Distributed Modeling and Parameter Estimation of Influenza Virus Replication During Vaccine Production

R. Dürr [*,⋆] **T. Müller** [**] **B. Isken** [*] **J. Schulze-Horsel** [*]
U. Reichl [*,**] **A. Kienle** [*,**]

[*] *Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany*
[**] *Otto-von-Guericke University, Magdeburg, Germany*

Abstract: This contribution is concerned with population balance modeling of influenza virus replication in mammalian cell cultures. The cells are heterogeneous with respect to intracellular compounds like viral proteins. The amount of viral NP protein can be measured directly by means of flow cytometry. The corresponding degree of fluorescence is introduced as internal coordinate for a distributed deterministic modeling approach. The resulting model includes kinetic processes like infection, virus replication and release, apoptosis and cell death. It consists of three partial differential equations describing the distribution dynamics which are coupled to two differential equations that characterize the concentration of active and inactive virions in the medium. Kinetic parameters are determined from experimental data. The parameters are assumed to depend on the internal coordinate. The emerging infinite dimensional parameter estimation problem is translated to a finite dimension using a hermite spline representation of the distributed parameters. Hence the resulting inverse problem can be solved in a weighted nonlinear least squares framework. Spline approaches of different complexity are discussed and the estimation results are compared.

Keywords: Influenza vaccine production, population balance modeling, partial differential equations, inverse problems .

1. INTRODUCTION

In influenza vaccine production the use of permanent mammalian cell lines becomes more and more important. Besides sophisticated cell culture technologies and downstream processing methods, mathematical modeling plays a crucial role in improving production efficiency. Most notably for analysis, experimental design and optimization of the process, the benefit of combining extensive experiments with mathematical modeling approaches becomes apparent.

In our previous work, infection of equine influenza A virus in Madin-Darby-Canine-Kidney (MDCK) cell cultures was investigated in either deterministic (Müller et al., 2008) or stochastic modeling frameworks (Sidorenko et al., 2008a,b). More recently, focus was on population balance modeling of human influenza replication in MDCK cell cultures (Müller et al., 2011). Distribution dynamics are measured by means of fluorescence intensity, which is proportional to the intracellular amount of viral protein NP. Interesting new phenomena like transient multimodality and reversal of propagation direction could be observed.

Our contribution refers to the lab-scale influenza virus production as described by Schulze-Horsel et al. (2009). Adherent MDCK cells are cultivated on microcarriers in small scale bioreactors as depicted in Figure 1. At the beginning of the process the cells are inoculated with human influenza A virus seed. The uninfected cells become infected and start synthesizing viral components. Completely assembled virions are released into the medium and can infect uninfected cells. With progress of infection cells become apoptotic. Apoptosis is programmed cell death which can be activated by a large variety of external and internal stimuli, particularly by viral infection. It invariably leads to cell lysis and has major influence on the process productivity.

To adapt the model to the data an inverse problem has to be solved. As the kinetic parameters are not only constants

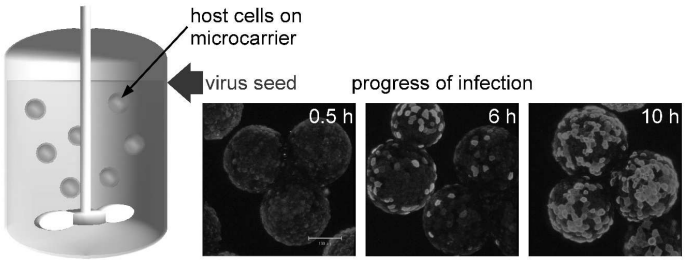

Fig. 1. Scheme of the process: Adherent MDCK cells growing on microcarriers and cells showing fluorescence due to intracellular accumulation of viral proteins

⋆ Correspondence to: R. Dürr, MPI for Dynamics of Complex Technical Systems, 39106 Magdeburg, Germany (e-mail: robert.duerr@ovgu.de)

but depend on the internal characteristics the inverse problem is set in an infinite dimension. In contrast to least-squares based parameter estimation in a finite dimension, which is extensively studied in literature (Walter and Pronzato, 1997), a relatively low number of publications deals with estimation techniques for infinite dimensional problems (Banks et al., 2011; Luzyanina et al., 2009).

2. MODEL FORMULATION

Heterogeneity of the cell population with respect to internal compounds like viral proteins gives rise to a distributed modeling approach with population balances. A degree of fluorescence $\varphi \in [\varphi_{min}, \varphi_{max}]$, which is proportional to the intracellular amount of viral protein NP is introduced as an internal coordinate. The fluorescence level of each cell is interpreted as degree of infection and can be measured by means of flow cytometry. The presented model comprises a system of three partial differential equations for description of the distribution dynamics. In addition, they are coupled to two ordinary differential equations that characterize virus particles in the medium. It is an extension of the model given by Müller et al. (2011).

