ARTICLE

Macroscopic modeling of bioreactors for recombinant protein producing *Pichia pastoris* in defined medium

Moo Sun Hong[1] | M. Lourdes Velez-Suberbie[2] | Andrew J. Maloney[1] |
Andrew Biedermann[1,3] | Kerry R. Love[3] | J. Christopher Love[1,3] |
Tarit K. Mukhopadhyay[2] | Richard D. Braatz[1]

[1]Department of Chemical Engineering,
Massachusetts Institute of Technology,
Cambridge, Massachusetts, USA

[2]Department of Biochemical Engineering,
University College London, London, UK

[3]Koch Institute for Integrative Cancer
Research, Massachusetts Institute of
Technology, Cambridge, Massachusetts, USA

Correspondence
Richard D. Braatz, Department of Chemical
Engineering, Massachusetts Institute of
Technology, 77 Massachusetts Avenue,
Cambridge, MA 02139, USA.
Email: braatz@mit.edu

Funding information
Bill and Melinda Gates Foundation

Abstract

The methylotrophic yeast *Pichia pastoris* is widely used as a microbial host for recombinant protein production. Bioreactor models for *P. pastoris* can inform understanding of cellular metabolism and can be used to optimize bioreactor operation. This article constructs an extensive macroscopic bioreactor model for *P. pastoris* which describes substrates, biomass, total protein, other medium components, and off-gas components. Species and elemental balances are introduced to describe uptake and evolution rates for medium components and off-gas components. Additionally, a pH model is constructed using an overall charge balance, acid/base equilibria, and activity coefficients to describe production of recombinant protein and precipitation of medium components. The extent of run-to-run variability is modeled by distributions of a subset of the model parameters, which are estimated using the maximum likelihood method. Model prediction from the extensive macroscopic bioreactor model well describes experimental data with different operating conditions. The probability distributions of the model predictions quantified from the parameter distribution are quantifiably consistent with the run-to-run variability observed in the experimental data. The uncertainty description in this macroscopic bioreactor model identifies the model parameters that have large variability and provides guidance as to which aspects of cellular metabolism should be the focus of additional experimental studies. The model for medium components with pH and precipitation can be used for improving chemically defined medium by minimizing the amount of components needed while meeting cellular requirements.

KEYWORDS
bioreactor modeling, defined medium, parameter estimation, *Pichia pastoris*, recombinant protein production

1 | INTRODUCTION

The methylotrophic yeast *Pichia pastoris* is widely used as a microbial host for recombinant protein production. This system has the advantages of (i) tight gene regulation and high product titers with methanol-induced alcohol oxidase (AOX1) promoter, (ii) growth on simple media at high cell density, (iii) eukaryotic posttranslational modification such as glycosylation and disulfide bond formation, and (iv) simplified downstream purification with foreign protein secretion and low amount of naturally secreted proteins (Cereghino & Cregg, 2000).

Bioreactor modeling can inform understanding of cellular metabolism—including aspects of growth kinetics, productivity, media consumption, and metabolite secretion—and can be used to optimize bioreactor operations. Models for bioreactors are developed with different levels of detail and complexity depending on their purpose. With simple mathematical formulations and low computational costs, macroscopic models can produce useful predictions of bioreactor operations while being suitable for parameter estimation and real-time model predictive control (Craven et al., 2014; Ramaswamy et al., 2005).

Although *P. pastoris* has been used to produce numerous recombinant proteins, limited macroscopic models have been published, with their main focus being on the substrate consumption, growth, and production (Barrigon et al., 2015; Çelik et al., 2009; Jahic et al., 2002; Ren et al., 2003). More comprehensive macroscopic models are necessary for *P. pastoris* to develop further understanding of cellular metabolism and optimization of bioreactor operations. In this study, an extensive macroscopic model for a bioreactor was constructed for substrates (glycerol and methanol), biomass, total protein, important medium components, and off-gas components (oxygen and carbon dioxide). The modeling of the medium components is facilitated by using a chemically defined medium (Matthews et al., 2018), which also reduces run-to-run variability and simplifies purification in the downstream processes. Species and elemental balances are introduced to describe uptake and evolution rates for medium components and off-gas components. Additionally, a pH model was constructed using an overall charge balance, acid/base equilibria, and activity coefficients to describe the dependence of recombinant protein production on pH, precipitation of medium components, and carbon dioxide to carbonate species reactions.

For a set of bioreactors operated in fed-batch with *P. pastoris*, variations among individual runs are observed, especially with respect to volumetric productivity for an exemplar recombinant subunit vaccine component for rotavirus. These variations are observed even with highly similar target operating conditions with controlled temperature, pH, and dissolved oxygen and reproducible substrate consumption and growth rates (Plantz et al., 2006; Surribas et al., 2006; Velez-Suberbie et al., 2020). These observations suggest that the experimental data for the bioreactors are not describable by a mathematical model with a single set of parameters. An additional consideration is that any macroscopic model does not describe all of the complex biology that occurs in the bioreactor; such models effectively lump multiple biological pathways into some of the model parameters. In this study, the extent of run-to-run variability is modeled by using distributions for a subset of the model parameters, in which each run is associated with a single set of model parameters. The

TABLE 1 Fermentation reference, pH, and methanol feed rate

Fermentation ID	pH	Methanol feed rate (ml/L per L of initial volume)
RDM 9.1, 9.2, & 10.2	4.0	7.6
RDM 9.4 & 10.4	6.5	7.6
RDM 9.3, 10.1, & 10.3	6.5–3.0–6.5	7.6
RDM 11.2 & 11.4	4.0	11.0
RDM 14.2 & 14.3	6.5–3.0–6.5	11.0
RDM 14.4	6.5–3.0–6.5	13.0

model parameters that have been observed in the literature to have high reproducibility from run to run including in different laboratories (substrate consumption and growth rate parameters) are set to literature values, and the distributions of the uncertain parameters in the macroscopic bioreactor model are estimated using the maximum likelihood method.

2 | MATERIALS AND METHODS

2.1 | Fed-batch fermentation

Fermentation was carried out with *P. pastoris* (*Komagataella phaffii* NRRL Y-11430) secreting a P[4] or a P[8] serotype of a non-replicating rotavirus VP8-derived subunit vaccine as previously reported (Velez-Suberbie et al., 2020). Fermenter operating conditions were 25°C, 25% dissolved oxygen tension (DOT), and pH 4.00 or 6.50 ± 0.15 (Table 1) controlled with 10% (v/v) ammonium hydroxide. In some fermentation, a pH pulse was implemented at the end of methanol adaptation (~30 h), which reduced pH to 3.00 ± 0.15, held low for ~4 h, and then ramped pH back up to original set point. Harvest was performed 3 h after the end of the pH pulse.

2.2 | Analytical methods

Biomass and protein concentration were measured as previously reported (Velez-Suberbie et al., 2020). Metabolite concentration in fermentation supernatant were determined off-line using (1) Dionex UltiMate 3000 ultrahigh performance liquid chromatography with refractive index detector (Thermo Fisher Scientific) with Aminex HPX-87H column (Bio-Rad) for glycerol and methanol, (2) CuBiAn HT2 70 photometric biochemistry analyser (Optocell Technology) for phosphate, and (3) BioProfile FLEX (Nova Biomedical) for potassium and ammonium.

3 | MODEL DEVELOPMENT

Bioreactors have been traditionally modeled as stirred-tank reactors. The material balance equation is

$$\frac{dy}{dt} = \frac{F_{in}}{V_L}(y_{in} - y) - \frac{F_{out}}{V_L}(\delta y - y) + r_y, \tag{1}$$

where y is the concentration in the reactor, V_L is the liquid medium volume, F is the media flow rate, the subscripts in and out referring to inlet and outlet streams, δ is the separation factor at the outlet, and r_y is the volumetric reaction rate (Enfors, 2011).

The developed macroscopic bioreactor model is unstructured and nonsegregated, that is, the cell population is treated as single compartment and all cells are considered to be identical. For fed-batch operation, differential balances for the concentrations of biomass X, substrate S (subscript g refers to glycerol and m refers to methanol), total protein P (including recombinant protein and host cell proteins), intracellular carbon C_X, and intracellular nitrogen N_X are

$$\frac{dX}{dt} = \frac{F_{in}}{V_L}(X_{in} - X) + \mu X, \frac{dS}{dt} = \frac{F_{in}}{V_L}(S_{in} - S) - q_S X,$$
$$\frac{dP}{dt} = -\frac{F_{in}}{V_L}P + q_P X, \tag{2}$$

$$\frac{dC_X}{dt} = q_S C_S - \mu C_X - q_P C_P - q_C C_C, \quad \frac{dN_X}{dt} = q_{N_X} - \mu N_X - q_P N_P, \tag{3}$$

where μ is the specific growth rate, q_S is the specific substrate consumption rate, q_P is the specific total protein production rate, q_C is the specific carbon dioxide evolution rate, q_{N_X} is the specific nitrogen uptake rate, C_S is the concentration of carbon in the substrate, C_P is the concentration of carbon in the protein, C_C is the concentration of carbon in the carbon dioxide, and N_P is the concentration of nitrogen in the protein.

