\boldsymbol{\mu}_\theta = \mathbf{0}_{n_a}$ ；但是一般情況下不必如此，並且此處的結果可以容易地擴展到 $\boldsymbol{\mu}_\theta \neq \mathbf{0}_{n_a}$ 的模型。高斯過程中協方差函數的作用類似於支持向量機（SVM）中使用的核的作用。協方差函數的常見選擇係高斯核，並且由下式給出：

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

其中， $k_\theta(\mathbf{a}_i, \mathbf{a}_j) \in \mathbb{R}_+$ 係輸入對 $\{\mathbf{a}_i, \mathbf{a}_j\}$ 之間的協方差。如果集合 $\{\mathbf{a}_i, \mathbf{a}_j\}$ 中

的輸入如等式（4）中的歐氏距離所定義的彼此「接近」，則高斯核  $k_{\theta}(\mathbf{a}_i, \mathbf{a}_j)$  指派較高的相關性。

【0053】對於選擇高斯核，等式（4）係正定對稱矩陣，使得  $\mathbf{k}_{\theta}(\cdot, \cdot) \in \mathbb{S}_{++}^{J \times J}$ 。在等式（4）中，集合  $\theta \equiv \{\beta, \{\alpha_l\}_{l=1}^{n_a}\}$  係一組超參數。實際上， $\alpha_l \in \mathbb{R}_+$  係長度尺度參數，並且  $\beta \in \mathbb{R}_+$  係信號方差參數。等式（4）中高斯協方差函數的選擇對應於  $f$  係平滑且連續的先前假設。因此，藉由改變協方差函數的超參數，可以改變  $f$  的「平滑性」。在此，假設具有高斯協方差函數的高斯過程。但是，一般情況下不必如此。

【0054】給定  $\mathbf{D}$ ，目的是學習高斯過程的超參數，包括任何其他未知模型參數。對於等式（1）中的高斯過程，一組未知參數為  $\gamma \equiv \{\theta, \sigma^2\} \in \Gamma \subseteq \mathbb{R}^{n_{\gamma}}$ 。可以藉由在未知參數的空間上最大化邊緣似然（或證據）函數來執行參數學習步驟。例如，對於等式（1）中的高斯過程，邊緣似然函數給出如下：

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

其中， $p(\bar{\mathbf{b}}|\bar{\mathbf{a}})$  係邊緣似然函數， $p(\bar{\mathbf{b}}|\mathbf{f}, \bar{\mathbf{a}})$  係由下式給出的似然函數：

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

並且， $p(\bar{\mathbf{f}}|\bar{\mathbf{a}})$  係等式（3）中給出的先驗密度函數。對於分別在等式（6）和（3）中的高斯似然和先驗密度，等式（5）中的積分具有封閉解，使得邊緣似然函數由下式給出：

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

【0055】現在給定等式（7），可以藉由求解以下優化問題來估計  $\gamma \equiv \{\theta, \sigma^2\} \in \Gamma \subseteq \mathbb{R}^{n_{\gamma}}$ ：

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

其中， $\gamma^* \in \Gamma$  係最優估計。根據等式（7），我們有：

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

其中， $\mathbf{k}_{\gamma} \equiv \mathbf{k}_{\theta}(\bar{\mathbf{a}}|\bar{\mathbf{a}}) + \sigma^2\mathbf{I}_{J \times J}$ 。為了求解等式（8）中的優化問題，關於 $\gamma$ 確定等式（9）的偏導數，使得對於所有 $r = 1, 2, \dots, n_{\gamma}$ ，

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

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

其中， $\alpha = \mathbf{k}_{\gamma}^{-1}\bar{\mathbf{b}}$ 。給定等式（7）中的邊緣似然函數及其在等式（10b）中的導數，可以使用梯度下降法來求解等式（8）。由於等式（8）通常是具有多個局部最優值的非凸優化問題，因此在求解優化問題時必須謹慎。此處假設 $\gamma^*$ 係已知的或者可以藉由求解等式（8）來計算。此外，除非另有說明，否則為了減輕標記法負擔，將在此假設 $\gamma$ 係最優估計 $\gamma^*$ 。

【0056】一旦訓練了等式（1）中的高斯過程光譜校準模型，就可以將其部署用於即時預測應用。如前所述，令 $\mathbf{D}$ 為用於訓練高斯過程模型的訓練資料集，並令 $\mathbf{a}^* \in \mathbb{R}^{n_{\alpha}}$ 為新的測試光譜信號。然後，目的是預測與測試輸入 $\mathbf{a}^*$ 相對應的輸出 $\mathbf{b}^* \in \mathbb{R}$ 。計算 $\mathbf{b}^*$ 的第一步係以訓練輸入集 $\bar{\mathbf{a}}$ 和測試輸入 $\mathbf{a}^*$ 為條件來構造所有訓練輸出集 $\bar{\mathbf{b}}$ 和測試高斯過程輸出 $f(\mathbf{a}^*)$ 的聯合密度。此聯合密度給出如下：

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

其中， $\mathbf{k}_{\gamma} \equiv \mathbf{k}_{\theta}(\bar{\mathbf{a}}, \bar{\mathbf{a}}) + \sigma^2\mathbf{I}_{J \times J}$ 。給定等式（11），在貝氏框架下，藉由構造所有高斯過程輸出上的分佈來計算高斯過程輸出 $f(\mathbf{a}^*)$ 。換言之，我們尋求高斯過程輸出 $f(\mathbf{a}^*)$ 的後驗分佈。當然， $f(\mathbf{a}^*)$ 上的後驗分佈只需要包括與訓練集 $\mathbf{D}$ 一致的那些函數。在概率設置下，可以藉由使等式（11）中的聯合分佈以訓練集 $\mathbf{D}$ 為條件來計算 $f(\mathbf{a}^*)$ 上的後驗分佈，從而給出下式：

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

其中， $p(f(\mathbf{a}^*)|\mathcal{D}, \mathbf{a}^*)$  係高斯過程輸出的後驗分佈，並且  $\mu_\theta^* \equiv \mathbb{E}[f(\mathbf{a}^*) | \mathcal{D}, \mathbf{a}^*]$  由下式給出：

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

並且  $k_\theta^* \equiv \mathbb{V}[f(\mathbf{a}^*) | \mathcal{D}, \mathbf{a}^*]$  由下式給出：

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

【0057】 給定等式 (12)，輸出  $b^*$  的預測後驗分佈可以計算如下：

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

其中， $\mu_\theta^*$  和  $k_\theta^*$  分別在等式 (13) 和 (14) 中給出。對於單個測試輸入  $\mathbf{a}^* \in \mathbb{R}^{n_\alpha}$ ，等式 (15) 中的高斯過程預測給出了實現非零概率的輸出的分佈。在諸如控制和監測等即時應用中，人們可能會對點估計感興趣，而不是對整個分佈感興趣。可以使用決策理論方法來計算點估計。可以表明，對於等式 (15) 中的高斯後驗分佈，均值函數使預期的絕對風險函數和平方風險函數兩者最小化，其中， $\hat{b} = \mu_\theta^*$  係輸入  $\mathbf{a}^*$  的最可能輸出。此外，對於選擇  $\hat{b} = \mu_\theta^*$  作為預測，由下式給出大約95%的可信度區間：

$$b^L = \left( \mu_\theta^* - 2(\sqrt{k_\theta^* + \sigma^2}) \right) \leq \hat{b} \leq \mu_\theta^* + 2(\sqrt{k_\theta^* + \sigma^2}) = b^U。 \quad \text{等式 (16)}$$

等式 (16) 中的區間可以用於評估高斯過程預測的品質，和/或用於設計基於高斯過程的模型預測控制或其他穩健的監測策略。

【0058】 現在轉到響應於查詢來選擇相關樣本（此處為觀察結果資料集），問題係對於給定的查詢點  $\mathbf{a}^* \in \mathbb{R}^{n_\alpha}$ 、以及包含  $L \in \mathbb{N}$  個輸入輸出對（觀察結果資料集）的中央資料庫/庫  $\mathcal{L} \equiv \{b_i, \mathbf{a}_i\}_{i=1}^{L_t}$ ，在時間  $t \in \mathbb{N}$  選擇包含  $D \in \mathbb{N}$  個樣本的本地訓練集  $\mathcal{D} \equiv \{b_j, \mathbf{a}_j\}_{j=1}^D$ ，其中， $D \ll L$ 。假設  $\mathcal{L}$  係動態的，並且在活動期間可能包括不同的條目。存在許多方法用於從  $\mathcal{L}$  構造

$\mathcal{D}$ 。出於此分析的目的，基於集合 $\mathcal{L}$ 中的光譜（例如，拉曼掃描向量）之間的歐氏距離來選擇 $\mathcal{D}$ 。儘管據報導，在某些情況下，JITL框架中的基於歐氏的相似性度量係次優的，但當使用高斯過程模型時，它們可能是有益的選擇。這係因為高斯過程模型本身基於歐氏距離。高斯核僅在集合 $\{\mathbf{a}_i, \mathbf{a}_j\}$ 中的輸入彼此「接近」時才指派較高的相關性。因此，藉由創建其中所有輸入都「接近」查詢點的本地訓練集 $\mathcal{D}$ ，可以確保本地高斯過程模型捕獲最大的「相關性」以預測查詢點處的輸出。

【0059】 在下面的演算法1中提供了示例演算法，其正式概述了用於從 $\mathcal{L}$ 創建本地訓練集 $\mathcal{D}$ 、使用此訓練集來訓練高斯過程模型，並使用已訓練的模型進行預測的方法：

- 1.     **Input:** 庫 $\mathcal{L} = \{(\mathbf{a}_i, b_i)\}_{i=1}^L$ ，查詢點 $\mathbf{a}^*$
- 2.     **Output:** 預測 $\hat{b}$ 和不確定性 $(b^L, b^U)$
- 3.     **for**  $t = 1$ 至 $T$ ，**do**
- 4.         設置 $I \leftarrow \text{sample\_index}(\mathcal{L})$ 和 $\mathcal{D} \leftarrow \{\emptyset\}$
- 5.         **for**  $d = 1$ 至 $D$  **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.         使用 $\mathcal{D}$ 來訓練等式（1）的高斯過程模型，並估計 $\gamma^*$
- 11.         使用等式（13）和（16）來計算 $\hat{b}$ 和 $(b^L, b^U)$
- 12.     **end for**

演算法1

【0060】現在轉到圖4，示出了當使用如本文所描述的JITL技術來分析生物製藥過程時可能發生的示例資料流程250。例如，資料流程250可能在圖1的系統100、或圖2的系統150內發生。在資料流程250中，由光譜儀/探針提供光譜資料252。例如，光譜資料252可以包括由拉曼分析儀106生成的拉曼掃描向量、或者NIR掃描向量等。查詢點254係（例如，由查詢單元140）基於光譜資料252生成的，並且被用於查詢總體資料集256，該總體資料集可以包括例如觀察結果資料庫136中的所有觀察結果資料集。基於該查詢，在總體資料集256內識別區域資料集258。如上所述，例如可以基於相關性標準（例如，歐氏距離）來選擇區域資料集258。

【0061】然後將區域資料集258（例如，由本地模型生成器142）用作訓練資料以校準本地模型260（例如，本地模型132）。然後（例如，由預測單元144）使用本地模型132來預測輸出（分析測量結果）262，該輸出諸如培養基成分濃度、培養基狀態（例如，葡萄糖、乳酸鹽、麩胺酸鹽、麩醯胺酸、氨、胺基酸、 $\text{Na}^+$ 、 $\text{K}^+$ 和其他營養素或代謝物、 $\text{pH}$ 、 $\text{pCO}_2$ 、 $\text{pO}_2$ 、溫度、滲透壓等）、活細胞密度、滴定度、關鍵品質屬性、細胞狀態等，並且還可能輸出可信度邊界或另一合適的置信度指標。

【0062】雖然（例如，如演算法1和資料流程250中的）基於JITL的本地模型提供了穩健的非線性建模框架，但這種方法不具有用於適應時變過程變化的固有機制。為了解決這個缺點，一些實施方式可以使用「自我調整」JITL（A-JITL）策略。如以上所指出的，當新樣本變得可用時，該等樣本可以包括在 $\mathcal{L}$ 中。在此類實施方式中（即， $\mathcal{L}$ 係動態的實施方式中）， $\mathcal{L}$ 可以表示為 $\mathcal{L}_t$ 。在一個這樣的實施方案中，實施了移動時間窗口

方法，在該方法中，將新獲得的樣本添加到 $\mathcal{L}_t$ 中並且從 $\mathcal{L}_t$ 中移除最舊的樣本。丟棄最舊的樣本可能是有益的，因為在自我調整策略中，保持 $\mathcal{L}_t$ 的大小對於確保整個JITL框架的計算易處理性可能至關重要。但是，這種方法的一個主要問題係，簡單地丟棄舊樣本可能會導致資訊丟失，因為舊樣本可能包含相關資訊。

**【0063】** 為了避免這種資訊丟失，在一個實施方式中，將新樣本添加到 $\mathcal{L}_t$ 中而不移除任何舊的/現有的樣本。因此，在新的分析測量結果變得可用時，中央資料庫 $\mathcal{L}_t$ 隨著樣本數的增加而擴展。在細胞培養過程應用中，擴展資料庫可能不會引起任何重大的計算問題，因為這樣的過程通常作為持續兩到三周的分批時間的分批過程來操作。這自然限制了要包括在 $\mathcal{L}_t$ 中的新樣本的數量。此外，在細胞培養過程分批的過程中通常僅採樣有限數量的分析測量結果（例如，不同於頻繁採樣分析測量結果的化學工業）。因此，通常僅適度增加資料庫 $\mathcal{L}_t$ 的大小，而不會對整個JITL框架的計算穩定性造成任何重大影響。

**【0064】** 雖然將新樣本包括在 $\mathcal{L}_t$ 中對於（上面的）演算法1的不斷自我調整非常重要，但此方法的成功依賴於選擇本地資料庫 $\mathcal{D}$ 中的那些新樣本用於本地模型校準。演算法1（該演算法基於歐氏距離從 $\mathcal{L}$ 中選擇用於 $\mathcal{D}$ 的樣本（例如，演算法1的第6行））可以被稱為「空間相關」方法，因為其僅對空間中相關（接近）的樣本進行優先順序排序。如果新樣本不接近查詢樣本，這可能是發生突然設定點變化（或其他突然過程條件變化）的情況，則演算法1可能無法將該等樣本包括在 $\mathcal{D}$ 中。另一方面，遞迴方法（例如，正則化的偏最小平方（RPLS）、遞迴最小平方（RLS）和遞迴N向偏最小平方（RNPLS））係「時間相關的」，因為它們對最近的測