3.1 | Cellular metabolism

The specific growth rate is modeled to be proportional to the specific substrate consumption rate, which is most commonly described with the Monod model (Monod, 1949),

$$\mu = (q_S - q_m)Y_{em}, q_S = q_{S,max}\frac{S}{K_S + S}, \tag{4}$$

where q_m is the maintenance coefficient, Y_{em} is the biomass yield coefficient for exclusive maintenance, $q_{S,max}$ is the maximum specific consumption rate, and K_S is the saturation constant.

The specific total protein production rate is the sum of the specific recombinant protein production rate and the specific host cell protein production rate, which are both related to the specific growth rate. The pH of the medium also affects the total protein concentration. Low pH conditions are often applied to reduce degradation of protein by inactivating proteases. However, low extracellular pH also affects the organization of the cell wall (Kapteyn et al., 2001), which may have impact on the protein secretion. For modeling this effect of the pH, it is assumed that there is an enzyme responsible for the secretion of the proteins and it is only active in its original deprotonated form E: $EH^+ \overset{K_{a,E}}{\rightleftarrows} E + H^+$ where $K_{a,E}$ is the acid dissociation constant for the enzyme. Then the active portion of total enzymes is included in the specific total production rate,

$$q_P = \frac{\alpha_P \mu}{10^{pK_{a,E}-pH} + 1}, \tag{5}$$

where α_P is the growth-associated proportional coefficient.

Elemental composition of *P. pastoris* has been reported from several past studies. Carnicer et al. (2009) reported the biomass composition in terms of macromolecular and chemical elements under different oxygenation conditions, as well as under recombinant protein (antibody fragment, Fab) producing and nonproducing conditions. The study showed that the control strain resulted in fairly constant protein and carbohydrate levels under all oxygenation conditions but a Fab-expressing strain resulted in significant increase in protein and decrease in carbohydrate content as the oxygen availability decreased. At the same time, extracellular and intracellular Fab content increased significantly when reducing the oxygen availability. These results suggest that recombinant protein production correlates with the variation in macromolecular and chemical compositions.

Therefore, concentrations of intracellular carbon and nitrogen were included as the states in the model and modeled based on the mass balance of biomass. The specific carbon uptake rate for anabolism and specific nitrogen uptake rate are modeled to be proportional to the specific growth rate,

$$q_{C_X} = q_S C_S - q_C C_C = \alpha_C \mu, \quad q_{N_X} = \alpha_N \mu, \tag{6}$$

where α_C and α_N are the growth-associated proportional coefficients. The concentration of carbon and nitrogen in the protein were calculated from the amino acid sequences of the host cell proteins of *P. pastoris* and the recombinant proteins.

3.2 | Medium components

Molar concentration of RDM components M include total phosphate species P_T, total sulfate species S_T, total ammonium species N_T, total glutamine species Q_T, total arginine species R_T, potassium $[K^+]$, magnesium $[Mg^{2+}]$, calcium $[Ca^{2+}]$, transition metals $[TM^{2+}]$, and chloride $[Cl^-]$:

$$\frac{dM}{dt} = \frac{F_{in}}{V_L}(M_{in} - M) - q_M X - r_{P,M}, \tag{7}$$

where q_M is the specific uptake rate and $r_{P,M}$ is the precipitation (and dissolution) rate for the related species.

Ammonium, glutamine, and arginine species are nitrogen sources that can be utilized simultaneously by *P. pastoris* (Matthews et al., 2018). Nitrogen regulation in *P. pastoris* is simplified as

$$q_M = \frac{n_M M}{N_T + 2Q_T + 4R_T}q_{N_X}, \tag{8}$$

where n_M is the number of nitrogen atoms. For the components without nitrogen, the specific uptake rate is modeled assuming constant elemental composition,

$$q_M = M_X \mu, \tag{9}$$

where M_X is the concentration of intracellular element (phosphorous for phosphate and sulfur for sulfate). This assumption is consistent with constant sulfur and ash composition under different conditions (Carnicer et al., 2009).

3.3 | pH and precipitation

Precipitation caused by medium components is one of several common problems in cell culture media. Here a pH model with precipitation is developed with the medium components. Acid and base species in the medium are described as $\left\{H_{n_{a,i}} A_i^{z_{a,i}+n_{a,i}}, ..., A_i^{z_{a,i}}\right\} = \mathbf{A}_i \in \mathfrak{A}$, where $\mathfrak{A}$ is the family of sets of the acid and base species $\mathbf{A}$, the subscripts i refer to each acid and its conjugate base(s), n_a is the number of potential protons to donate, and z_a is the charge when all potential protons are donated. Remaining species in the medium are described as $S_i^{z_{s,i}} \in \mathbf{S}$, where $\mathbf{S}$ is the set of remaining species. The pH model is based on the total charge balance and acid/base equilibria:

$$0 = [H^+] - [OH^-] + \sum_{\mathbf{A}_i \in \mathfrak{A}} \sum_{n=0}^{n_{a,i}} (z_{a,i} + n) \left[H_n A_i^{z_{a,i}+n} \right] + \sum_{S_i \in \mathbf{S}} z_{s,i} \left[S_i^{z_{s,i}} \right], \tag{10}$$

$$K_{a(n_{a,i}-n+1),i} = \frac{[H^+] \left[H_{n-1} A_i^{z_{a,i}+n-1} \right]}{\left[H_n A_i^{z_{a,i}+n} \right]} \frac{\gamma_1 \gamma_{z_{a,i}+n-1}}{\gamma_{z_{a,i}+n}}, \quad n = 1, ..., n_{a,i}, \quad \mathbf{A}_i \in \mathfrak{A}, \tag{11}$$

where K_a is the acid dissociation constant and γ_z is the activity coefficient for z-charged ions. Activity is applied instead of concentration in the acid/base equilibria due to the high ionic strength I in the fermentation conditions, and is modeled by the Davis equation (Davies, 1962):

$$\gamma_z = 10^{-Az^2 \left(\frac{\sqrt{I}}{1+\sqrt{I}} - 0.3I \right)}, \tag{12}$$

$$I = \frac{1}{2} \left(\sum_{\mathbf{A}_i \in \mathfrak{A}} \sum_{n=0}^{n_{a,i}} (z_{a,i} + n)^2 \left[H_n A_i^{z_{a,i}+n} \right] + \sum_{S_i \in \mathbf{S}} c_{s,i}^2 \left[S_i^{z_{s,i}} \right] \right), \tag{13}$$

where the parameter A depends on temperature T and dielectric constant ϵ as $A = 1.82 \times 10^6 (\epsilon T)^{-3/2}$ and $A \approx 0.51$ for water at 25°C.

The precipitate species are described as

$$P_i \overset{K_{sp,i}}{\rightleftarrows} \sum_{n=1}^{n_{p,i}} s_{n,i} P_{n,i}, \quad P_i \in \mathbf{P}, \tag{14}$$

where $\mathbf{P}$ is the set of the precipitate species, K_{sp} is the solubility product constant, n_p is the number of different ion species forming the precipitate, and s is the stoichiometric coefficient. Then the concentration of precipitates and the precipitation (and dissolution) rate can be described as

$$\frac{d[P_i]}{dt} = \frac{F_{in}}{V_L} ([P_i]_{in} - [P_i]) + r_{P,i}, \tag{15}$$

$$r_{P,i} = \begin{cases} 0, & \text{if } Q_{sp,i} < K_{sp,i} \text{ and } [P_i] = 0, \\ k_i(Q_{sp,i} - K_{sp,i}), & \text{otherwise}, \end{cases} \tag{16}$$

where Q_{sp} is the reaction quotient and k is the reaction rate constant.

3.4 | Gas phase

The remaining reactor volume comprises a volume of the homogeneously dispersed bubbles and volume of the headspace gas above the liquid medium. These two gas phases can be modeled as stirred tanks connected in series, but combining the gas together as a sole gas phase is a reasonable approximation. Gas transfer between headspace gas and medium is negligible compared to gas transfer between bubbles and medium because of order-of-magnitude smaller surface area, resulting in negligible difference between headspace gas and bubbles (de Jonge et al., 2014). Then for the gas phase, differential balances for mole fraction of oxygen O and carbon dioxide C including water W are

$$\frac{dx_i}{dt} = \frac{p_{in} Q_{in} T}{p V_G T_{in}} \left(x_{i,in} - \frac{1 - x_{W,in}}{1 - x_W} x_i \right) + \frac{V_L RT}{p V_G} \left(\frac{x_i}{1 - x_W} \sum_{j=O,C} j TR - i TR \right), \tag{17}$$

where x is the mole fraction, the subscripts i refer to the components, Q is the gas flow rate, V_G is the total gas volume, p is the pressure, T is the temperature, and i is the gas transfer rate.