量結果進行優先順序排序而不考慮空間相關性。使用最近的樣本更新本地模型可以使遞迴方法成功地適應當前過程條件。

【0065】 一個這樣的實施方式（在本文中被稱為「自我調整」JITL（A-JITL））對在空間和時間上都相關的樣本進行優先順序排序。令  $\mathcal{G}^- = \{\{\mathbf{a}_i^-, \mathbf{b}_i^-\}\}_{i=1}^L$  表示從當前實驗（即發生查詢  $\mathbf{a}^*$  的實驗/過程）開始之前可用的一組  $L$  個歷史測量結果，並且令  $\mathcal{G}^+ = \{\{\mathbf{a}_j^+, \mathbf{b}_j^+\}\}_{j=1}^n$  表示從當前實驗中可用的一組  $n$  個測量結果，可以按以下方式重新分配樣本：

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

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

其中， $\mathcal{L}_t$  表示中央資料庫，並且  $\mathcal{K}$  表示一組最近的（最新的） $k$  個測量結果。在一些實施方式中， $\mathcal{K}$  包含來自當前實驗/過程的最近的  $k$  個樣本，並且  $\mathcal{L}_t$  包含來自先前實驗/過程的樣本、以及（潛在地）來自當前實驗/過程的早於該最近  $k$  個樣本的樣本。以上等式（17a）和（17b）係針對給定查詢  $\mathbf{a}^*$  定義的。對於在另一個時刻進行的查詢，資料集  $\mathcal{L}_t$  和  $\mathcal{K}$  可能包含不同的樣本，這取決於在這個時刻可用的測量結果的數量。例如，一旦樣本  $(\mathbf{a}_{n+1}^+, \mathbf{b}_{n+1}^+)$  可用，就從  $\mathcal{K}$  中移除  $(\mathbf{a}_{n-k+1}^+, \mathbf{b}_{n-k+1}^+)$  並且將  $(\mathbf{a}_{n+1}^+, \mathbf{b}_{n+1}^+)$  包括在  $\mathcal{K}$  中。然後將丟棄的樣本  $(\mathbf{a}_{n-k+1}^+, \mathbf{b}_{n-k+1}^+)$  包括在  $\mathcal{L}_t$  中以防止任何資訊丟失。用最近的測量結果來更新  $\mathcal{K}$  確保了  $\mathcal{K}$  至少反映一些當前條件。

【0066】 給定  $\mathcal{L}_t$  和  $\mathcal{K}$ ，目的是選擇  $\mathcal{D}$ 。如以上所指出的，對於 A-JITL，空間相關樣本和時間相關樣本都包括在  $\mathcal{D}$  中。如果假設  $\mathcal{D}$  可以分解為：

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

其中， $\mathcal{D}_S$  和  $\mathcal{D}_T$  分別是空間相關集合和時間相關集合，則目標係選擇  $\mathcal{D}_S$  和

$\mathcal{D}_T$ 。首先，假設  $\mathcal{D}_S \cap \mathcal{D}_T = \emptyset$ ，使得  $\mathcal{D}$  僅包含唯一的樣本。為了設計  $\mathcal{D}_S$ ，根據基於距離的（空間）度量（諸如「相似性指數」或「 $s$ 值」）從  $\mathcal{L}_t$  中選擇  $D - k$  個樣本：

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

【0067】 例如，等式 (19) 可以用作上述（非自我調整）JITL 技術中的相似性度量。因此，例如，可以從  $\mathcal{L}_t$  中選擇具有最大  $s$  值的  $D - k$  個樣本以包括在  $\mathcal{D}_S$  中。為了設計  $\mathcal{D}_T$ ，如果假設來自當前實驗/過程的最近  $k$  個樣本係時間相關的，則在一些實施方式中  $\mathcal{D}_T$  可以被定義為等於  $\mathcal{K}$ 。應當注意，與確定  $\mathcal{D}_S$  中的樣本成員資格的  $s$  值不同，基於採樣時間來確定  $\mathcal{D}_T$  中的成員資格。當然，取決於場景， $\mathcal{D}_T$  中的樣本可能會表現出較大的  $s$  值。不論  $s$  值如何， $\mathcal{D}_T$  僅被假設為係時間相關的。類似地， $\mathcal{D}_S$  僅是空間相關的，因為藉由構造， $\mathcal{L}_t$  不具有時間相關性。應當注意的是， $\mathcal{D}_S$  和  $\mathcal{D}_T$  係針對給定查詢  $\mathbf{a}^*$  定義的， $\mathcal{D}_S$  中的樣本係基於它們關於  $\mathbf{a}^*$  計算的  $s$  值來選擇的，並且  $\mathcal{D}_T$  中的樣本係基於它們關於  $\mathbf{a}^*$  的採樣時間計算的採樣時間來選擇的。為了方便， $\mathcal{D}_S$  和  $\mathcal{D}_T$  一般定義如下：

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

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

其中， $\bar{\mathbf{a}}_S$  和  $\bar{\mathbf{a}}_T$  分別是來自拉曼光譜儀的空間相關樣本和時間相關樣本，並且  $\bar{\mathbf{b}}_S$  和  $\bar{\mathbf{b}}_T$  分別是來自分析儀器的空間相關樣本和時間相關樣本，使得

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

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

【0068】 將等式 (20a) 和 (20b) 代入等式 (18) 中得到集合  $\mathcal{D}$ ，一般表示為  $\mathcal{D} \equiv \{\bar{\mathbf{a}}, \bar{\mathbf{b}}\}$ ，其中  $\bar{\mathbf{a}} \equiv [\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T]^T$  並且  $\bar{\mathbf{b}} \equiv [\bar{\mathbf{b}}_S, \bar{\mathbf{b}}_T]^T$ 。與上面討論的

（非自我調整）JITL技術相比，本地庫/資料集 $\mathcal{D}$ 對空間相關且時間相關的樣本進行優先順序排序。給定 $\mathcal{D}_T$ 和查詢 $\mathbf{a}^*$ ，可以校準等式（1）中的高斯過程模型（例如，本地模型132）。 $\mathbf{a}^*$ 處的點估計和可信度區間可以分別使用等式（13）和（16）來計算，其中 $\mathbf{k}_\gamma(\bar{\mathbf{a}}, \bar{\mathbf{a}})$ 和 $\mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}})$ 由下式給出：

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

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

其中， $\mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) \in \mathcal{S}_+^{(D-k)}$ 和 $\mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_T) \in \mathcal{S}_+^k$ 係分別與 $\mathcal{D}_S$ 和 $\mathcal{D}_T$ 相關聯的協方差函數，並且其中 $\mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) \in \mathbb{R}^{(D-k)k}$ 係 $\mathcal{D}_S$ 與 $\mathcal{D}_T$ 之間的協方差。

【0069】 在下面的演算法2中提供了正式概述A-JITL技術的示例演算法：

1.     **Input:** 庫 $\mathcal{L}_t = \{(\mathbf{a}_i, \mathbf{b}_i)\}_{i=1}^L$ ，查詢點 $\mathbf{a}^*$
2.     **Output:** 預測 $\hat{b}$ 和不確定性（ $b^L, b^U$ ）
3.     設置 $\mathcal{K} \leftarrow \{\emptyset\}$
4.     **for**  $t = 1$ 至 $T$ ，**do**
5.         設置 $I \leftarrow \text{sample\_index}(\mathcal{L}_t)$ ， $\mathcal{D}_S \leftarrow \{\emptyset\}$ ， $\mathcal{D}_T \leftarrow \{\emptyset\}$
6.         **for**  $d = 1$ 至 $D - \text{set\_cardinality}(\mathcal{K})$  **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.      使用 $\mathcal{D}$ 來訓練等式（1）中的高斯過程模型，並估計 $\gamma^*$ 
16.      使用等式（13）和（16）來計算 $\hat{b}$ 和 $(b^L, b^U)$ 
17.      if  $b^*$  可用，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
```

演算法2

【0070】 因此，演算法2將JITL（空間相關）與遞迴學習（時間相關）相結合。例如，對於 $|\mathcal{D}_T| = 0$ ，使用演算法2對本地模型132的校準類似於空間相關的JITL，而對於 $|\mathcal{D}_S| = 0$ ，使用演算法2對本地模型132的校準類似於遞迴學習。因此，藉由調整 $|\mathcal{D}_S|$ 和 $|\mathcal{D}_T|$ ，可以對（非遞迴）JITL和遞迴學習進行適當的平衡。

【0071】 現在轉到圖5，示出了當使用如本文所描述的A-JITL技術來分析生物製藥過程時可能發生的示例資料流程300。例如，資料流程300可能在圖1的系統100、或圖2的系統150內發生。在資料流程300中，由光譜儀/探針提供光譜資料302。例如，光譜資料302可以包括由拉曼分析儀106生成的拉曼掃描向量、或者NIR掃描向量等。查詢點304係（例

如，由查詢單元140）基於光譜資料302生成的，並且被用於查詢總體資料集306，該總體資料集可以包括例如觀察結果資料庫136中的所有觀察結果資料集。總體資料集306在邏輯上被分為最近的 $k$ 個條目307A（例如，全部來自當前實驗/過程）和在該最近 $k$ 個條目307A之前的所有條目307B（例如，來自先前的實驗/過程、以及還可能來自當前實驗/過程）。可以基於查詢點304的樣本數來確定 $k$ 的值。如本文所使用的，術語「樣本數」可以廣義地指代與給定樣本/觀察結果相關聯的時間或相對時間的任何指標。基於與查詢點304的空間相似性（例如，歐氏距離）將條目307B中的某些條目添加到區域資料集308中，同時可以將所有條目307A添加到區域資料集308中而不考慮空間相似性。例如，可以根據演算法2從條目307A和條目307B生成區域資料集308。

【0072】 然後將區域資料集308（例如，由本地模型生成器142）用作訓練資料以校準本地模型310（例如，本地模型132）。然後使用本地模型310（例如，由預測單元144）來預測輸出（分析測量結果）312，該輸出諸如培養基成分濃度、培養基狀態（例如，葡萄糖、乳酸鹽、麩胺酸鹽、麩醯胺酸、氨、胺基酸、 $\text{Na}^+$ 、 $\text{K}^+$ 和其他營養素或代謝物、 $\text{pH}$ 、 $\text{pCO}_2$ 、 $\text{pO}_2$ 、溫度、滲透壓等）、活細胞密度、滴定度、關鍵品質屬性、細胞狀態等，並且還可能輸出可信度邊界或另一合適的置信度指標。

【0073】 如果實際分析測量結果（例如，由諸如（多個）分析儀器104之一等分析儀器進行的測量結果）可用，則創建新條目314並將其添加到總體資料集306中。例如，此類測量結果可以是基於週期性採樣（例如，每天一次或兩次）可用的，和/或可以是響應於具有可變定時的觸發而變得可用的（例如，如果連續一定數量的預測具有不可接受的寬可信度

邊界等），如下面進一步討論的。

【0074】雖然將空間相關樣本和時間相關樣本包括在 $\mathcal{D}$ 中對於上面討論的A-JITL方法的不斷自我調整係必要的，但是由A-JITL實現的總體適應程度取決於 $\mathcal{D}$ 用於本地模型校準的有效程度。對於查詢樣本/查詢點 $\mathbf{a}^*$ ，空間相關樣本 $(\mathbf{a}_i, b_i) \in \mathcal{D}_S$ 提供了函數 $(f(\mathbf{a}^*), f(\mathbf{a}_i))$ 之間的高度相關性。這係因為，對於查詢 $\mathbf{a}^*$ ， $(\mathbf{a}_i, b_i)$ 的空間相關性以及 $(f(\mathbf{a}^*), f(\mathbf{a}_i))$ 之間的相關性都是基於 $(\mathbf{a}_i, \mathbf{a}^*)$ 之間的歐氏距離來計算的。因此，對於等式（19）中基於歐氏的相似性度量和等式（4）中基於歐氏的核的選擇， $\mathcal{D}_S$ 中的樣本預期提供較高的函數相關性。相反，時間相關樣本 $(\mathbf{a}_j, b_j) \in \mathcal{D}_T$ 可能無法提供函數 $(f(\mathbf{a}^*), f(\mathbf{a}_j))$ 之間的強烈相關性。這係因為，如以上所指出的， $\mathcal{D}_T$ 中的樣本不一定是空間相關的。結果，如果 $(\mathbf{a}_j, b_j)$ 的空間相關性很小，則 $(f(\mathbf{a}^*), f(\mathbf{a}_j))$ 之間的歸因於等式（4）中高斯核的相關性將很小。從建模的角度來看，用帶有較小相關性的樣本來訓練等式（1）中的高斯過程模型係不希望的，因為這會導致較差的模型性能。從數學上講，這可以證明如下。

【0075】對於查詢 $\mathbf{a}^*$ 以及演算法2的經校準高斯過程模型，可以使用等式（13）來計算模型預測 $\hat{b}$ 。不失一般性，如果 $\sigma^2 = 0$ （在無雜訊的情況下），則可以如下寫出等式（13）：

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

【0076】如果 $(\bar{\mathbf{a}}_T, \bar{\mathbf{b}}_T)$ 具有可忽略的空間相關性（即， $\bar{\mathbf{a}}_T$ 與 $\mathbf{a}^*$ 之間的 $s$ 值無限大），則等式4得出 $\mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_T) \approx \mathbf{0}_{1 \times k}$ 。此外，藉由構造，由於 $\bar{\mathbf{a}}_S$ 與 $\mathbf{a}^*$ 的距離近於與 $\bar{\mathbf{a}}_T$ 的距離，因此結果係 $\mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) \approx \mathbf{0}_{(D-k) \times k}$ 並且 $\mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_S) \approx \mathbf{0}_{k \times (D-k)}$ 。將它們代入等式（23）中，得出

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

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

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

從等式 (24c) 可以清楚地看出，點估計與  $\mathcal{D}_T$  無關。類似地，可以表明等式 (16) 也與  $\mathcal{D}_T$  無關。例如，等式 (16) 中的  $k_\theta^*$  可以計算如下：

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

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

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

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

從等式 (25b) 和 (25c) 可以看出，使用了幾種近似，包括  $\mathbf{k}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_T) \approx 0_{k \times 1}$ 、 $\mathbf{k}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) \approx 0_{(D-k) \times k}$  和  $\mathbf{k}_\theta(\bar{\mathbf{a}}_T, \bar{\mathbf{a}}_S) \approx 0_{k \times (D-k)}$ 。從等式 (20a) 和 (20b) 明顯可以看出，如果  $\mathcal{D}_T$  具有有限的空間相關性，則演算法2無法很好地利用該集合。

**【0077】** 在一些實施方式中，為了確保  $\mathcal{D}$  中的空間相關樣本和時間相關樣本能夠做出貢獻，使用「時空」JITL (ST-JITL) 方法，並使用以下時空拉曼模型 (例如，作為本地模型132)：

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

其中， $g: \mathbb{R}^{n_a} \times \mathbb{N} \rightarrow \mathbb{R}$  係時空拉曼模型並且  $t_i$  係  $\mathbf{a}_i$  的樣本數，並且  $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$  係具有零均值和未知方差  $\sigma^2 \in \mathbb{R}_+$  的獨立高斯隨機變數序列。

與等式 (1) 相比，等式 (26) 的時空模型取決於光譜信號及其採樣時

間。如上所述，假設 $g$ 係被建模為高斯過程的隱性函數，使得對於任何輸入 $(\mathbf{a}, t)$ ，

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

係隨機函數。為了方便起見，假設等式 (27) 中的均值函數為零，但是一般情況下不必如此。此外，對於任意輸入 $(\mathbf{a}_i, t_i)$ 和 $(\mathbf{a}_j, t_j)$ ，協方差函數 $r_\theta(\mathbf{a}_i, \mathbf{a}_j, t_i, t_j)$ 可以定義如下：

$$r_\theta(\mathbf{a}_i, \mathbf{a}_j, t_i, t_j) = k_{\text{空間}}(\mathbf{a}_i, \mathbf{a}_j) + k_{\text{時間}}(t_i, t_j), \quad \text{等式 (28)}$$

其中， $k_{\text{空間}}(\mathbf{a}_i, \mathbf{a}_j) \in \mathbb{R}_+$ 和 $k_{\text{時間}}(t_i, t_j) \in \mathbb{R}_+$ 分別是 $(g(\mathbf{a}_i, t_i), g(\mathbf{a}_j, t_j))$ 之間的空間協方差和時間協方差。應當注意，對於查詢 $(\mathbf{a}^*, t^*)$ ，如果樣本 $(\mathbf{a}_j, t_j) \in \mathcal{D}_T$ 具有可忽略的空間相關性，則 $k_{\text{空間}}(\mathbf{a}_j, \mathbf{a}^*) \approx 0$ 但 $k_{\text{時間}}(t_j, t^*) > 0$ ，使得等式 (28) 定義 $(g(\mathbf{a}^*, t^*), g(\mathbf{a}_j, t_j))$ 之間的非零相關性。最後，應當注意，等式 (28) 係有效的協方差函數，因為兩個獨立核的總和也是核。假設 $k_{\text{空間}}$ 和 $k_{\text{時間}}$ 係高斯核，使得對於任何輸入對 $(\mathbf{a}_i, t_i)$ 和 $(\mathbf{a}_j, t_j)$ ，

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

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

其中， $\theta \equiv [\alpha_1, \alpha_2, \beta_1, \beta_2]^T \in \Theta \in \mathbb{R}^4$ 係核參數。給定等式 (29a) 和 (29b)，如果 $(\mathbf{a}_i, t_i), (\mathbf{a}_j, t_j)$ 彼此接近，則等式 (28) 歸因於 $(g(\mathbf{a}_i, t_i), g(\mathbf{a}_j, t_j))$ 之間的高度相關性。如果 $\bar{\mathbf{t}}_S = [t_1, \dots, t_{D-k}]^T$ 和 $\bar{\mathbf{t}}_T = [t_{D-k+1}, \dots, t_D]^T$ 分別表示 $\mathcal{D}$ 中的狀態相關樣本和時間相關樣本的樣本數，使得 $\bar{\mathbf{t}} = [\bar{\mathbf{t}}_S; \bar{\mathbf{t}}_T]$ ，則對於查詢 $(\mathbf{a}^*, t^*)$ ，等式 (28) 中的協方差函數 $r_\theta$ 可以寫成：

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

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

【0078】應當注意，與變數 $\mathbf{a}$ 和 $b$ 不同，等式（30a）和（30b）中 $t$ 的作用僅僅是為了提高 $\mathcal{D}_T$ 的貢獻。實際上，給定 $\mathbf{a}$ ，變數 $t$ 不會對 $b$ 造成影響。因此，如果 $\bar{\mathbf{t}}_T = [t_{D-k+1}, \dots, t_D]^T$ 被定義為與 $\mathcal{D}_T$ 中的樣本相對應的樣本數，則 $\bar{\mathbf{t}}_S = [t_1, \dots, t_{D-k}]^T$ 可以被定義為使得其滿足以下等式：

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

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

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

對於所有 $i, j \in \{1, \dots, D-k\}$ 和 $k \in \{D-k+1, \dots, D\}$ ，其中， $M, N, P \in \mathbb{R}_+$ 係任意的大正常數。此外，如果假設 $\bar{\mathbf{t}}_T$ 和 $t^*$ 為使得 $k_{\text{時間}}(\bar{\mathbf{t}}_T, \bar{\mathbf{t}}_T > 0)$ 和 $k_{\text{時間}}(t^*, \bar{\mathbf{t}}_T > 0)$ ，則對於上述 $\bar{\mathbf{t}}_T$ 和 $\bar{\mathbf{t}}_S$ ， $\mathbf{r}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S, \bar{\mathbf{t}}_S, \bar{\mathbf{t}}_S)$ 可以寫成如下：

$$\mathbf{r}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S, \bar{\mathbf{t}}_S, \bar{\mathbf{t}}_S) = \mathbf{k}_{\text{空間}}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_S) + \mathbf{k}_{\text{時間}}(\bar{\mathbf{t}}_S, \bar{\mathbf{t}}_S), \quad \text{等式（32a）}$$

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

其中，等式（32b）來自等式（31a），其使 $k_{\text{時間}}(\bar{\mathbf{t}}_S, \bar{\mathbf{t}}_S)$ 中的非對角項趨於零。類似地，協方差 $\mathbf{r}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_S, t^*, \bar{\mathbf{t}}_S)$ 和 $\mathbf{r}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T, \bar{\mathbf{t}}_S, \bar{\mathbf{t}}_T)$ 可以計算如下：

$$\mathbf{r}_\theta(\mathbf{a}^*, \bar{\mathbf{a}}_S, t^*, \bar{\mathbf{t}}_S) = \mathbf{k}_{\text{空間}}(\mathbf{a}^*, \bar{\mathbf{a}}_S) + \mathbf{k}_{\text{時間}}(t^*, \bar{\mathbf{t}}_S), \quad \text{等式（33a）}$$

$$\approx \mathbf{k}_{\text{空間}}(\mathbf{a}^*, \bar{\mathbf{a}}_S), \quad \text{等式（33b）}$$

$$\mathbf{r}_\theta(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T, \bar{\mathbf{t}}_S, \bar{\mathbf{t}}_T) = \mathbf{k}_{\text{空間}}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) + \mathbf{k}_{\text{時間}}(\bar{\mathbf{t}}_S, \bar{\mathbf{t}}_T), \quad \text{等式（33c）}$$

$$\approx \mathbf{k}_{\text{空間}}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T), \quad \text{等式（33d）}$$

其中，等式（33b）基於等式（31b），並且等式（33d）基於等式（31c）。將等式（32b）、（33b）和（33d）代入等式（30a）和（30b）中得出：

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

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

【0079】從等式（30a）和（30b）可以很容易地確認協方差 $\mathbf{r}_\theta$ 包括 $\mathbf{k}_{\text{空間}}$ 和 $\mathbf{k}_{\text{時間}}$ 兩者的貢獻。給定等式（30a）和（30b）中的時空拉曼模型的協方差函數，可以藉由最大化下式來估計核參數 $\theta$ 和雜訊方差 $\sigma^2$ ：

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

其中 $\gamma = [\theta, \sigma^2]^T \in \Gamma \subset \mathbb{R}^5$ ， $\log p(\bar{\mathbf{b}}|\bar{\mathbf{a}}, \bar{\mathbf{t}})$ 係對數邊緣似然函數，並且 $\mathbf{r}_\gamma = \mathbf{r}_\theta + \mathbf{I}_{D \times D}$ 。在 $\Gamma$ 上最大化等式（35）得出最優估計 $\gamma^*$ 。對於基於梯度的優化器，可以以類似於等式（10b）的方式來計算等式（35）相對於 $\gamma$ 的梯度。給定 $\gamma^*$ ，針對查詢 $(\mathbf{a}^*, t^*)$ 的點估計和後驗方差可以計算為：

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

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

其中，協方差函數在等式（34a）和（34b）中給出。類似地，等式（36a）中的點估計上的可信度邊界（ $b^L \leq \hat{b} \leq b^U$ ）可以計算如下：

$$b^L = \hat{b} - 2\sqrt{r_\gamma^*}, \quad \text{等式（37a）}$$

$$b^U = \hat{b} + 2\sqrt{r_\gamma^*}, \quad \text{等式（37b）}$$

其中， $r_\gamma^* = r_\theta^* + \sigma^2$ 。從等式（36a）、（37a）和（37b）可以很容易地看出，空間相關樣本和時間相關樣本都對模型預測和可信度邊界計算有貢獻。最後，將等式（34a）和（34b）代入等式（36a）和（36b）中分別得出後驗均值和方差。應當注意，與演算法2的情況不同，即使在 $\mathcal{D}_T$ 不具有空間相關性的情況下，等式（36a）中的模型預測以及等式（37a）和（37b）中的可信度區間也取決於 $\mathcal{D}_T$ 。例如，當 $\mathcal{D}_T$ 不具有空間相關性（即， $\mathbf{k}_{\text{空間}}(\bar{\mathbf{a}}_S, \bar{\mathbf{a}}_T) \approx \mathbf{0}_{(D-k) \times k}$ 和 $\mathbf{k}_{\text{空間}}(\mathbf{a}^*, \bar{\mathbf{a}}_T) \approx \mathbf{0}_{1 \times k}$ ）時，則等式（36a）和（36b）可以寫為：

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

等式 (38a)

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

等式 (38b)

【0080】從上面可以看出，等式 (38a) 和 (38b) 仍然包括來自  $\mathbf{k}_{\text{空間}}$  和  $\mathbf{k}_{\text{時間}}$  兩者的貢獻。在下面的演算法3中提供了正式概述ST-JITL技術的示例演算法：

1.     **Input:** 庫  $\mathcal{L}_t = \{(\mathbf{a}_i, \mathbf{b}_i)\}_{i=1}^L$ ，查詢點  $\mathbf{a}^*$
2.     **Output:** 預測  $\hat{b}$  和不確定性  $(b^L, b^U)$
3.     設置  $\mathcal{K} \leftarrow \{\emptyset\}$  和  $\bar{\mathbf{t}}_T \leftarrow \{\emptyset\}$
4.     **for**  $t = 1$  至  $T$ ，**do**
5.         設置  $I \leftarrow \text{sample\_index}(\mathcal{L}_t)$ ， $\mathcal{D}_S \leftarrow \{\emptyset\}$ ， $\mathcal{D}_T \leftarrow \{\emptyset\}$
6.         **for**  $d = 1$  至  $D - \text{set\_cardinality}(\mathcal{K})$  **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.         根據等式 (31a) 至等式 (31c) 設置  $\bar{\mathbf{t}}_S$
16.         設置  $\bar{\mathbf{t}} \leftarrow [\bar{\mathbf{t}}_S; \bar{\mathbf{t}}_T]$
17.         使用  $\mathcal{D}$  和  $\bar{\mathbf{t}}$  來訓練等式 (28) 中的高斯過程模型，並估計  $\gamma^*$

```
18.      使用等式 ( 36a ) 來計算  $\hat{b}$  , 並且使用等式 ( 37a ) 和
( 37b ) 來計算  $(b^L, b^U)$ 

19.      if  $b^*$  可用 , 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.               $\mathcal{K} \leftarrow \mathcal{K} \cup \{a^*, b^*\}$ 