Oxygen transfer is especially important for system with *P. pastoris* because of the high oxygen uptake rate at high densities of cells. The bioreactor is designed to diffuse oxygen from sparged bubbles into the bulk liquid. The oxygen transfer rate (OTR) can be calculated from the differential balance for the concentration of dissolved oxygen (DO) and a scaling analysis that shows that terms with DO can be neglected due to the low solubility of oxygen in water: $OTR \approx q_O X$, where q_O is the specific oxygen uptake rate.

Carbon dioxide produced from the cell metabolism has a net flux in the opposite direction of oxygen, from the bulk liquid to sparged bubbles. The carbon dioxide transfer rate (CTR) can be calculated from the differential balance for the concentration of dissolved carbon dioxide (DC):

$$CTR = -q_C X + r_C + \frac{dDC}{dt} - \frac{F_{in}}{V_L} (DC_{in} - DC), \tag{18}$$

where r_C is the reaction rate from dissolved carbon dioxide to carbonate species described by the chemical reactions: $CO_2 + H_2O \rightleftarrows H_2CO_3, CO_2 + OH^- \rightleftarrows HCO_3^-$

The approximation that was used for oxygen is not applicable for carbon dioxide because of its high solubility in water. DC will be generally higher than the saturated concentration DC^* since the concentration gradient is needed for carbon dioxide transfer. Considering that CTR and OTR are usually in the same order of magnitude (Royce & Thornhill, 1991),

$$(DC - DC^*) \sim \frac{k_L a_O}{k_L a_C} (DO^* - DO), \tag{19}$$

where $k_L a$ is the volumetric mass transfer coefficient. Because the values of $k_L a_O$ and $k_L a_C$ are of the same order of magnitude (Royce & Thornhill, 1991), scaling analysis similar to what was applied to the oxygen shows that DC is close to its saturated value: $DC \approx DC^* = p x_C / H_C$, where H is the Henry's law constant. Then
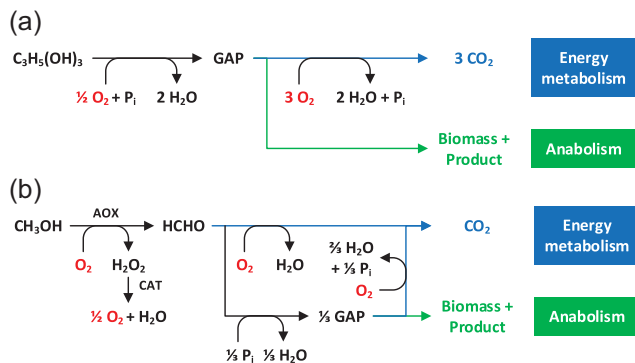
FIGURE 1 (a) Glycerol metabolism and (b) methanol metabolism of *Pichia pastoris*. AOX, alcohol oxidase; CAT, catalase; GAP, glyceraldehyde-3-phosphate [Color figure can be viewed at wileyonlinelibrary.com]

$$CTR \approx -q_C X + r_C + \frac{p}{H_C} \frac{dx_C}{dt}. \tag{20}$$

For oxygen uptake rate, Jahic et al. (2002) proposed a metabolic flux model based on the division of metabolic flux into anabolism and energy metabolism. A modification of their model to have a more detailed carbon balance and reaction network is

$$q_O = \frac{q_C}{C_S} Y_{O/S,\mathrm{an}} + \left(q_S - \frac{q_C}{C_S}\right) Y_{O/S,\mathrm{en}}, \tag{21}$$

where $Y_{O/S,\mathrm{an}}$ is the oxygen yield coefficient for anabolism, and $Y_{O/S,\mathrm{en}}$ is the oxygen yield coefficient for energy metabolism. The metabolic fluxes of glycerol and methanol are simplified in the reaction networks (see Figure 1), which also specify the oxygen yield coefficients.

3.5 | Maximum likelihood estimation and sensitivity analysis

Most of the parameters in the bioreactor model are taken from the literature; this section describes the procedure for the estimation of the model parameters that were fit to experimental data. Maximum likelihood (ML) estimation is commonly applied in literature assuming that there exists constant parameter set θ for all experiments (Ashyraliyev et al., 2009). Then the model equation for the system can be determined by

$$y_i = x_i + \epsilon_{m_i} = F(u_i; \theta) + \epsilon_{m_i}, \quad \epsilon_{m_i} \sim N(0, V_{m_i}), \quad i = 1, ..., d, \tag{22}$$

where y is the vector of measured states, x is the vector of states, u is the vector of inputs, ϵ_m is the vector of measurement error which is a white noise vector with covariance matrix V_m, the subscript i refers to each experimental data set, and d is number of experimental data sets. Then the maximum likelihood estimate is given as

$$\theta^* = \arg\min_{\theta} \sum_{i=1}^{d} (y_i - F(u_i; \theta))^{\top} V_{m_i}^{-1} (y_i - F(u_i; \theta)). \tag{23}$$

In this study, however, the parameters were estimated assuming that there exists a distribution of parameter sets between different experiments:

$$y_i = x_i + \epsilon_{m_i} = F(u_i; \theta_i) + \epsilon_{m_i}, \quad \theta_i = \theta_0 + \epsilon_{\theta_i}, \quad \epsilon_{\theta_i} \sim N(0, V_\theta),$$
$$\epsilon_{m_i} \sim N(0, V_{m_i}), \quad i = 1, ..., d, \tag{24}$$

where θ_i is constant parameter set for each of the experiment. By linearization of the model, the covariance matrix of the measured states can be expressed as

$$V_{y_i} \approx \frac{\partial x_i}{\partial \theta} V_\theta \frac{\partial x_i}{\partial \theta}^{\top} + V_{m_i}, \tag{25}$$

and then the maximum likelihood estimation can be formulated as

$$\arg\min_{\theta_0, V_\theta} \sum_{i=1}^{d} \ln \det(V_{y_i}) + (y_i - F(u_i; \theta_0))^{\top} V_{y_i}^{-1} (y_i - F(u_i; \theta_0)). \tag{26}$$

Sensitivities of the states with respect to the parameters are computed via sensitivity equations:

$$\frac{dx(t)}{dt} = f(x(t), u(t); \theta), \tag{27}$$

$$\frac{d}{dt}\left(\frac{\partial x}{\partial \theta}\right) = \frac{\partial}{\partial \theta}\left(\frac{dx}{dt}\right) = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial \theta}. \tag{28}$$

4 | RESULTS AND DISCUSSION

4.1 | pH and precipitation

Precipitation caused by medium components is a common problem in cell culture media and was observed in bioreactor runs when base ammonium hydroxide was added for pH adjustment before the inoculation. Concentration measurements of media components before and after pH adjustment indicated that the ammonium and phosphate ions are included in the precipitates (data not shown). This result suggests magnesium ammonium phosphate ($MAP, MgNH_4PO_4$) as the candidate precipitates.