24.          end if

25.           $\mathcal{K} \leftarrow \mathcal{K} \cup \{a^*, b^*\}$ 

26.      end if

27.  end for
```

**演算法3**

【0081】 應當注意，對於 $\beta_1 = 0$ 的情況，A-JITL和ST-JITL（分別在演算法2和演算法3中）可以是相同的。這係因為，對於 $\beta_1 = 0$ ， $k_{時間} = 0$ ，使得 $r_{\theta} = k_{空間} = k_{\theta}$ （如從等式（28）和（29b）看出）。

【0082】 現在轉到圖6，示出了當使用如本文所描述的ST-JITL技術來分析生物製藥過程時可能發生的示例資料流程350。例如，資料流程350可能在圖1的系統100、或圖2的系統150內發生。在資料流程350中，由光譜儀/探針提供光譜資料352。例如，光譜資料352可以包括由拉曼分析儀106生成的拉曼掃描向量、或者NIR掃描向量等。查詢點354係（例如，由查詢單元140）基於光譜資料352生成的，並且被用於查詢總體資料集356，該總體資料集可以包括例如觀察結果資料庫136中的所有觀察結果資料集。總體資料集356在邏輯上被分為最近的 $k$ 個條目357A（例

如，全部來自當前實驗/過程）和在該最近 $k$ 個條目357A之前的所有條目357B（例如，來自先前的實驗/過程、以及還可能來自當前實驗/過程）。可以基於查詢點354的樣本數來確定 $k$ 的值。例如，可以根據演算法3從條目357A和條目357B生成區域資料集358。

【0083】 然後將區域資料集358（例如，由本地模型生成器142）用作訓練資料以校準本地模型360（例如，本地模型132）。然後（例如，由預測單元144）使用本地模型360來預測輸出（分析測量結果）362，該輸出諸如培養基成分濃度、培養基狀態（例如，葡萄糖、乳酸鹽、麩胺酸鹽、麩醯胺酸、氨、胺基酸、 $\text{Na}^+$ 、 $\text{K}^+$ 和其他營養素或代謝物、 $\text{pH}$ 、 $\text{pCO}_2$ 、 $\text{pO}_2$ 、溫度、滲透壓等）、活細胞密度、滴定度、關鍵品質屬性、細胞狀態等，並且還可能輸出可信度邊界或另一合適的置信度指標。

【0084】 如果實際分析測量結果（例如，由諸如（多個）分析儀器104之一等分析儀器進行的測量結果）可用，則創建新條目364（包括其樣本數）並將其添加到總體資料集356中。例如，此類測量結果可以是基於週期性採樣（例如，每天一次或兩次）可用的，和/或可以是響應於具有可變定時的觸發而變得可用的（例如，如果連續一定數量的預測具有不可接受的寬可信度邊界等）。

【0085】 如以上所指出的，可以基於一個或多個本地模型（例如，本地模型132、260、310或360）的當前和/或最新的性能來調度/觸發分析測量，以便維持或提高預測準確性，同時減少資源使用（例如分析儀器的使用）。例如，這種技術可以與A-JITL、ST-JITL或單純的JITL一起使用。

【0086】 在一個實施方式中，可信度區間被用來觸發模型維護。具

體地，如果圍繞給定模型預測（例如，圍繞由本地模型132、260、310或360進行的最新預測）的可信度區間的寬度（例如，如使用等式（16）或等式（37a）、（37b）計算的可信度邊界之間的距離）大於預定義閾值，則資料庫維護單元146可以生成請求消息，並使電腦110將該消息發送到（多個）分析儀器104以請求測量結果。在圖3的示例結果中，例如，資料庫維護單元146可能會在日期12/08/17、12/09/17和12/14/17結束附近觸發新的分析測量結果，其中陰影區域208指示較寬可信度區間（即， $b^U - b^L$ 中的較大值）。

【0087】響應於該請求消息，（多個）分析測量104執行（多次）測量，並將（多個）測量結果提供給電腦110。然後，資料庫維護單元146可以將該（多個）測量結果以及從拉曼分析儀106接收到的（多個）相應拉曼掃描向量發送到資料庫伺服器112，以存儲在觀察結果資料庫136中。例如，可以將該（多個）測量結果和（多個）掃描向量添加到上面討論的庫 $\mathcal{L}$ （對於單純的JITL）或庫 $\mathcal{K}$ （對於A-JITL或ST-JITL）中。

【0088】相反，如果圍繞給定模型預測的可信度區間的寬度不大於預定義閾值，則資料庫維護單元146可以不請求新的分析測量結果，在這種情況下，觀察結果資料庫136中的庫保持不變。在（多個）分析儀器104包括測量不同的特性（諸如培養基成分濃度、培養基狀態（例如，葡萄糖、乳酸鹽、麩胺酸鹽、麩醯胺酸、氨、胺基酸、 $\text{Na}^+$ 、 $\text{K}^+$ 和其他營養素或代謝物、pH、 $\text{pCO}_2$ 、 $\text{pO}_2$ 、溫度、滲透壓等）、活細胞密度、滴定度、關鍵品質屬性、細胞狀態等）的多個儀器、並且單獨的本地模型被用來預測各種不同特性值的實施方案中，可以針對每個預測的屬性以及測量此屬性的分析儀器分別實施調度過程，其中可能每個屬性具有不同的可

信度區間寬度閾值。

【0089】 在數學上，資料庫維護單元146可以在以下條件下在查詢點 $a^*$ 處調度/觸發（多次）新的分析測量：

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

其中， $THR$ 係用戶定義的閾值。在一些實施方式中， $THR$ 可以由用戶調整以適合特定的應用或用例。例如，用戶可以為模型可靠性至關重要的應用設置相對較小的 $THR$ 值（由資料庫維護單元146使用），從而使得更頻繁地發生模型/庫維護操作。通常，可以基於過程關鍵性、基於正在預測的參數（諸如培養基成分濃度、培養基狀態（例如，葡萄糖、乳酸鹽、麩胺酸鹽、麩醯胺酸、氨、胺基酸、 $Na^+$ 、 $K^+$ 和其他營養素或代謝物、 $pH$ 、 $pCO_2$ 、 $pO_2$ 、溫度、滲透壓等）、活細胞密度、滴定度、關鍵品質屬性、細胞狀態等）、和/或基於當前時間段（例如，在培養後期使用與初期相比較低的 $THR$ ）將 $THR$ 設置為不同的值。 $THR$ 的選擇表示模型準確性與資源（分析儀器）使用之間的折衷，其中較低的閾值往往以增加資源使用為代價來提高模型準確性。

【0090】 這個調度協議的變型也是可能的。例如，在一個實施方式中，資料庫維護單元146可以將一個或多個模型性能標準不僅應用於當前（最新）預測，而且還應用於一個或多個其他最新預測（例如，最新的 $N$ 個預測，其中， $N > 1$ ）。作為這樣的實施方式的示例，資料庫維護單元146可以為最新的 $N$ 個預測（ $N \geq 1$ ）計算可信度區間的平均寬度，並且然後將此平均寬度與閾值 $THR$ 進行比較。作為另一示例，資料庫維護單元146可以在最近的 $Y$ 個預測中識別出 $X$ 個最大的可信度區間寬度（ $X < Y$ ），並且僅當這 $X$ 個寬度中的每一個都大於閾值 $THR$ 時才調度/觸發新的

分析測量。

【0091】 圖7係用於分析生物製藥過程（例如，用於監測和/或控制目的）的示例方法400的流程圖。例如，方法400可以由諸如圖1或圖2的電腦110等電腦（例如，藉由處理單元120執行JITL預測器應用程式130的指令）來設施，和/或由諸如圖1或圖2的資料庫伺服器112等伺服器來實施。

【0092】 在框402處，確定與由光譜系統（例如，由系統100或系統150的拉曼分析儀104和拉曼探針106）對生物製藥過程進行的掃描相關聯的查詢點。例如，可以至少部分地基於由光譜系統在掃描生物製藥過程時所生成的光譜掃描向量（例如，拉曼掃描向量或NIR掃描向量）來確定查詢點。取決於實施方式，可以基於原始光譜掃描向量或在對原始光譜掃描向量進行合適的濾波預處理之後確定查詢點。在一些實施方案中，例如，還基於其他資訊來確定查詢點，該等其他資訊諸如與生物製藥過程相關聯的培養基配置（例如，流體類型、特定營養素、pH水平等）和/或分析生物製藥過程時的一個或多個操作條件（例如，代謝物濃度設定點等）。

【0093】 在框404處，查詢觀察結果資料庫（例如，觀察結果資料庫136）。觀察結果資料庫可以包含與許多生物製藥過程的過去觀察結果相關聯的觀察結果資料集。每個觀察結果資料集可以包括光譜資料（例如，拉曼掃描向量或NIR掃描向量）和相應的分析測量結果（或者在一些實施方式中，兩個或更多個分析測量結果）。例如，分析測量結果可以是培養基成分濃度、培養基狀態（例如，葡萄糖、乳酸鹽、麩胺酸鹽、麩醯胺酸、氨、胺基酸、Na<sup>+</sup>、K<sup>+</sup>和其他營養素或代謝物、pH、pCO<sub>2</sub>、pO<sub>2</sub>、溫度、滲透壓等）、活細胞密度、滴定度、關鍵品質屬性和/或細胞

狀態。

【0094】 框404可以包括從該等觀察結果資料集中選擇滿足關於該查詢點的一個或多個相關性標準的那些觀察結果資料集作為訓練資料。例如，如果查詢點包括光譜掃描向量，則框404可以包括將該光譜掃描向量與同表示在觀察結果資料庫中的每個過去觀察結果相關聯的光譜掃描向量進行比較（例如，藉由計算（1）確定該查詢點所基於的光譜掃描向量與（2）同過去觀察結果相關聯的每個光譜掃描向量之間的歐氏距離或其他距離，並且然後選擇與該等過去觀察結果相關聯的光譜掃描向量中的、被確定為在確定該查詢點所基於的光譜掃描向量的閾值距離之內的所有光譜掃描向量作為訓練資料）。

【0095】 在框406處，使用所選訓練資料來校準特定於正在監測的生物製藥過程的本地模型。在框406處，訓練本地模型（例如，本地模型132）以基於光譜資料登錄（例如，拉曼光譜掃描向量或NIR光譜掃描向量）來預測分析測量結果。在一些實施方式中，該本地模型係高斯過程機器學習模型。

【0096】 在框408處，使用本地模型來預測生物製藥過程的分析測量結果。框408可以包括使用本地模型來分析光譜系統在掃描生物製藥過程時所生成的光譜資料（例如，拉曼掃描向量或NIR掃描向量）。例如，框408可以包括藉由使用本地模型來處理查詢點所基於的同一掃描向量或其他光譜資料來預測分析測量結果。取決於實施方式，本地模型可以用於分析原始光譜資料（例如，原始拉曼掃描向量），或者用於在對原始光譜資料進行合適的濾波預處理之後分析光譜資料。在一些實施方式中，框408還包括確定與生物製藥過程的所預測分析測量結果相關聯的置信度指

標（例如，可信度邊界、置信度評分等）。在一些實施方式中，在框408處，本地模型還預測一個或多個附加分析測量結果。

**【0097】** 在一些實施方式中，方法400包括圖5中未示出的一個或多個附加框。例如，方法400可以包括附加框，在該附加框中，至少部分地基於在框408處預測的分析測量結果來控制生物製藥過程的至少一個參數。取決於實施方式，該參數可以是與所預測的分析測量結果相同的類型（例如，基於所預測的葡萄糖濃度來控制葡萄糖濃度），或者可以是不同的類型。例如，可以使用模型預測控制（MPC）技術來控制（一個或多個）參數。

**【0098】** 作為另一示例，方法400可以包括第一附加框和第二附加框，在該第一附加框中，獲得了生物製藥過程的實際分析測量結果（例如，如上面所討論的，響應於確定預測的分析測量結果、以及可能還有一個或多個較早/最新的測量結果不滿足一個或多個模型性能標準而由或從（多個）分析儀器104之一獲得）；在該第二附加框中，將（1）光譜系統在獲得實際分析測量結果時所生成的光譜資料以及（2）生物製藥過程的實際分析測量結果添加到觀察結果資料庫中（例如，藉由將光譜資料和分析測量結果發送到資料庫伺服器（諸如資料庫伺服器112），或者藉由直接將光譜資料和分析測量結果添加到本地觀察結果資料庫等）。在預測多種類型的分析測量結果的實施方式中，可以獲得多個實際分析測量結果並將其添加到觀察結果資料庫。

**【0099】** 作為又一個示例，方法400可以包括一個或多個附加的框組，每個框類似於框402至408。在該等附加的框組的每個框組中，可以藉由查詢觀察結果資料庫（或另一個觀察結果資料庫）來校準本地模型，

並將該本地模型用於預測不同類型的分析測量結果。

**【0100】** 現在將解決與本揭露有關的其他考慮。

**【0101】** 術語「多肽 (polypeptide)」或「蛋白質 (protein)」在全文中可互換使用，並且係指包括藉由肽鍵彼此連結的兩個或更多個胺基酸殘基的分子。多肽和蛋白質還包括具有天然序列的胺基酸殘基的一個或多個缺失、插入和/或取代的大分子，即包括由天然存在的非重組細胞產生的多肽或蛋白質；或藉由基因工程細胞或重組細胞產生，並且包括具有天然蛋白質的胺基酸序列的胺基酸殘基的一個或多個缺失、插入和/或取代的分子。多肽和蛋白質還包括如下胺基酸聚合物，其中一種或多種胺基酸為相應天然存在的胺基酸和聚合物的化學類似物。多肽和蛋白質還包括修飾，所述修飾包括但不限於糖基化、脂質附著、硫酸化、麩胺酸殘基的 $\gamma$ -羧化、羥基化和ADP核糖基化。

**【0102】** 多肽和蛋白質可能具有科學意義或商業意義，包括基於蛋白質的治療法。蛋白質尤其包括分泌型蛋白質、非分泌型蛋白質、胞內蛋白質或膜結合蛋白質。多肽和蛋白質可以使用細胞培養方法藉由重組動物細胞系產生，並且可以被稱為「重組蛋白質」。所表現的（多種）蛋白質可以在細胞內產生或被分泌到培養基中，從培養基中可以回收和/或收集所述蛋白質。蛋白質包括藉由結合靶、特別是下面列出的那些中的靶而發揮治療作用的蛋白質，包括從其衍生的靶、與其相關的靶及其修飾。

**【0103】** 蛋白質「抗原結合蛋白」。「抗原結合蛋白」係指包括抗原結合區或抗原結合部分的蛋白質或多肽，該抗原結合區或抗原結合部分對與其結合的另一分子（抗原）具有強親和力。抗原結合蛋白涵蓋抗體、肽體、抗體片段、抗體衍生物、抗體類似物、融合蛋白（包括單鏈可變片

段（scFv）和雙鏈（雙價）scFv、突變蛋白、xMAb和嵌合抗原受體（CAR））。

【0104】 scFv係單鏈抗體片段，它具有連接在一起的抗體重鏈和輕鏈的可變區。參見美國專利案號7,741,465、和6,319,494以及Eshhar等人, Cancer Immunol Immunotherapy [癌症免疫學免疫療法] (1997) 45: 131-136。scFv保留了親本抗體與靶抗原特異性相互作用的能力。

【0105】 術語「抗體」包括任何同種型或亞類的糖基化免疫球蛋白和非糖基化免疫球蛋白，或者其與完整抗體競爭特異性結合的抗原結合區。除非另有說明，否則抗體包括人類的、人源化的、嵌合的、多特異性的、單株的、多株的、heteroIgG、XmAb、雙特異性的抗體、及其寡聚物或抗原結合片段。抗體包括IgG1型、IgG2型、IgG3型或IgG4型。還包括具有抗原結合片段或抗原結合區的蛋白質，諸如Fab、Fab'、F(ab')<sub>2</sub>、Fv、雙抗體、Fd、dAb、最大抗體（maxibody）、單鏈抗體分子、單結構域VHH、互補決定區（CDR）片段、scFv、雙抗體、三抗體、四抗體和至少包含足以使特異性抗原與靶多肽結合的免疫球蛋白的一部分的多肽。

【0106】 還包括人類的、人源化的和其他抗原結合蛋白，諸如人類抗體和人源化抗體，該等抗原結合蛋白當施用于人類時不會產生明顯有害的免疫反應。

【0107】 還包括肽體，該等肽體是包括視需要地經由接頭與Fc結構域連結在一起的一個或多個生物活性肽的多肽。參見美國專利案號6,660,843、美國專利案號7,138,370和美國專利案號7,511,012。

【0108】 蛋白質還包括基因工程受體，諸如嵌合抗原受體（CAR或

CAR-T) 和 T 細胞受體 (TCR)。CAR 通常將抗原結合結構域 (諸如 scFv) 與一個或多個共刺激 (「信號傳導」) 結構域和一個或多個激活結構域串聯在一起。

【0109】還包括雙特異性 T 細胞接合物 (BiTE®) 抗體構建體，其係由兩個柔性連接的抗體衍生的結合結構域製成的重組蛋白構建體 (參見 WO 99/54440 和 WO 2005/040220)。該構建體的一個結合結構域對靶細胞上的所選腫瘤相關表面抗原具有特異性；第二結合結構域對 CD3 (T 細胞上的 T 細胞受體複合物的亞基) 具有特異性。BiTE® 構建體還可以包括在 CD3s 鏈的 N 末端處結合背景無關表位 (context independent epitope) 的能力 (WO 2008/119567)，以更特異性地激活 T 細胞。半衰期延長 BiTE® 構建體包括小雙特異性抗體構建體與較大蛋白質的融合物，其較佳地不會干擾 BiTE® 抗體構建體的治療效果。雙特異性 T 細胞接合物的這種進一步發展的示例包括雙特異性 Fc- 分子，例如 US 2014/0302037、US 2014/0308285、WO 2014/151910 和 WO 2015/048272 中所述。替代性策略係使用與雙特異性分子融合的人血清白蛋白 (HAS) 或者僅人白蛋白結合肽的融合物 (參見例如 WO 2013/128027、WO2014/140358)。另一種 HLE BiTE® 策略包括融合與靶細胞表面抗原結合的第一結構域、與人和/或獼猴 CD3e 鏈的胞外表位結合的第二結構域以及作為特異性 Fc 模式的第三結構域 (WO 2017/134140)。

【0110】還包括經修飾的蛋白質，諸如經非共價鍵、共價鍵或者共價鍵和非共價鍵兩者化學修飾的蛋白質。還包括進一步包含一種或多種譯後修飾的蛋白質，其可以藉由細胞修飾系統或由酶和/或化學方法離體引入或以其他方式引入的修飾製得。

【0111】蛋白質還可以包括重組融合蛋白，該重組融合蛋白包括例如多聚化結構域，諸如亮胺酸拉鍊、捲曲螺旋、免疫球蛋白的Fc部分等。還包括包含分化抗原的全部或部分胺基酸序列的蛋白質（稱為CD蛋白質）或其配位基或與該等中的任一個實質上相似的蛋白質。

【0112】在一些實施方式中，蛋白質可以包括群落刺激因子，諸如顆粒性白血球群落刺激因子（G-CSF）。此類G-CSF試劑包括但不限於Neupogen®（非格司亭（filgrastim））和Neulasta®（培非格司亭（pegfilgrastim））。還包括紅血球生成刺激劑（ESA），諸如Epogen®（依伯汀（epoetin） $\alpha$ ），Aranesp®（達貝泊汀（darbepoetin） $\alpha$ ），Dynepo®（依伯汀 $\delta$ ），Mircer®（甲氧基聚乙二醇-依伯汀 $\beta$ ），Hematide®，MRK-2578，INS-22，Retacrit®（依伯汀 $\zeta$ ），Neorecormon®（依伯汀 $\beta$ ），Silapo®（依伯汀 $\zeta$ ），Binocrit®（依伯汀 $\alpha$ ），epoetin alfa Hexal，Abseamed®（依伯汀 $\alpha$ ），Ratioepo®（依伯汀 $\theta$ ），Eporatio®（依伯汀 $\theta$ ），Biopoin®（依伯汀 $\theta$ ），依伯汀 $\alpha$ ，依伯汀 $\beta$ ，依伯汀 $\zeta$ ，依伯汀 $\theta$ 和依伯汀 $\delta$ ，依伯汀 $\omega$ ，依伯汀 $\iota$ ，組織纖溶酶原活化劑，GLP-1受體激動劑，以及前述任何內容的分子或其變體或類似物和生物仿製藥。