For the pH model with precipitation, pH of the rich defined medium was measured as base 10% (v/v) ammonium hydroxide was added (Figure 2). Model prediction of pH was shifted to match initial pH data to deal with run-to-run variability of measured initial pH. The acid dissociation constants were taken from the literature (Kern, 1960; Lide, 2004). The solubility product of MAP was estimated with least squares ($pK_{sp} = 12.24$) and the estimated value reasonably agreed with values from the literature (Aage et al., 1997; Ohlinger et al., 1998; Snoeyink & Jenkins, 1980). The differences may be due to the different activity models and considered chemical species. The pH measurement data of rich defined medium at different base concentrations were well described by the pH model with precipitation, with a root-mean-squared error (RMSE) of 0.042 which is within the accuracy of the pH measurement (Rosemount Analytical, 2010). It was also observed in the experiment that the precipitation occurred when the volume of base added was 0.3% to 0.4% of volume of the rich defined medium, which also agreed with the model prediction.
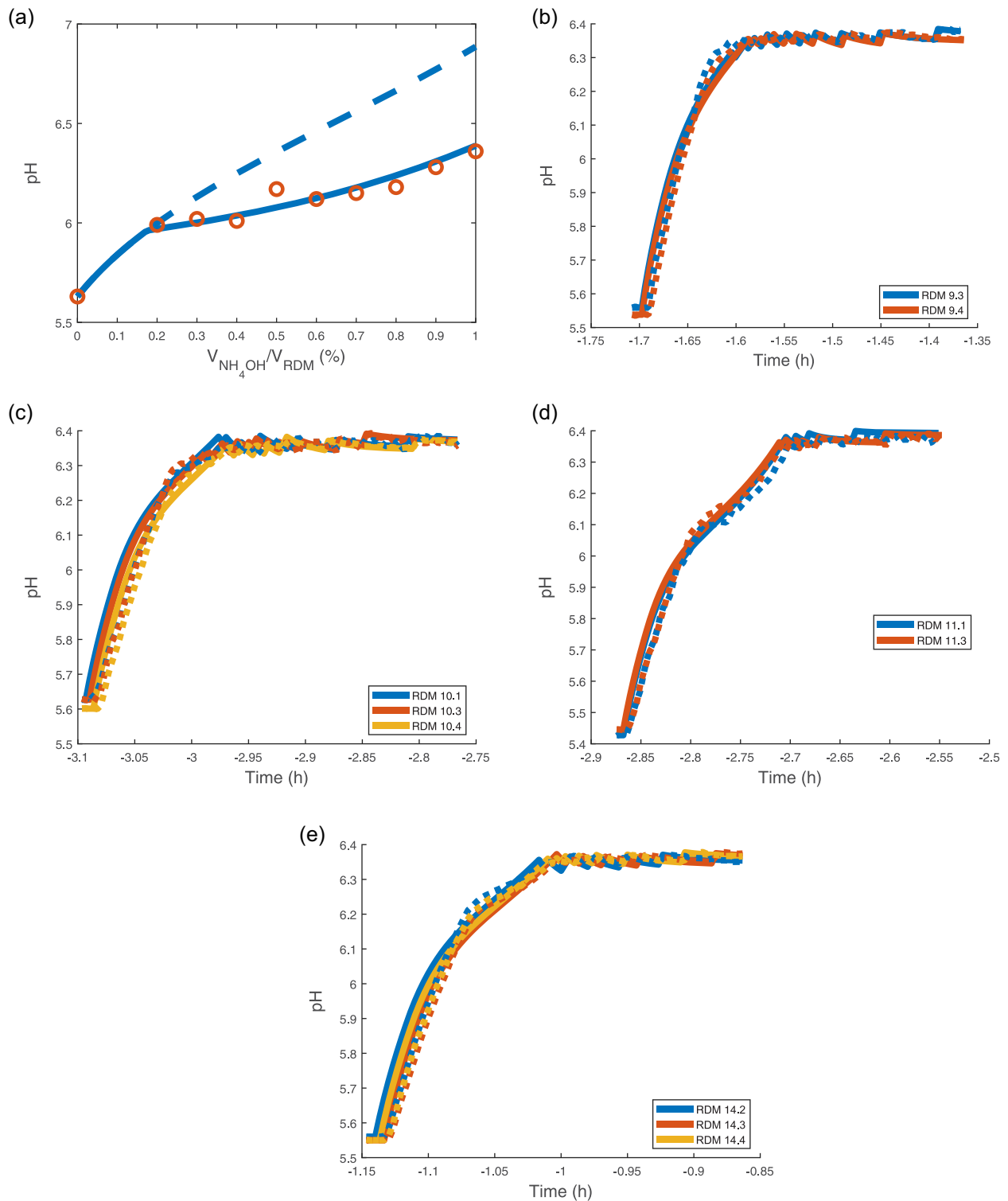
FIGURE 2 (a) Model predictions with precipitation (solid line) and without precipitation (dashed line) and experimental data (open circles) for pH as base 10% (v/v) ammonium hydroxide is added to the rich defined medium. Model prediction with precipitation (solid lines) and experimental data (dotted lines) during the pH adjustment before the inoculation for (b) RDM 9 ($N_{T,in}/N_{T,in}^{ref} = 1.12$), (c) RDM 10 ($N_{T,in}/N_{T,in}^{ref} = 1.04$), (d) RDM 11 ($N_{T,in}/N_{T,in}^{ref} = 1.04$), and (e) RDM 14 ($N_{T,in}/N_{T,in}^{ref} = 0.87$), where $N_{T,in}$ is the corrected concentration of base ammonium hydroxide and $N_{T,in}^{ref}$ is the intended concentration [Color figure can be viewed at wileyonlinelibrary.com]

Base ammonium hydroxide used for pH control in the experiments can have a different concentration than was intended. Lab solutions that are purchased in a concentrated form usually report lower bounds of concentration in the label (e.g., ≥25%). Furthermore, ammonia gas liberates over time, which decreases the concentration. To take this phenomenon into account, concentration of base ammonium hydroxide was corrected based on the pH measurements before inoculation (Figure 2). The concentration of the base used for pH control was initially higher than the intended concentration but the concentration decreased over time as ammonia gas evaporated. The estimated reaction rate constant for precipitation was $k = 1.3 \times 10^{11} \, \text{M}^{-2} \, \text{s}^{-1}$.

4.2 | Cellular metabolism

Measurements of biomass and total protein concentration were used to validate substrate consumption, growth, and production in the macroscopic bioreactor model. Bioreactor model predictions are compared with experimental data for the biomass and total protein concentration in Figure 3. Parameters required for the growth and substrate consumption were taken from the literature (Jahic et al., 2002). The biomass concentration data with different operating conditions (pH) and different strains (products) were well described by the model with the reported parameters, with an RMSE of 3.1 g/L, which is within 5% to 10% of the measured biomass.

The parameters for the total protein production estimated with maximum likelihood using pH values from the measurements were $\alpha_{P,g} = 23.2 \pm 1.4 \, \text{mg/g}$, $\alpha_{P,m} = 69.0 \pm 15.3 \, \text{mg/g}$, $pK_{a,E} = 3.42 \pm 0.01$. The total protein production during the glycerol period was consistent with most of the proteins produced being host cell proteins, with a RMSE of 49 mg/L, which is within 5% to 8% of the measured total protein production. Run-to-run variability of total protein production occurred mainly during the methanol period, which is associated with recombinant protein production (RMSE of 250 mg/L) and may be due to unmeasured cell properties such as viability and copy number variation.

The pH dependence of the total protein production is evident from total protein concentration for different pH conditions during the glycerol period and the trend of total protein concentration after
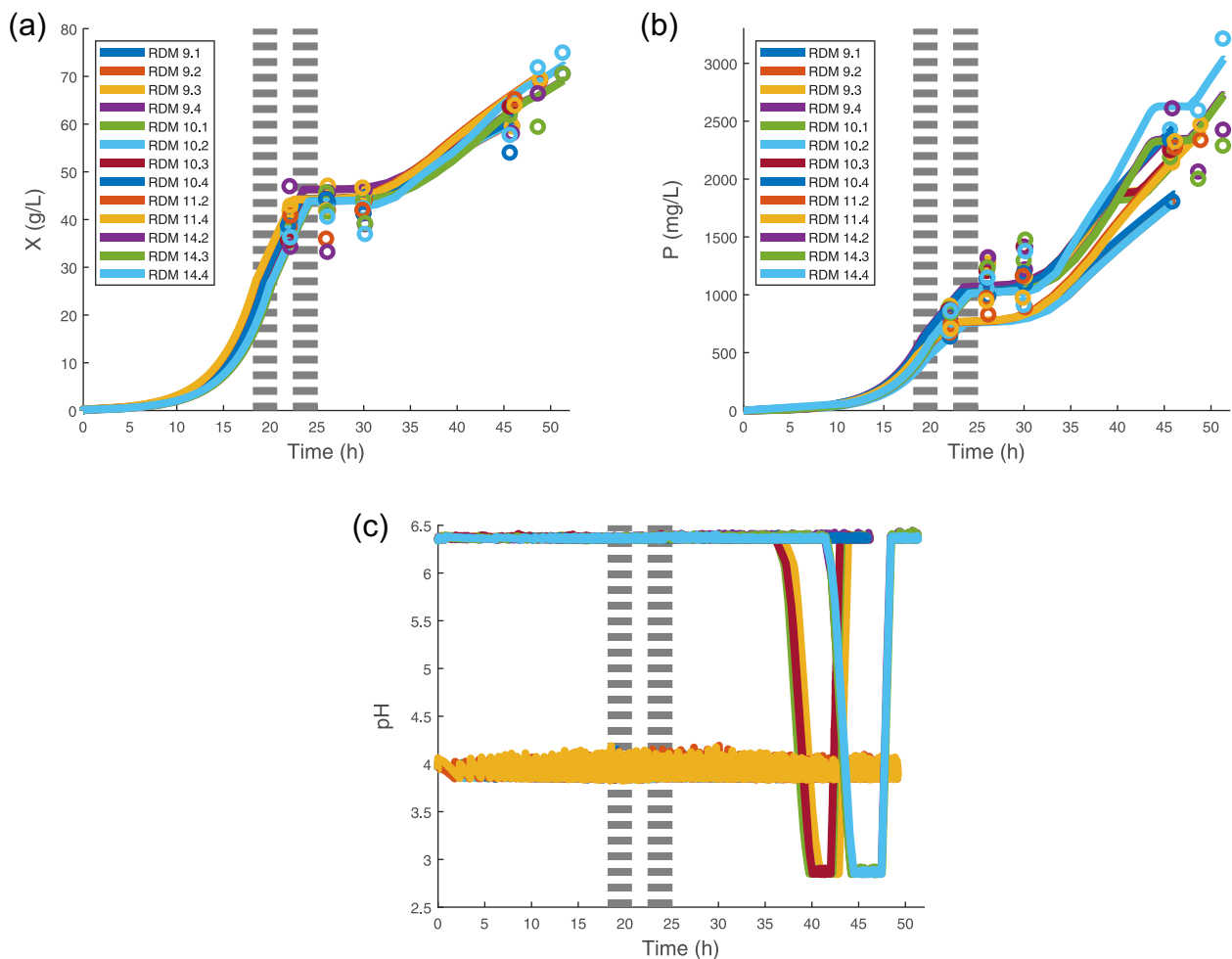
FIGURE 3 Model predictions (solid lines) and experimental data (open circles) of (a) biomass and (b) total protein concentration, and (c) pH measurement data of *Pichia pastoris*. Vertical dotted lines represent times when the operation mode changed [Color figure can be viewed at wileyonlinelibrary.com]

pH shift during the methanol period. This dependence suggests that lowering pH for inactivating proteases can negatively affect the quantitative productivity of recombinant proteins.