【0113】在一些實施方式中，蛋白質可以包括與一種或多種CD蛋白質、HER受體家族蛋白質、細胞黏著分子、生長因子、神經生長因子、成纖維細胞生長因子、轉變生長因子（TGF）、胰島素樣生長因子、骨誘導因子、胰島素和胰島素相關蛋白、凝血和凝血相關蛋白、群落刺激因子（CSF）、其他血液和血清蛋白血型抗原特異性結合的蛋白質；受體、受體相關蛋白、生長激素、生長激素受體、T細胞受體；神經滋養因子、神

經滋養蛋白、鬆弛素（relaxin）、干擾素、介白素、病毒抗原、脂蛋白、整合素、類風濕因子、免疫毒素、表面膜蛋白、運輸蛋白、歸巢受體、位址素、調節蛋白和免疫黏附素。

【0114】 在一些實施方式中，蛋白質可以包括單獨或以任何組合結合以下一種或多種蛋白質的蛋白質：CD蛋白質（包括但不限於CD3、CD4、CD5、CD7、CD8、CD19、CD20、CD22、CD25、CD30、CD33、CD34、CD38、CD40、CD70、CD123、CD133、CD138、CD171和CD174）、HER受體家族蛋白質（包括例如HER2、HER3、HER4和EGF受體）、EGFRvIII、細胞黏著分子（例如LFA-1、Mol、p150,95、VLA-4、ICAM-1、VCAM和 $\alpha$  v/ $\beta$  3整合素）、生長因子（包括但不限於例如血管內皮生長因子（「VEGF」））；VEGFR2、生長激素、甲狀腺刺激素、卵泡刺激素、黃體生成激素、生長激素釋放因子、甲狀旁腺激素、米勒管抑制物質（mullerian-inhibiting substance）、人類巨噬細胞炎性蛋白（MIP-1- $\alpha$ ）、紅血球生成素（EPO）、神經生長因子（諸如NGF- $\beta$ ）、血小板源性生長因子（PDGF）、成纖維細胞生長因子（包括例如aFGF和bFGF）、表皮生長因子（EGF）、Cripto、轉變生長因子（TGF）（其中包括TGF- $\alpha$ 和TGF- $\beta$ （包括TGF- $\beta$ 1、TGF- $\beta$ 2、TGF- $\beta$ 3、TGF- $\beta$ 4或TGF- $\beta$ 5））、胰島素樣生長因子-I和胰島素樣生長因子-II（IGF-I和IGF-II）、des(1-3)-IGF-I（腦IGF-I）和骨誘導因子、胰島素和胰島素相關蛋白（包括但不限於胰島素、胰島素A鏈、胰島素B鏈、胰島素原和類胰島素生長因子結合蛋白）；（凝血蛋白和凝血相關蛋白，尤其諸如，VIII因子、組織因子、範威爾邦德（von Willebrand）因子、蛋白質C、 $\alpha$ -1-抗胰蛋白酶、纖溶酶原活化劑（諸如尿激酶和組織纖溶酶原

活化劑（「t-PA」））、邦巴辛（bombazine）、凝血酶、血小板生成素和血小板生成素受體、群落刺激因子（CSF）（尤其包括以下物質：M-CSF、GM-CSF和G-CSF）、其他血液和血清蛋白（包括但不限於白蛋白、IgE和血型抗原）、受體和受體相關蛋白（包括例如flk2/flt3受體、肥胖（OB）受體、生長激素受體和T細胞受體）；(x) 神經滋養因子，包括但不限於骨源性神經滋養因子（BDNF）和神經滋養蛋白-3、神經滋養蛋白-4、神經滋養蛋白-5或神經滋養蛋白-6（NT-3、NT-4、NT-5或NT-6）；(xi) 鬆弛素A鏈、鬆弛素B鏈和鬆弛素原、干擾素（包括例如干擾素 $\alpha$ 、干擾素 $\beta$ 和干擾素 $\gamma$ ）、介白素（IL）（例如IL-1至IL-10、IL-12、IL-15、IL-17、IL-23、IL-12/IL-23、IL-2Ra、IL1-R1、IL-6受體、IL-4受體和/或IL-13受體、IL-13RA2或IL-17受體、IL-1RAP）；(xiv) 病毒抗原，包括但不限於AIDS包膜病毒抗原、脂蛋白、降鈣素、升糖素、心房利尿鈉因子、肺表面活性劑、腫瘤壞死因子- $\alpha$ 和腫瘤壞死因子- $\beta$ 、腦啡肽酶、BCMA、IgKappa、ROR-1、ERBB2、間皮素、RANTES（受激活調節的正常T細胞表現與分泌因子）、小鼠促性腺激素相關肽、DNA酶、FR- $\alpha$ 、抑制素和激活素、整合素、蛋白質A或D、類風濕因子、免疫毒素、骨成形性蛋白質（BMP）、超氧化物歧化酶、表面膜蛋白、衰退加速因子（DAF）、AIDS包膜、運輸蛋白、歸巢受體、MIC（MIC-a、MIC-B）、ULBP 1-6、EPCAM、位址素、調節蛋白、免疫黏附素、抗原結合蛋白、生長激素、CTGF、CTLA4、伊紅趨素（eotaxin）-1、MUC1、CEA、c-MET、密連蛋白（Claudin）-18、GPC-3、EPHA2、FPA、LMP1、MG7、NY-ESO-1、PSCA、神經節苷脂GD2、神經節苷脂GM2、BAFF、OPGL（RANKL）、肌生成抑制素、Dickkopf-1（DKK-

1) 、Ang2、NGF、IGF-1受體、肝細胞生長因子(HGF)、TRAIL-R2、c-Kit、B7RP-1、PSMA、NKG2D-1、計劃性細胞死亡蛋白1和配位基、PD1和PDL1、甘露糖受體/hCG $\beta$ 、丙型肝炎病毒、間皮素dsFv[PE38 軛合物、呼吸道病菌(lly)、IFN  $\gamma$ 、 $\gamma$ 干擾素誘導蛋白10(IP10)、IFNAR、TALL-1、胸腺基質淋巴細胞生成素(TSLP)、前蛋白轉化酶枯草桿菌蛋白酶/Kexin 9型(PCSK9)、幹細胞因子、Flt-3、抑鈣素基因系胜肽(CGRP)、OX40L、 $\alpha$ 4 $\beta$ 7、血小板特異性(血小板糖蛋白Iib/IIIb(PAC-1)、轉變生長因子 $\beta$ (TFG $\beta$ )、透明帶精子結合蛋白3(ZP-3)、TWEAK、血小板衍生的生長因子受體 $\alpha$ (PDGFR $\alpha$ )、硬化蛋白(sclerostin)以及任何前述內容的生物活性片段或變體。

【0115】在另一個實施方式中，蛋白質包括阿昔單抗(abciximab)、阿達木單抗(adalimumab)、阿德木單抗(adecatumumab)、阿柏西普(aflibercept)、阿侖單抗(alemtuzumab)、阿利庫單抗(alirocumab)、阿那白滯素(anakinra)、阿塞西普(atacicept)、巴厘昔單抗(basiliximab)、貝利木單抗(belimumab)、貝伐單抗(bevacizumab)、生物素單抗(biosozumab)、博納吐單抗(blinatumomab)、本妥昔單抗(brentuximab vedotin)、布羅達單抗(brodalumab)、莫坎妥珠單抗(cantuzumab mertansine)、康納單抗(canakinumab)、西妥昔單抗(cetuximab)、塞妥珠單抗(certolizumab pegol)、可那木單抗(conatumumab)、達利珠單抗(daclizumab)、迪諾舒單抗(denosumab)、依庫麗單抗(eculizumab)、依決洛單抗(edrecolomab)、依法利珠單抗(efalizumab)、依帕珠單抗

( epratuzumab ) 、 依 那 西 普 ( etanercept ) 、 依 伏 庫 單 抗 ( evolocumab ) 、 加 利 昔 單 抗 ( galiximab ) 、 蓋 尼 塔 單 抗 ( ganitumab ) 、 吉 妥 珠 單 抗 ( gemtuzumab ) 、 戈 利 木 單 抗 ( golimumab ) 、 替 伊 莫 單 抗 ( ibritumomab tiuxetan ) 、 英 夫 利 昔 單 抗 ( infliximab ) 、 易 普 利 姆 瑪 ( ipilimumab ) 、 樂 地 單 抗 ( lerdelimumab ) 、 魯 昔 單 抗 ( lumiliximab ) 、 左 旋 單 抗 ( lxdkizumab ) 、 馬 帕 木 單 抗 ( mapatumumab ) 、 磷 酸 莫 特 沙 尼 ( motesanib diphosphate ) 、 莫 羅 單 抗 -CD3 ( muromonab-CD3 ) 、 那 他 珠 單 抗 ( natalizumab ) 、 奈 西 立 肽 ( nesiritide ) 、 尼 妥 珠 單 抗 ( nimotuzumab ) 、 納 武 單 抗 ( nivolumab ) 、 奧 瑞 珠 單 抗 ( ocrelizumab ) 、 奧 法 木 單 抗 ( ofatumumab ) 、 奧 馬 珠 單 抗 ( omalizumab ) 、 奧 普 瑞 介 白 素 ( oprelvekin ) 、 帕 利 珠 單 抗 ( palivizumab ) 、 帕 尼 單 抗 ( panitumumab ) 、 派 姆 單 抗 ( pembrolizumab ) 、 帕 妥 珠 單 抗 ( pertuzumab ) 、 培 克 珠 單 抗 ( pexelizumab ) 、 蘭 尼 單 抗 ( ranibizumab ) 、 利 妥 木 單 抗 ( rilotumumab ) 、 利 妥 昔 單 抗 ( rituximab ) 、 羅 米 司 亭 ( romiplostim ) 、 洛 莫 索 珠 單 抗 ( romosozumab ) 、 沙 格 司 亭 ( sargamostim ) 、 托 珠 單 抗 ( tocilizumab ) 、 托 西 莫 單 抗 ( tositumomab ) 、 曲 妥 單 抗 ( trastuzumab ) 、 優 特 克 單 抗 ( ustekinumab ) 、 維 多 珠 單 抗 ( vedolizumab ) 、 維 西 珠 單 抗 ( visilizumab ) 、 伏 洛 昔 單 抗 ( volociximab ) 、 紮 木 單 抗 ( zanolimumab ) 、 紮 魯 木 單 抗 ( zalutumumab ) 、 以 及 前 述 任 何 內 容 的 生 物 仿 製 藥 。

【0116】 蛋白質涵蓋所有前述內容，並且進一步包括包含上述任何抗體的1、2、3、4、5或6個互補決定區（CDR）的抗體。還包括這樣的變體，其包括與感興趣蛋白質的參考胺基酸序列具有70%或更高、特別是80%或更高、更特別是90%或更高、再更特別是95%或更高、具體是97%或更高、更具體是98%或更高、再更具體是99%或更高同一性的胺基酸序列的區。在這方面的同一性可以使用多種眾所周知的且容易獲得的胺基酸序列分析軟體來確定。較佳軟體包括實施史密斯-沃特曼（Smith-Waterman）演算法的那些軟體，所述軟體被認為是搜索和比對序列問題的令人滿意的解決方案。還可以採用其他演算法，特別是在速度係重要考慮因素的情況下。可以用於此方面的用於DNA、RNA和多肽的比對和同源性匹配的常用程式包括FASTA、TFASTA、BLASTN、BLASTP、BLASTX、TBLASTN、PROSRCH、BLAZE和MPSRCH，後者係用於在MasPar製造的大規模並行處理器上執行的史密斯-沃特曼演算法的實施方式。

【0117】 本文描述的一些圖展示了具有一個或多個功能部件的示例框圖。應理解的是，這種框圖係出於說明的目的並且所描述和示出的設備可以比所展示的具有額外的、更少的、或替代的部件。此外，在不同實施方式中，部件（以及由相應部件提供的功能）可以與任何合適部件相關聯或以其他方式集成為其一部分。

【0118】 本揭露的實施方案係關於非暫態電腦可讀存儲介質，在該非暫態電腦可讀存儲介質上具有用於執行不同電腦實施操作的電腦代碼。術語「電腦可讀存儲介質」在本文中用於包括能夠存儲或編碼用於執行操作、方法、和本文描述的技術的一系列指令或電腦代碼的任何介質。介質

和電腦代碼可以是為了本揭露的實施方式的目的而特別設計和構造的介質和電腦代碼，或者它們可以是熟悉電腦軟體技術者公知和可獲得的類型。電腦可讀存儲介質的示例包括、但不限於：磁性介質，諸如硬碟、軟碟、和磁帶；光學介質，諸如CD-ROM和全息設備；磁光介質，諸如光碟；以及硬體設備，該等硬體設備被特別配置成存儲和執行程式碼，諸如ASIC、可程式設計邏輯器件（「PLD」）、以及ROM和RAM設備。

**【0119】** 電腦代碼的示例包括諸如由編譯器產生的機械代碼、以及包含由電腦使用解譯器或編譯器執行的較高級代碼的檔。例如，可以使用Java、C++、或其他物件導向的程式設計語言和開發工具實施本揭露的實施方式。電腦代碼的附加示例包括加密代碼和壓縮代碼。此外，本揭露的實施方式可以作為電腦程式產品被下載，該電腦程式產品可以經由發送通道從遠端電腦（例如，伺服器電腦）傳遞至請求電腦（例如，用戶端電腦或不同的伺服器電腦）。本揭露的另一個實施方式可以用硬接線電路系統代替機器可執行軟體指令或與其組合來實施。

**【0120】** 如本文使用的，除非上下文另有明確指明，否則單數術語「一（a、an）」和「該（the）」可以包括複數引用物。

**【0121】** 如本文使用的，術語「連接」、「連接的」和「連接件」指代聯接或連結操作。連接的部件可以例如藉由另一組部件直接或間接地彼此聯接。

**【0122】** 如本文使用的，術語「大致」、「實質上」、「實質」和「約」用來描述並且解釋小的變化。當結合事件或情況一起使用時，該等術語可以指代事件或情況恰好發生的情形以及事件或情況近似發生的情形。例如，當結合數值使用時，該等術語可以指代那個數值的小於或等於

$\pm 10\%$ 、諸如小於或等於  $\pm 5\%$ ，小於或等於  $\pm 4\%$ 、小於或等於  $\pm 3\%$ 、小於或等於  $\pm 2\%$ 、小於或等於  $\pm 1\%$ 、小於或等於  $\pm 0.5\%$ 、小於或等於  $\pm 0.1\%$ 、或小於或等於  $\pm 0.05\%$  的變化範圍。例如，如果值之間的差小於或等於值的平均值的  $\pm 10\%$ 、諸如小於或等於  $\pm 5\%$ 、小於或等於  $\pm 4\%$ 、小於或等於  $\pm 3\%$ 、小於或等於  $\pm 2\%$ 、小於或等於  $\pm 1\%$ 、小於或等於  $\pm 0.5\%$ 、小於或等於  $\pm 0.1\%$ 、或小於或等於  $\pm 0.05\%$ ，則可以認為兩個數值「基本上」相同。

**【0123】** 此外，數量、比率、以及其他數值有時以範圍格式在本文中呈現。應理解的是這種範圍格式係為了方便和簡潔性而使用的並且應靈活地理解為包括明確指定為範圍極限的數值，但是也包括包含在那個範圍內的所有單獨數值和子範圍，就好像每個數值或子範圍明確指定了一樣。

**【0124】** 雖然已經參考本揭露的特定實施方式描述和展示了本揭露，但是該等描述和圖示不限制本揭露。熟悉該項技術者應理解的是：在不脫離由所附申請專利範圍限定之本揭露之真實精神和範圍的情況下，可以進行各種改變並且可以替換等同物。該等圖示可能不一定是按比例繪製的。由於製造製程、容差和/或其他原因，本揭露中的藝術再現與實際設備之間可能存在不同。可以存在沒有具體地展示的本揭露的其他實施方式。說明書（除了申請專利範圍之外）和圖式應被視為說明性的而非限制性的。可以進行修改以使特定情況、材料、物質組成、技術、或過程適應本揭露的目的、精神和範圍。所有的該等變化旨在落入所附申請專利範圍的範圍內。雖然已經參考按特定順序進行的特定操作描述了本文揭露的技術，但是應理解的是，該等操作可以組合、細分、或重新排序以在不脫離本揭露的教導的情況下形成等同的技術。因此，除非本文具體地指示，否

則操作的順序和分組並不是對本揭露的限制。

【符號說明】

【0125】

|     |             |
|-----|-------------|
| 100 | 系統          |
| 102 | 生物反應器       |
| 104 | 一個或多個分析儀器   |
| 106 | 拉曼分析儀       |
| 108 | 拉曼探針        |
| 110 | 電腦          |
| 112 | 資料庫伺服器      |
| 114 | 網路          |
| 120 | 處理單元        |
| 122 | 網路介面        |
| 124 | 顯示器         |
| 126 | 用戶輸入裝置      |
| 128 | 記憶體         |
| 130 | JITL預測器應用程式 |
| 132 | 本地模型        |
| 136 | 觀察結果資料庫     |
| 140 | 查詢單元        |
| 142 | 本地模型生成器     |
| 144 | 預測單元        |
| 146 | 資料庫維護單元     |

|      |             |
|------|-------------|
| 150  | 系統          |
| 152  | 控制單元        |
| 154  | 葡萄糖泵        |
| 202  | 虛線          |
| 204  | 圓圈          |
| 206  | 實線          |
| 208  | 陰影區域        |
| 250  | 資料流程        |
| 252  | 光譜數據        |
| 254  | 查詢點         |
| 256  | 總體資料集       |
| 258  | 區域資料集       |
| 260  | 本地模型        |
| 262  | 所預測輸出       |
| 300  | 資料流程        |
| 302  | 光譜數據        |
| 304  | 查詢點         |
| 306  | 總體資料集       |
| 307A | 最近k個條目      |
| 307B | 最近k個條目之前的條目 |
| 308  | 區域資料集       |
| 310  | 本地模型        |
| 312  | 所預測輸出       |

|      |             |
|------|-------------|
| 314  | 新條目         |
| 350  | 資料流程        |
| 352  | 光譜資料        |
| 354  | 查詢點         |
| 356  | 總體資料集       |
| 357A | 最近k個條目      |
| 357B | 最近k個條目之前的條目 |
| 358  | 區域資料集       |
| 360  | 本地模型        |
| 362  | 所預測輸出       |
| 364  | 新條目         |
| 400  | 方法          |
| 402  | 方法步驟框       |
| 404  | 方法步驟框       |
| 406  | 方法步驟框       |
| 408  | 方法步驟框       |

## 【發明申請專利範圍】

### 【第1項】

一種用於監測和/或控制生物製藥過程的電腦實施之方法，該方法包括：

由一個或多個處理器確定與由光譜系統對該生物製藥過程進行的掃描相關聯的查詢點；

由該一個或多個處理器查詢觀察結果資料庫，該觀察結果資料庫包含與生物製藥過程的過去觀察結果相關聯的多個觀察結果資料集，其中，該等觀察結果資料集中的每一個包括光譜資料和相應的實際分析測量結果，並且其中，查詢該觀察結果資料庫包括從該多個觀察結果資料集中選擇滿足關於該查詢點的一個或多個相關性標準的觀察結果資料集作為訓練資料；

由該一個或多個處理器並使用所選訓練資料來校準特定於該生物製藥過程的本地模型，訓練該本地模型以基於光譜資料登錄來預測分析測量結果；以及