4.3 | Medium components

The metabolite concentrations in fermentation supernatant were measured in bioreactor runs to validate the modeling of media consumption. The bioreactor model predictions are compared with experimental measurements of the concentrations of ammonium, phosphate, and potassium ions in Figure 4. For ammonium and phosphate, data with low pH conditions were used because precipitation occurs at higher pH conditions. Elemental compositions of biomass were estimated using the maximum likelihood method, and estimated values reasonably agreed with values from the literature (Carnicer et al., 2009; Tavasoli et al., 2017; van Eunen et al., 2010) (Table 2). The parameter for specific nitrogen uptake rate was also estimated: $\alpha_{N,m} = 0.133 \pm 0.007$ g/g.

TABLE 2 Maximum likelihood estimates and literature-reported values of parameters for elemental compositions of biomass

Parameters	This study	Carnicer et al. (2009)	Tavasoli et al. (2017)	van Eunen et al. (2010)
$C_{X,0}$ (mg/g)	413 ± 16	443 ± 22	446	
$N_{X,0}$ (mg/g)	90.4 ± 1.5	75 ± 11	79	
P_X (mg/g)	21.8 ± 1.2		25.7	20 ± 1
K_X (mg/g)	20.0 ± 1.8		29.2	28 ± 2

The potassium and phosphate data were well described by the model (RMSE of 6.5 and 4.0 mM, respectively), which demonstrates that constant elemental composition was a reasonable assumption. Ammonium data have more run-to-run variability compared to other medium components (RMSE of 13 mM). Run-to-run variability of ammonium consumption occurred mainly during the methanol period while the concentration of intracellular nitrogen is increasing due to recombinant protein production.
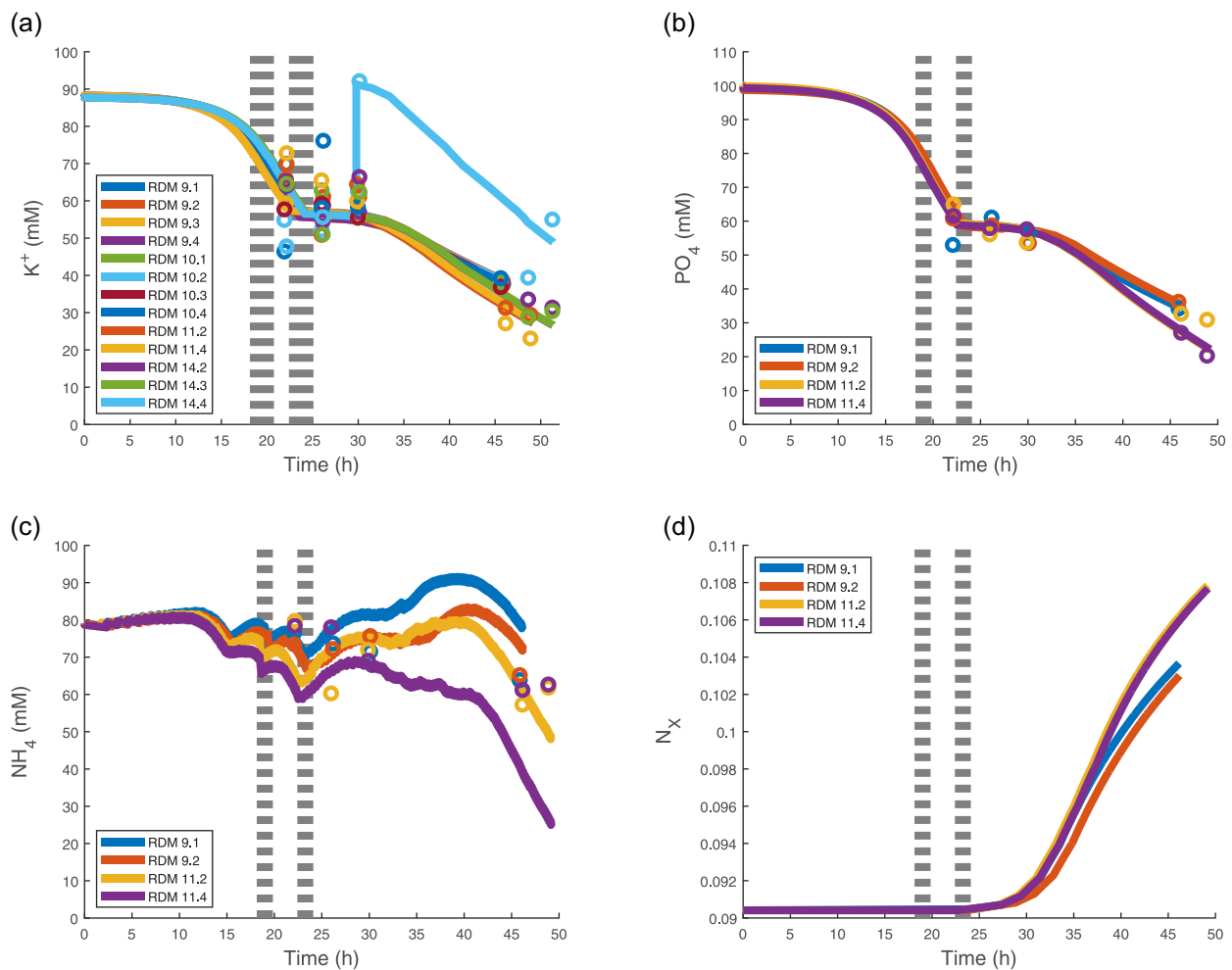
FIGURE 4 Model predictions (solid lines) and experimental data (open circles) of (a) phosphate, (b) potassium, and (c) ammonium concentration and (d) model predictions of nitrogen composition of biomass. Vertical dotted lines represent times when the operation mode changed [Color figure can be viewed at wileyonlinelibrary.com]

4.4 | Gas phase

Lastly, off-gas data were used to validate the modeling of oxygen uptake and carbon dioxide evolution. Figure 5 compares bioreactor model predictions with experimental measurements of off-gas compositions of oxygen and carbon dioxide. The Henry's law constant and carbonate species reaction constants were taken from the literature (Kern, 1960; Sander, 2015). Inlet air and oxygen were dry ($x_{W,\text{in}} \approx 0$) and the mole fraction of water vapor in the reactor was assumed to be $x_W = 0.01$ based on the steady-state off-gas oxygen concentration before inoculation. Carbon composition of biomass during glycerol period was estimated using maximum likelihood and the estimated values reasonably agreed with values from the literature (Carnicer et al., 2009; Tavasoli et al., 2017) (Table 2). The parameter for specific carbon uptake rate for anabolism was also estimated: $\alpha_{C,m} = 0.594 \pm 0.013$ g/g.

The off-gas data of oxygen and carbon dioxide were well described by the model (RMSE of 0.76% and 0.73%, respectively), which demonstrates that metabolic flux model for the oxygen uptake rate is a reasonable simplification. A sudden increase in both oxygen uptake rate and carbon dioxide evolution rate were observed in all experiments at the beginning of the methanol induction, however, causing disagreement between model prediction and experimental data during that time window. This finding suggests that the actual cell metabolism was more focused on energy metabolism than what was predicted from the model and may be due to dramatic changes in cell metabolism for adaptation to methanol.

4.5 | Parameter distributions and sensitivity analysis

The developed macroscopic bioreactor model with pH and precipitation was implemented and compared with the entire experimental data set of bioreactor runs (Figure 6). Elemental compositions of biomass for the remaining medium components were taken from the literature (van Eunen et al., 2010). Including precipitation into the macroscopic bioreactor model resulted in good agreement of model predictions and experimental data of medium components and pH. Model prediction for medium components with pH and precipitation can be applied for improving the chemically defined medium by minimizing the concentration of components needed while meeting cellular requirements, which is important for extended operations in perfusion mode. For example, phosphate was the limiting component in the current medium for bioreactor run RDM 14.2. The model predictions that extended fed-batch operation would result in phosphate depletion at 60.5 ± 2.5 hr. However, if monopotassium phosphate (KH_2PO_4) in the medium was increased from 12 to 15 g/L, the PO_4 concentration would shift up to the dashed line in Figure 6, and the model predicts that phosphate depletion would be delayed to 71.2 ± 3.2 hr.