由該一個或多個處理器預測該生物製藥過程的分析測量結果，其中，預測該生物製藥過程的分析測量結果包括使用該本地模型來分析該光譜系統在掃描該生物製藥過程時所生成的光譜資料。

### 【第2項】

如請求項1之電腦實施之方法，其中至少下列一者：

a) 該光譜系統係拉曼光譜系統；或

b) 確定查詢點包括至少部分地基於光譜掃描向量來確定該查詢點，該光譜掃描向量係由該光譜系統在掃描該生物製藥過程時生成的；並且

選擇滿足關於該查詢點的一個或多個相關性標準的觀察結果資料集作為訓練資料包括：將確定該查詢點所基於的光譜掃描向量與同該等生物製藥過程的過去觀察結果相關聯的光譜掃描向量進行比較。

**【第3項】**

如請求項2之電腦實施之方法，其中：

確定查詢點進一步包括：基於與該光譜掃描向量相關聯的樣本數來確定該查詢點；並且

選擇滿足關於該查詢點的一個或多個相關性標準的觀察結果資料集作為訓練資料包括：(i) 將確定該查詢點所基於的光譜掃描向量與同該等生物製藥過程的過去觀察結果相關聯的光譜掃描向量進行比較；以及 (ii) 將與該查詢點相關聯的樣本數與同該等生物製藥過程的過去觀察結果相關聯的樣本數進行比較。

**【第4項】**

如請求項3之電腦實施之方法，其中，選擇滿足關於該查詢點的一個或多個相關性標準的觀察結果資料集作為訓練資料包括：

選擇最新的 $k$ 個觀察結果資料集以包括在該訓練資料中。

**【第5項】**

如請求項2至4中任一項之電腦實施之方法，其中至少下列一者：

a) 預測該生物製藥過程的分析測量結果包括：

使用該本地模型來分析確定該查詢點所基於的光譜掃描向量；或

b) 選擇滿足關於該查詢點的一個或多個相關性標準的觀察結果資料集作為訓練資料包括：

計算 (i) 確定該查詢點所基於的光譜掃描向量與 (ii) 同該等生物製

藥過程的過去觀察結果相關聯的光譜掃描向量之間的距離；以及

選擇與該等過去觀察結果相關聯的光譜掃描向量中的、在確定該查詢點所基於的光譜掃描向量的閾值距離之內的所有光譜掃描向量作為訓練資料。

**【第6項】**

如請求項1至4中任一項之電腦實施之方法，其中至少下列一者：

a) 確定查詢點包括：

至少部分地基於以下中的一項或兩項來確定該查詢點：(i) 與該生物製藥過程相關聯的培養基配置，以及 (ii) 分析該生物製藥過程時的一個或多個操作條件；

b) 校準特定於該生物製藥過程的本地模型包括：

校準特定於該生物製藥過程的高斯過程機器學習模型；

c) 校準特定於該生物製藥過程的本地模型包括：

校準作為給定觀察結果資料集的光譜資料以及樣本數兩者的函數的模型；

d) 預測該生物製藥過程的分析測量結果包括：

使用該本地模型來確定與該生物製藥過程的所預測分析測量結果相關聯的置信度指標；

e) 該方法進一步包括：由該一個或多個處理器並至少部分地基於該生物製藥過程的所預測分析測量結果來控制該生物製藥過程的至少一個參數；

f) 該生物製藥過程的所預測分析測量結果係培養基成分濃度、培養基狀態、活細胞密度、滴定度、關鍵品質屬性或細胞狀態；

g) 該生物製藥過程的所預測分析測量結果係葡萄糖、乳酸鹽、麩胺酸鹽、麩醯胺酸、氨、胺基酸、 $\text{Na}^+$ 或 $\text{K}^+$ 的濃度；

h) 該生物製藥過程的所預測分析測量結果係pH、 $\text{pCO}_2$ 、 $\text{pO}_2$ 、溫度或滲透壓；

i) 該方法進一步包括：由分析儀器獲得該生物製藥過程的實際分析測量結果；以及

由該一個或多個處理器將 (i) 該光譜系統在獲得該實際分析測量結果時所生成的光譜資料和 (ii) 該生物製藥過程的實際分析測量結果添加到該觀察結果資料庫；或

j) 該生物製藥過程係細胞培養過程。

#### 【第7項】

如請求項6之電腦實施之方法，進一步包括：

由該一個或多個處理器確定至少所預測分析測量結果不滿足一個或多個模型性能標準，

其中，獲得該實際分析測量結果係響應於確定至少所預測分析測量結果不滿足該一個或多個模型性能標準而執行的。

#### 【第8項】

如請求項7之電腦實施之方法，其中，確定至少所預測分析測量結果不滿足該一個或多個模型性能標準包括：

生成與所預測分析測量結果相關聯的可信度區間；以及

將該可信度區間與預定義閾值進行比較。

#### 【第9項】

一種用於監測和/或控制生物製藥過程之光譜系統，該光譜系統包

括：

一個或多個光譜探針，該一個或多個光譜探針被共同配置成：(i) 將源電磁輻射遞送到該生物製藥過程，以及 (ii) 在將該源電磁輻射遞送到該生物製藥過程的同時收集電磁輻射；

一個或多個記憶體，該一個或多個記憶體共同存儲觀察結果資料庫，該觀察結果資料庫包含與生物製藥過程的過去觀察結果相關聯的多個觀察結果資料集，其中，該等觀察結果資料集中的每一個包括光譜資料和相應的實際分析測量結果；以及

一個或多個處理器，該一個或多個處理器被配置成：

確定與由該光譜系統對該生物製藥過程進行的掃描相關聯的查詢點；

至少藉由從該多個觀察結果資料集中選擇滿足關於該查詢點的一個或多個相關性標準的觀察結果資料集作為訓練資料來查詢該觀察結果資料庫；

使用所選訓練資料來校準特定於該生物製藥過程的本地模型，訓練該本地模型以基於光譜資料登錄來預測分析測量結果；以及

至少藉由使用該本地模型分析該光譜系統在利用該一個或多個光譜探針掃描該生物製藥過程時所生成的光譜資料來預測該生物製藥過程的分析測量結果。

#### 【第10項】

如請求項9之光譜系統，其中至少下列一者：

- a) 該光譜系統係拉曼光譜系統；
- b) 該一個或多個處理器被配置成：

至少部分地基於光譜掃描向量來確定該查詢點，該光譜掃描向量係由該光譜系統在掃描該生物製藥過程時生成的；以及

至少藉由將確定該查詢點所基於的光譜掃描向量與同該等生物製藥過程的過去觀察結果相關聯的光譜掃描向量進行比較來選擇該訓練資料；

c) 該本地模型係高斯過程機器學習模型；

d) 該本地模型係給定觀察結果資料集的光譜資料以及樣本數兩者的函數；

e) 該一個或多個處理器進一步被配置成：使用該本地模型來確定與該生物製藥過程的所預測分析測量結果相關聯的置信度指標；

f) 該一個或多個處理器進一步被配置成：至少部分地基於該生物製藥過程的所預測分析測量結果來控制該生物製藥過程的至少一個參數；

g) 該生物製藥過程的所預測分析測量結果係培養基成分濃度、培養基狀態、活細胞密度、滴定度、關鍵品質屬性或細胞狀態；

h) 該生物製藥過程的所預測分析測量結果係葡萄糖、乳酸鹽、麩胺酸鹽、麩醯胺酸、氨、胺基酸、 $\text{Na}^+$ 或 $\text{K}^+$ 的濃度；

i) 該生物製藥過程的所預測分析測量結果係pH、 $\text{pCO}_2$ 、 $\text{pO}_2$ 、溫度或滲透壓；

j) 進一步包括：分析儀器，該分析儀器被配置成獲得該生物製藥過程的實際分析測量結果，且其中該一個或多個處理器進一步被配置成將 (i) 該光譜系統在獲得該實際分析測量結果時所生成的光譜資料和 (ii) 該生物製藥過程的實際分析測量結果添加到該觀察結果資料庫；

k) 該一個或多個處理器進一步被配置成：確定至少所預測分析測量結果不滿足一個或多個模型性能標準；以及

響應於確定至少所預測分析測量結果不滿足該一個或多個模型性能標準而從該分析儀器獲得該實際分析測量結果；

l) 該一個或多個處理器被配置成至少藉由以下操作來確定至少所預測分析測量結果不滿足該一個或多個模型性能標準：

生成與所預測分析測量結果相關聯的可信度區間；以及

將該可信度區間與預定義閾值進行比較；或

m) 該生物製藥過程係細胞培養過程。

### 【第11項】

如請求項10之光譜系統，其中，該一個或多個處理器被配置成：

至少部分地基於與該光譜掃描向量相關聯的樣本數來確定該查詢點；以及

部分地藉由以下操作來選擇滿足關於該查詢點的一個或多個相關性標準的觀察結果資料集作為訓練資料：(i) 將確定該查詢點所基於的光譜掃描向量與同該等生物製藥過程的過去觀察結果相關聯的光譜掃描向量進行比較；以及 (ii) 將與該查詢點相關聯的樣本數與同該等生物製藥過程的過去觀察結果相關聯的樣本數進行比較。

### 【第12項】

如請求項11之光譜系統，其中，該一個或多個處理器被配置成部分地藉由以下操作來選擇滿足關於該查詢點的一個或多個相關性標準的觀察結果資料集作為訓練資料：

選擇最新的 $k$ 個觀察結果資料集以包括在該訓練資料中。

### 【第13項】

一種非暫態電腦可讀存儲介質，存儲用於監測和/或控制生物製藥過

程的指令，其中，該等指令當被一個或多個處理器執行時使所述一個或多個處理器執行如請求項1至8中任一項之方法。

**【第14項】**

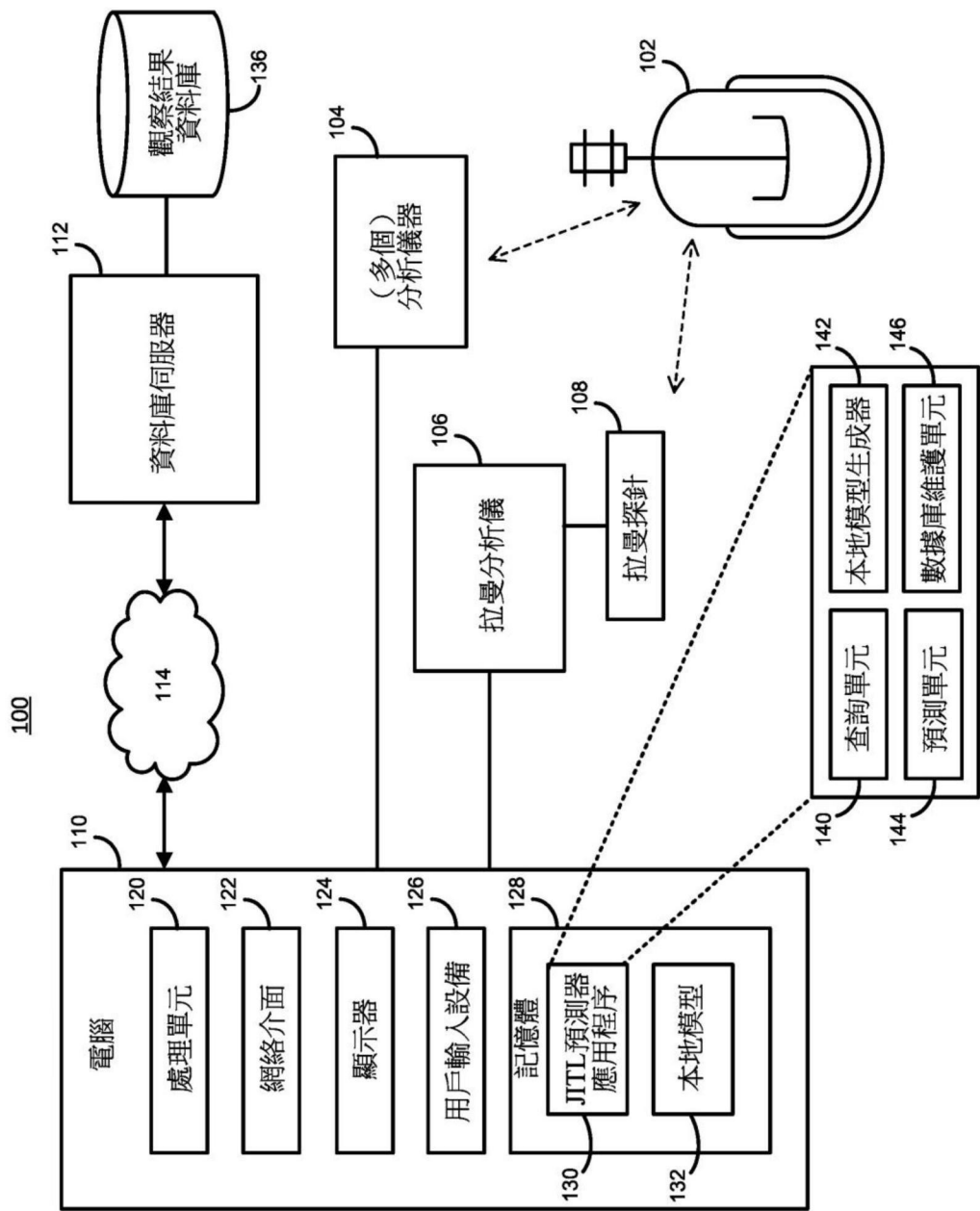
一種生物反應器系統，包括：

生物反應器室，該生物反應器室被配置成用於容納生物製藥過程；及如請求項9至12中任一項之光譜系統。

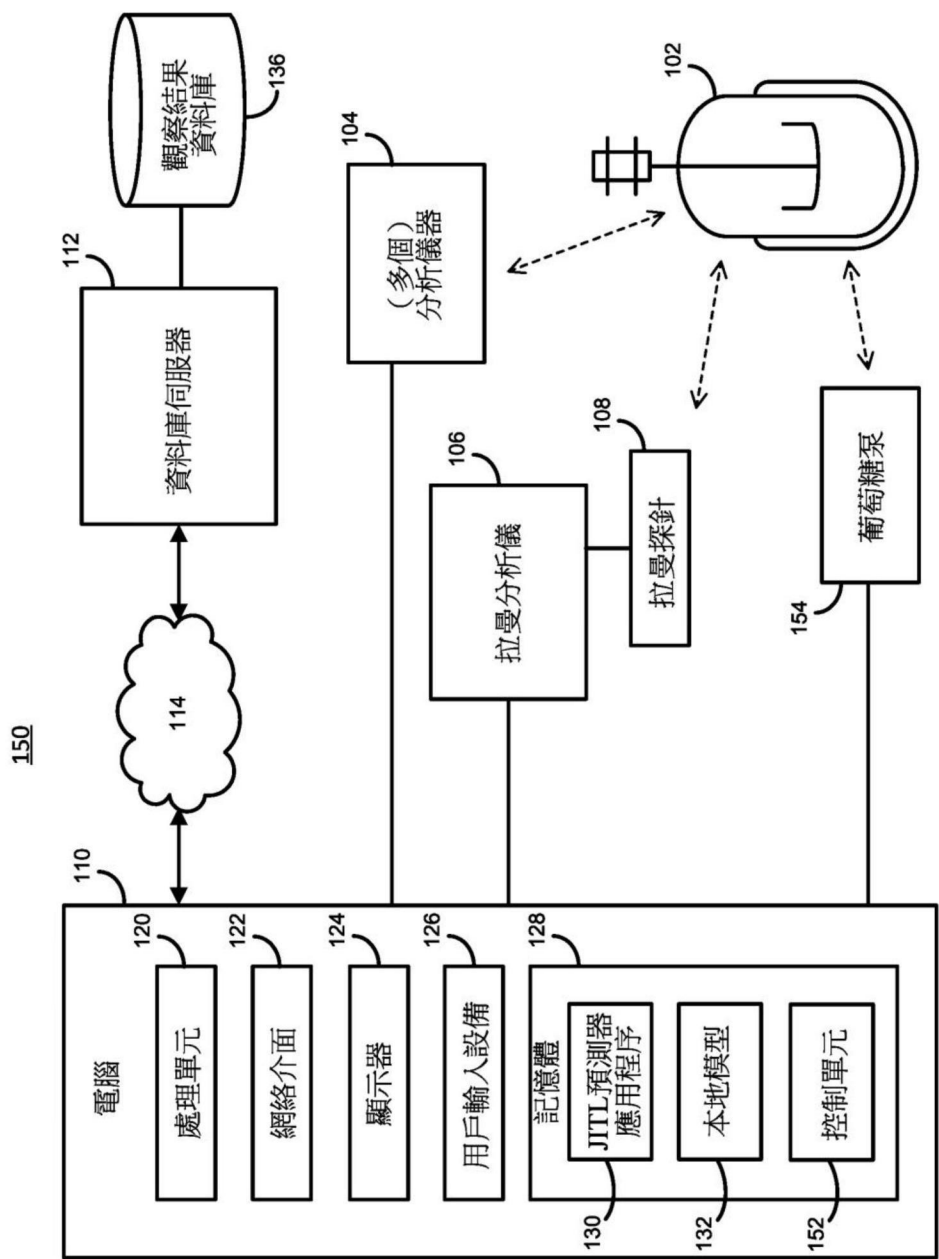
**【第15項】**

如請求項14之生物反應器系統，其被配置成於細胞培養過程中產生重組蛋白質。

【發明圖式】

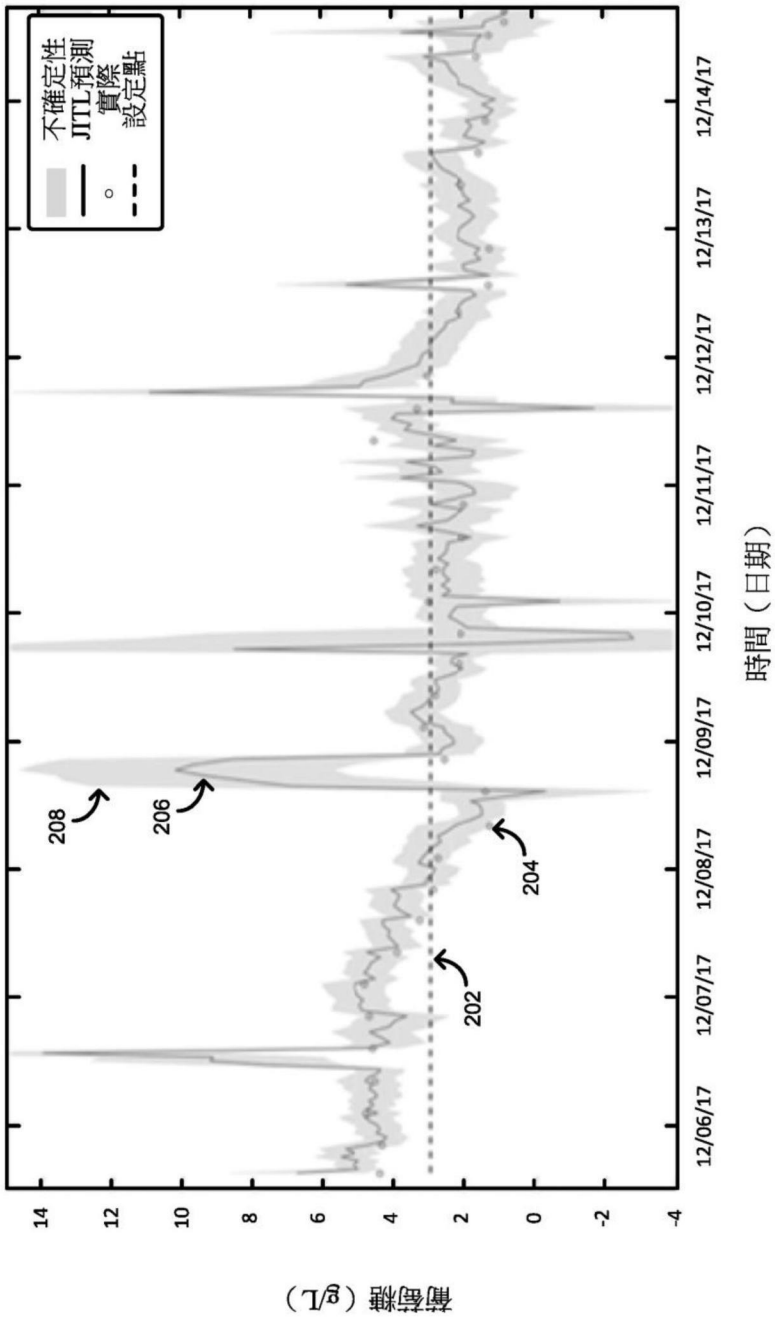


【圖1】



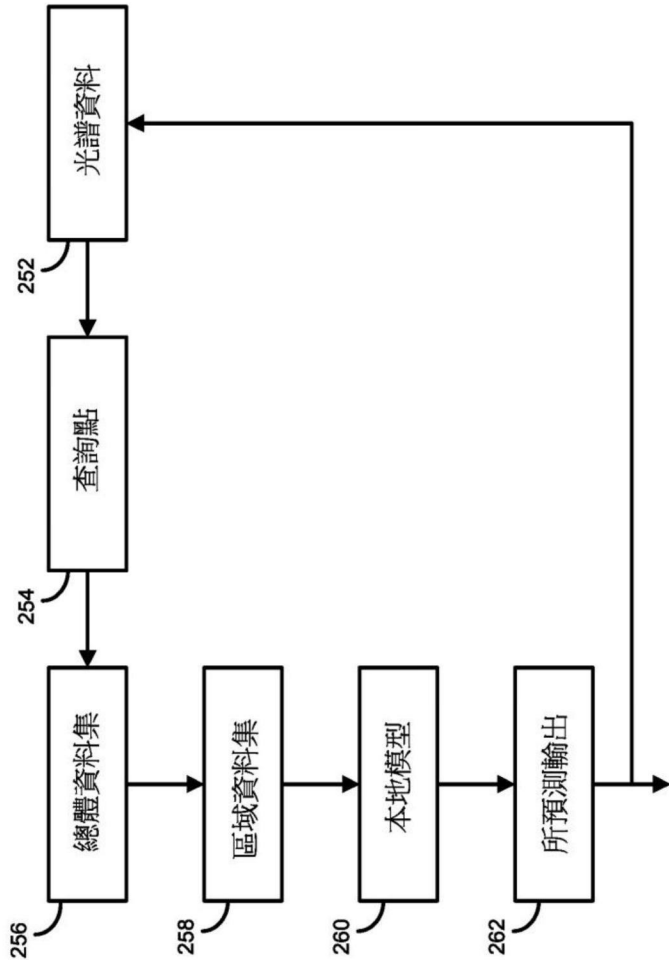
【圖2】

200

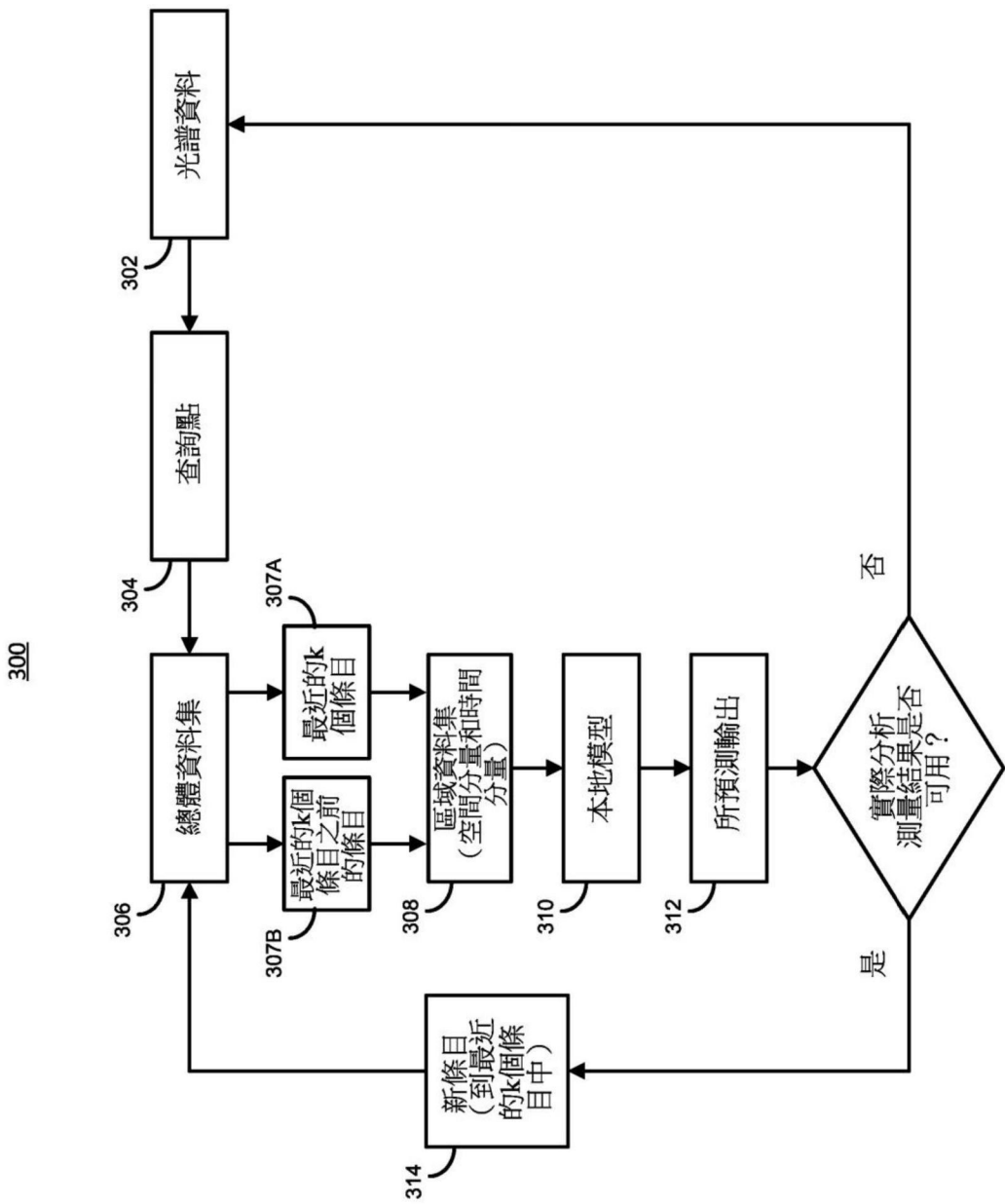


【圖3】

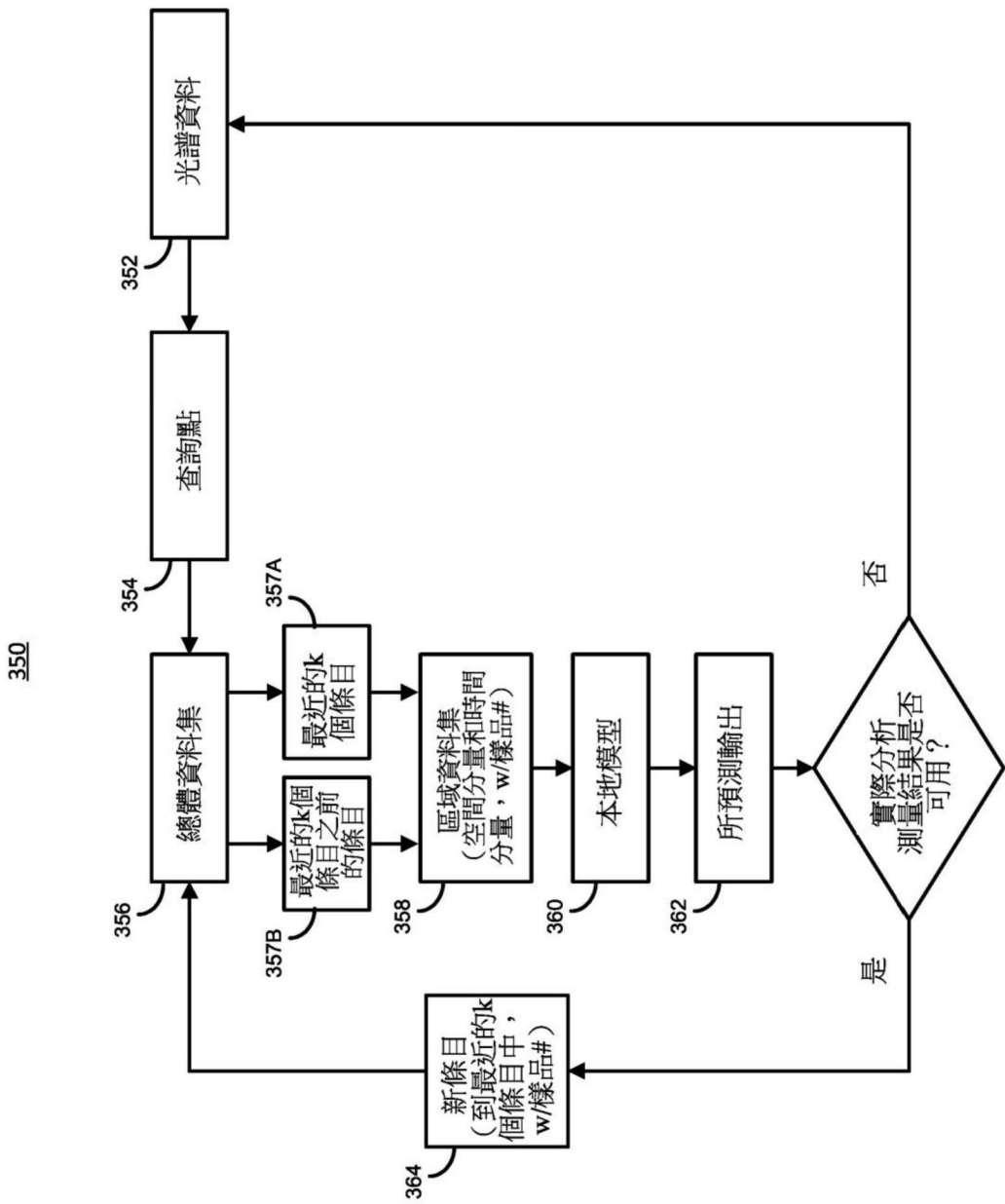
250



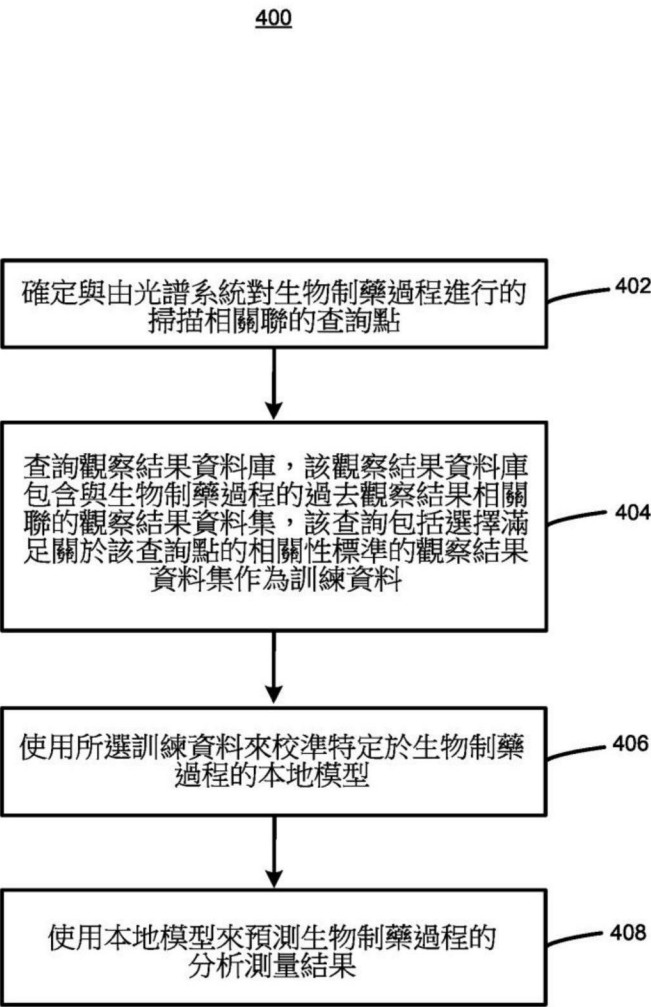
【圖4】



【圖5】



【圖6】



【圖7】