Furthermore, the effects of the parameter distributions estimated by the maximum likelihood method on the predicted variables were quantified. Their probability distribution functions were constructed by Monte Carlo simulation with 10,000 random parameter sets sampled from the parameter distributions (Figure 7). The probability distribution functions capture the run-to-run variability in experimental data, especially for total protein and ammonium concentration.

Sensitivity analysis identifies key parameters for the states of interest in the macroscopic bioreactor model (Figure 8). The total protein concentration was sensitive to α_P of each substrate during the period of its consumption, while being sensitive to $pK_{a,E}$ only when operating in low pH conditions. For medium components and off-gas components, key parameters are those related to elemental composition of biomass. Medium components are also sensitive to inlet acid and base concentrations because of their high concentration. This result indicates that the acid and base concentrations should be accurately identified for the model application.

5 | CONCLUSION

This article constructs an extensive macroscopic bioreactor model for *P. pastoris* producing recombinant proteins in defined medium. The model describes substrates, biomass, total protein, other medium components, and off-gas components based on species and elemental balances. The model also describes pH based on overall charge balance, acid/base equilibria, and activity coefficients to describe total protein production and precipitation of medium components. The distributions of uncertain model parameters were estimated using the maximum likelihood method to model the extent of run-to-run variability.

Model predictions from the extensive macroscopic bioreactor model well describe experimental data with different operating conditions. The pH dependence of the total protein production suggests that the low pH condition for inactivating proteases may decrease the quantitative productivity of recombinant proteins. The model predicts that the specific total production rate drops from pH 6.5 by 21% to pH 4, 45% to pH 3.5, and 72% to pH 3. Elemental compositions of biomass are identified as the most sensitive parameters of the medium and off-gas components. Precipitation of medium components is also critical for accurate prediction of concentration of medium components and pH. The model of medium components with pH and precipitation can be used for optimizing a chemically defined medium, especially for extended operations in perfusion mode. The metabolic flux model well describes oxygen uptake rate and carbon dioxide evolution rate and provides an understanding of the state of cell metabolism from the off-gas measurement data.

The probability distribution of model predictions from the parameter distribution quantifies the run-to-run variability observed in the experimental data. The uncertainty description in extensive macroscopic bioreactor model identifies large variability of the model parameters related to the total protein production and nitrogen consumption during the
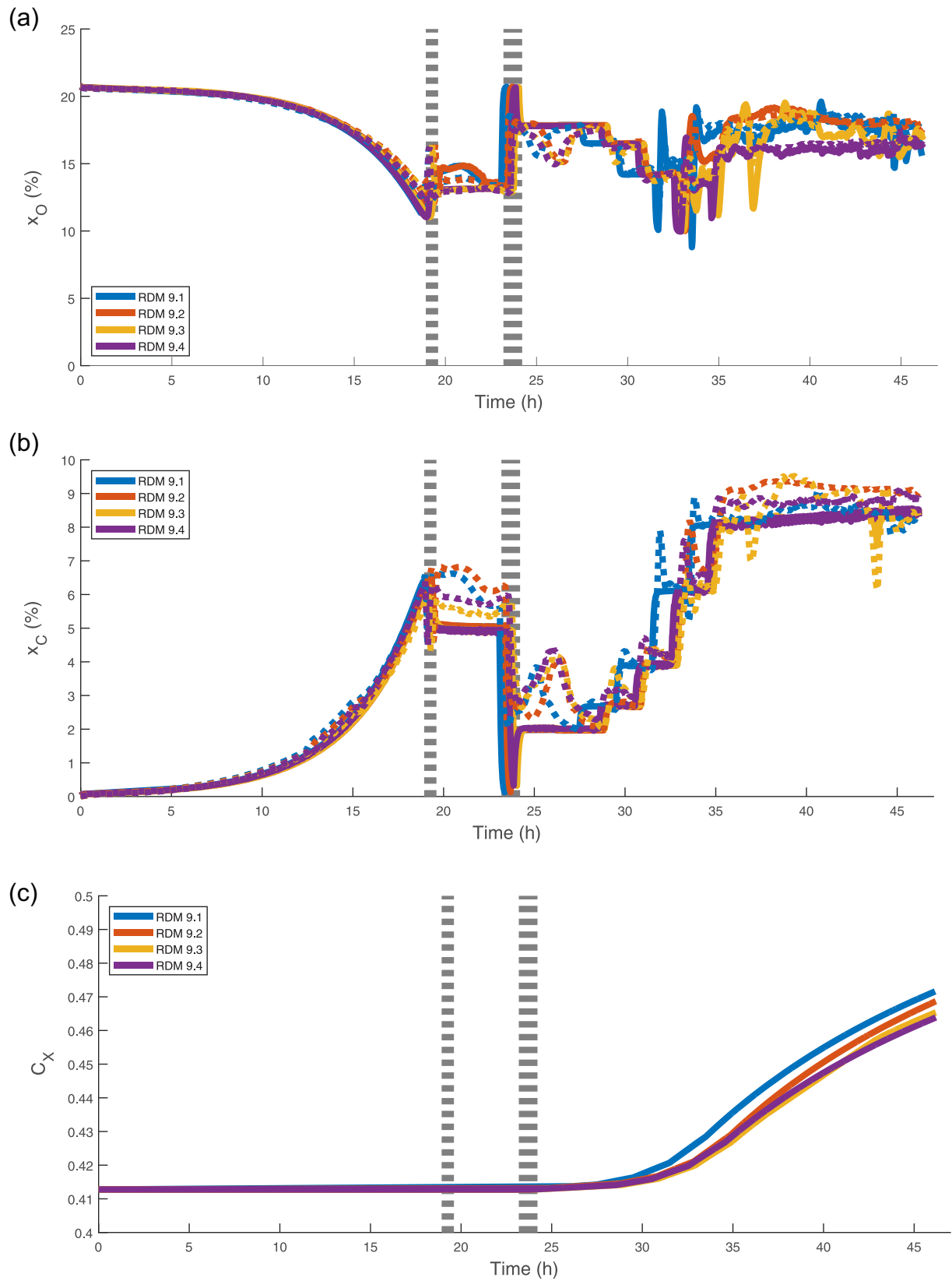
FIGURE 5 (a) Model predictions (solid lines) and experimental data (dotted lines) of the off-gas composition of oxygen and (b) carbon dioxide and (c) model predictions of carbon composition of biomass. Bioreactor runs RDM 9 are shown for demonstration. Vertical dotted lines represent times when the operation mode changed [Color figure can be viewed at wileyonlinelibrary.com]
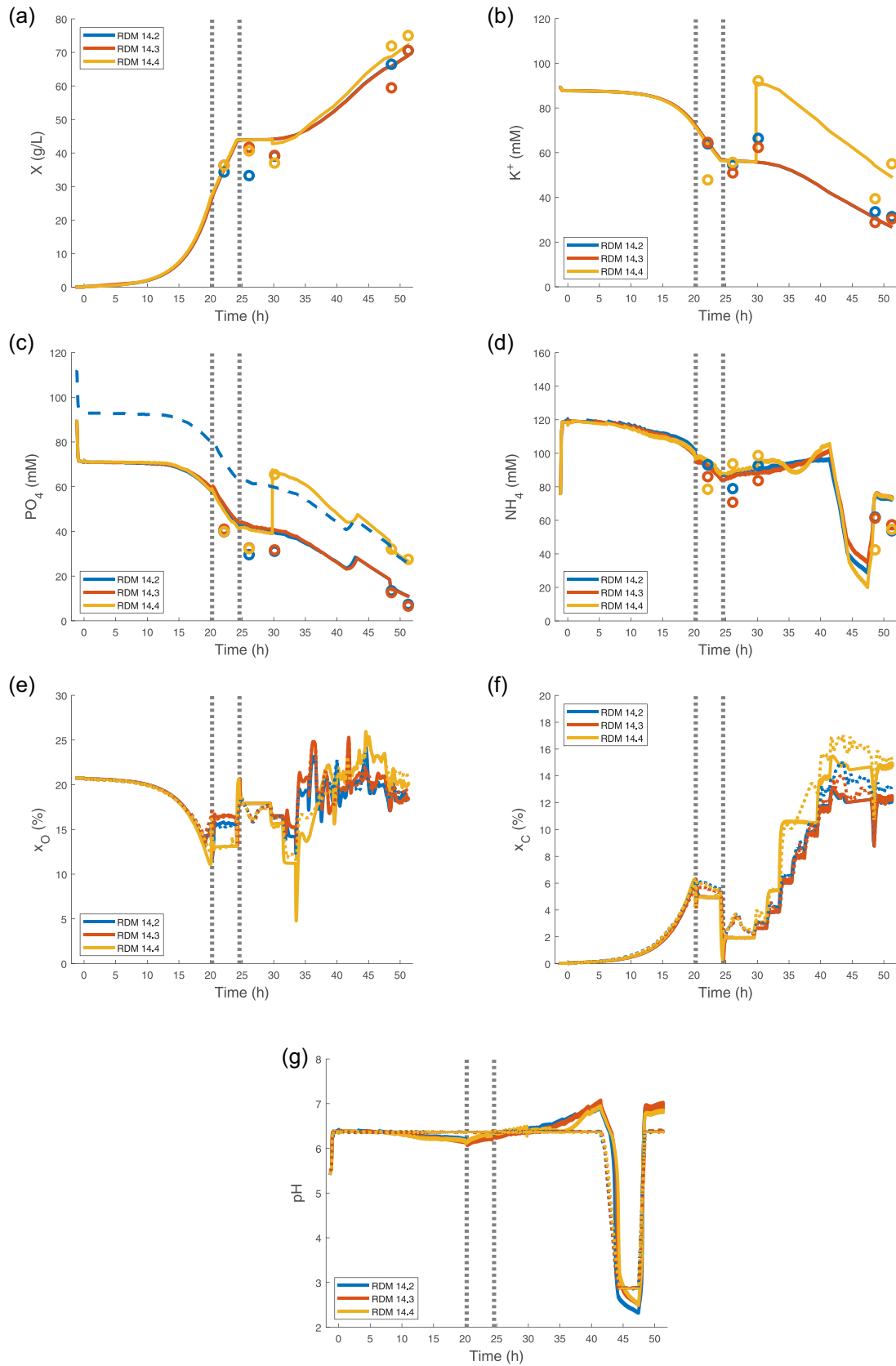
FIGURE 6 Model predictions (solid lines) and experimental data (open circles or dotted lines) of bioreactor runs RDM 14. The vertical dotted lines represent times when the operation mode changed. The additional dashed line in the PO_4 plot shows the model prediction if monopotassium phosphate (KH_2PO_4) in the medium was increased from 12 to 15 g/L [Color figure can be viewed at wileyonlinelibrary.com]
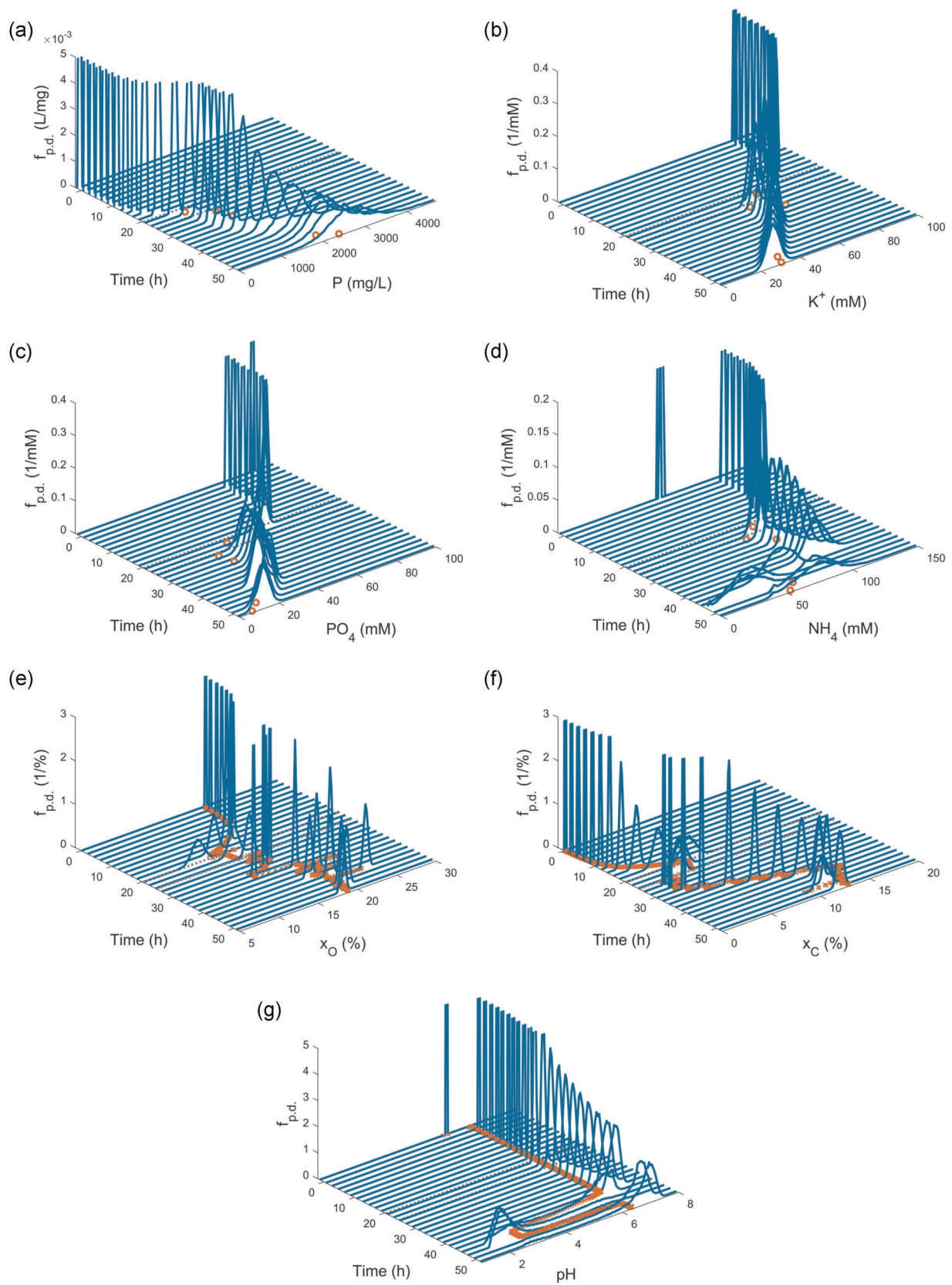
FIGURE 7 Probability distribution functions from Monte Carlo simulations (solid lines) and experimental data (open circles or crosses) of bioreactor run RDM 14.2 [Color figure can be viewed at wileyonlinelibrary.com]
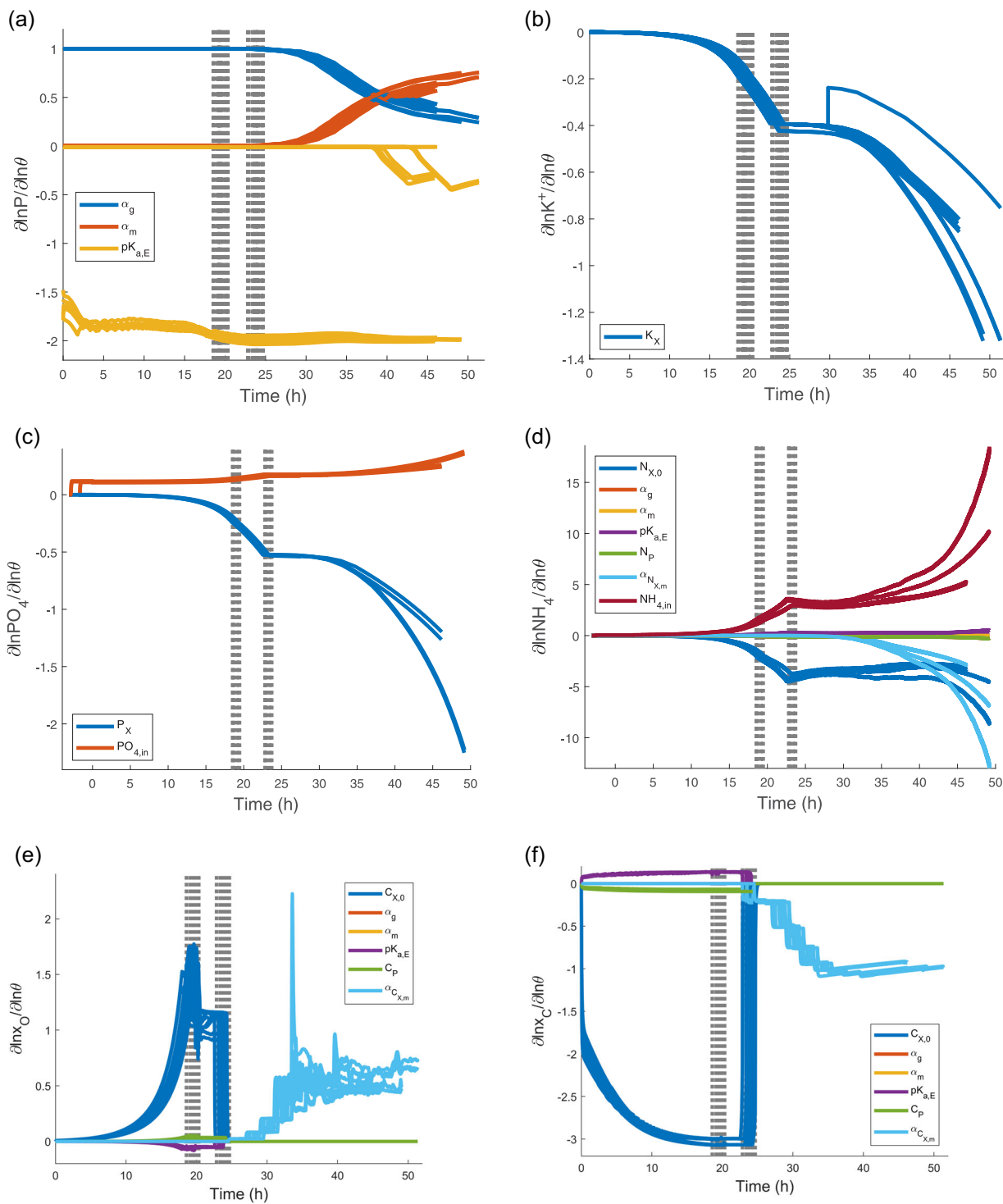
FIGURE 8 Sensitivities of states of interest in the macroscopic bioreactor model (concentration of total protein, medium components, and off-gas components) with respect to related parameters [Color figure can be viewed at wileyonlinelibrary.com]

methanol phase. This provides guidance as to which aspects of cellular metabolism should be the focus of additional experimental studies.

AUTHOR CONTRIBUTIONS

Conceptualization: Moo Sun Hong, Andrew J. Maloney, Richard D. Braatz; *Modeling*: Moo Sun Hong. *Experiments*: M. Lourdes Velez-Suberbie. *Cell

REFERENCES

Aage, H. K., Andersen, B. L., Blom, A., & Jensen, I. (1997). The solubility of struvite. *Journal of Radioanalytical and Nuclear Chemistry*, 223(1–2), 213-215. https://doi.org/10.1007/bf02223387

Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J. A., & Blom, J. G. (2009). Systems biology: Parameter estimation for biochemical models. *The FEBS Journal*, 276(4), 886-902. https://doi.org/10.1111/j.1742-4658.2008.06844.x

Barrigon, J. M., Valero, F., & Montesinos, J. L. (2015). A macrokinetic model-based comparative meta-analysis of recombinant protein production by Pichia pastoris under AOX1 promoter. *Biotechnology and Bioengineering*, 112(6), 1132-1145. https://doi.org/10.1002/bit.25518

Carnicer, M., Baumann, K., Töplitz, I., Sánchez-Ferrando, F., Mattanovich, D., Ferrer, P., & Albiol, J. (2009). Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels. *Microbial Cell Factories*, 8, 65. https://doi.org/10.1186/1475-2859-8-65

Çelik, E., Çalık, P., & Oliver, S. G. (2009). A structured kinetic model for recombinant protein production by Mut^+ strain of *Pichia pastoris*. *Chemical Engineering Science*, 64(23), 5028-5035. https://doi.org/10.1016/j.ces.2009.08.009

Cereghino, J. L., & Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiology Reviews*, 24(1), 45-66. https://doi.org/10.1111/j.1574-6976.2000.tb00532.x

Craven, S., Whelan, J., & Glennon, B. (2014). Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. *Journal of Process Control*, 24(4), 344-357. https://doi.org/10.1016/j.jprocont.2014.02.007

Davies, C. W. (1962). *Ion association*. Butterworths.

de Jonge, L. P., Heijnen, J. J., & van Gulik, W. M. (2014). Reconstruction of the oxygen uptake and carbon dioxide evolution rates of microbial cultures at near-neutral pH during highly dynamic conditions. *Biochemical Engineering Journal*, 83, 42-54. https://doi.org/10.1016/j.bej.2013.12.002

Enfors, S.-O. (2011). *Fermentation process engineering*. Royal Institute of Technology.

Jahic, M., Rotticci-Mulder, J., Martinelle, M., Hult, K., & Enfors, S.-O. (2002). Modeling of growth and energy metabolism of *Pichia pastoris* producing a fusion protein. *Bioprocess and Biosystems Engineering*, 24(6), 385-393. https://doi.org/10.1007/s00449-001-0274-5

Kapteyn, J. C., TerRiet, B., Vink, E., Blad, S., De Nobel, H., Van Den Ende, H., & Klis, F. M. (2001). Low external pH induces HOG1-dependent changes in the organization of the Saccharomyces cerevisiae cell wall. *Molecular Microbiology*, 39(2), 469-480. https://doi.org/10.1046/j.1365-2958.2001.02242.x

Kern, D. M. (1960). The hydration of carbon dioxide. *Journal of Chemical Education*, 37(1), 14-23. https://doi.org/10.1021/ed037p14

Lide, D. R. (2004). *Crc handbook of chemistry and physics* (Vol. 85). CRC Press.

Matthews, C. B., Kuo, A., Love, K. R., & Love, J. C. (2018). Development of a general defined medium for *Pichia pastoris*. *Biotechnology and Bioengineering*, 115, 103-113. https://doi.org/10.1002/bit.26440

Monod, J. (1949). The growth of bacterial cultures. *Annual Review of Microbiology*, 3, 371-394. https://doi.org/10.1146/annurev.mi.03.100149.002103

Ohlinger, K. N., Young, T. M., & Schroeder, E. D. (1998). Predicting struvite formation in digestion. *Water Research*, 32(12), 3607-3614. https://doi.org/10.1016/S0043-1354(98)00123-7

Plantz, B. A., Sinha, J., Villarete, L., Nickerson, K. W., & Schlegel, V. L. (2006). *Pichia pastoris* fermentation optimization: Energy state and testing a growth-associated model. *Applied Microbiology and Biotechnology*, 72(2), 297-305. https://doi.org/10.1007/s00253-005-0271-8

Ramaswamy, S., Cutright, T. J., & Qammar, H. K. (2005). Control of a continuous bioreactor using model predictive control. *Process Biochemistry*, 40(8), 2763-2770. https://doi.org/10.1016/j.procbio.2004.12.019

Ren, H. T., Yuan, J. Q., & Bellgardt, K. -H. (2003). Macrokinetic model for methylotrophic *Pichia pastoris* based on stoichiometric balance. *Journal of Biotechnology*, 106(1), 53-68. https://doi.org/10.1016/j.jbiotec.2003.08.003

Rosemount Analytical. (2010). *Theory and practice of pH measurement*. Emerson Process Management. (PN 44-6033/rev. D).

Royce, P. N., & Thornhill, N. F. (1991). Estimation of dissolved carbon dioxide concentrations in aerobic fermentations. *AIChE Journal*, 37(11), 1680-1686. https://doi.org/10.1002/aic.690371111

Sander, R. (2015). Compilation of Henryas law constants (version 4.0) for water as solvent. *Atmospheric Chemistry & Physics*, 15(8), 4399-4981. https://doi.org/10.5194/acp-15-4399-2015

Snoeyink, V. L., & Jenkins, D. (1980). *Water chemistry*. Wiley.

Surribas, A., Geissler, D., Gierse, A., Scheper, T., Hitzmann, B., Montesinos, J. L., & Valero, F. (2006). State variables monitoring by in situ multi-wavelength fluorescence spectroscopy in heterologous protein production by *Pichia pastoris*. *Journal of Biotechnology*, 124(2), 412-419. https://doi.org/10.1016/j.jbiotec.2006.01.002

Tavasoli, T., Arjmand, S., Siadat, S. O. R., Shojaosadati, S. A., & Lotfi, A. S. (2017). Enhancement of alpha 1-antitrypsin production in *Pichia pastoris* by designing and optimizing medium using elemental analysis. *Iranian Journal of Biotechnology*, 15(4), 224-231. https://doi.org/10.15171/ijb.1808

Tomáas-Gamisans, M., Ferrer, P., & Albiol, J. (2016). Integration and validation of the genomescale metabolic models of *Pichia pastoris*: A comprehensive update of protein glycosylation pathways, lipid and energy metabolism. *PLOS One*, 11(1), e0148031. https://doi.org/10.1371/journal.pone.0148031

van Eunen, K., Bouwman, J., Daran-Lapujade, P., Postmus, J., Canelas, A. B., Mensonides, F. I. C., Orij, R., Tuzun, I., van den Brink, J., Smits, G. J., van Gulik, W. M., Brul, S., Heijnen, J. J., de Winde, J. H., de Mattos, M. J. T., Kettner, C., Nielsen, J., Westerhoff, H. V., & Bakker, B. M. (2010). Measuring enzyme activities under standardized in vivo-like conditions for systems biology. *The FEBS Journal*, 277(3), 749-760. https://doi.org/10.1111/j.1742-4658.2009.07524.x

Velez-Suberbie, M. L., Morris, S. A., Kaur, K., Hickey, J. M., Joshi, S. B., Volkin, D. B., Bracewell, D. G., & Mukhopadhyay, T. K. (2020). Holistic process development to mitigate proteolysis of a subunit rotavirus vaccine candidate produced in Pichia pastoris by means of an acid pH pulse during fed-batch fermentation. *Biotechnology Progress*, e2966. https://doi.org/10.1002/btpr.2966