On contact with active virions the uninfected cells U_c become infected with the rate constant k_{vi}. Due to medium exchange, limited space on microcarriers and fast progression of infection, growth and death of the uninfected cells are neglected.

$$\frac{\partial U_c(\varphi,t)}{\partial t} = -k_{vi}\, U_c(\varphi,t)\, V_{ac}(t) \tag{1}$$

Infected cells I_c originate from uninfected cells and start replicating and releasing virions. The first kinetic process increases and the latter decreases the degree of fluorescence. Both are summed up in a net rate coefficient $k_{net}(\varphi)$ that may depend on the fluorescence level and which is strictly positive in $\varphi \in [\varphi_{min}, \varphi_{max}]$. A time delay τ is considered to summarize intracellular processes that cause a lag between infection and virus replication and release. In the course of the infection more and more cells become apoptotic. This is accounted for by the rate of apoptosis $k_{apo}(\varphi)$ which is also considered to depend nonlinearly on the degree of fluorescence.

$$\begin{aligned}\frac{\partial I_c(\varphi,t)}{\partial t} = &- \frac{\partial\left\{k_{net}(\varphi)\, I_c(\varphi,t)\right\}}{\partial \varphi} - k_{apo}(\varphi)\, I_c(\varphi,t) \\ &+ \; k_{vi}\, U_c(\varphi, t-\tau)\, V_{ac}(t-\tau)\end{aligned} \tag{2}$$

After the infected cells have become apoptotic their replication ability is altered significantly. Its fluorescence degree is assumed to be generally decreasing which results in a negative net rate $k_{net,apo}(\varphi)$ in $\varphi \in [\varphi_{min}, \varphi_{max}]$. In addition, the apoptotic cells drop out of the process with lysis rate $k_{cd}(\varphi)$.

$$\begin{aligned}\frac{\partial A_c(\varphi,t)}{\partial t} = &- \frac{\partial\left\{k_{net,apo}(\varphi)\, A_c(\varphi,t)\right\}}{\partial \varphi} \\ &+ \; k_{apo}(\varphi)\, I_c(\varphi,t) - k_{cd}(\varphi)\, A_c(\varphi,t)\end{aligned} \tag{3}$$

Fully assembled virions are released to the medium by infected as well as apoptotic cells with the release rate k_{rel}. Due to errors in the virus replication processes, only a small fraction of released virus particles is able to infect uninfected cells. This fact is accounted for by the constant parameter P_{eff}. The so called active virions V_{ac} can become inactive with the constant degradation rate k_{deg}.

$$\begin{aligned}\frac{dV_{ac}(t)}{dt} = &- k_{vi}\, V_{ac}(t) \int_{\varphi} U_c(\varphi,t)\, d\varphi - k_{deg}\, V_{ac}(t) \\ &+ \; P_{eff}\, k_{rel} \int_{\varphi} \left(I_c(\varphi,t) + A_c(\varphi,t)\right) d\varphi\end{aligned} \tag{4}$$

Inactive virions V_{inac} accumulate in the medium and the ratio of inactive virions to the overall amount of released virions is given by $(1 - P_{eff})$.

$$\begin{aligned}\frac{dV_{inac}(t)}{dt} = &\quad (1-P_{eff})\, k_{rel} \int_{\varphi} \left(I_c(\varphi,t) + A_c(\varphi,t)\right) d\varphi \\ &+ \; k_{deg}\, V_{ac}(t)\end{aligned} \tag{5}$$

3. NUMERICAL SOLUTION

For simulation, the model was discretized with a finite volume method with 128 logarithmically distributed control volumes with respect to the internal coordinate. The delay was approximated with a transport system with 50 control volumes. The overall system to be solved consists of the discretized partial differential equations (1)-(3), the transport system and the equations (4) and (5). The resulting large scale system of ordinary differential equations is solved with the MATLAB solver *ode15s*. Computations could be speeded up significantly by providing analytical patterns of the Jacobian with the solver option *"JPattern"*.

4. ESTIMATION OF KINETIC PARAMETERS

4.1 Preprocessing of flow cytometric measurements

The flow cytometric data comprises the numbers of cells with a specific fluorescence intensity on a logarithmic fluorescence grid with 1024 compartments. As the model is solved numerically on an logarithmic grid with 128 finite volumes, the measurement records are transformed to the same grid by simply averaging the cell concentrations over eight channels. For comparability, the flow cytometric measurements have to be converted into number density functions as described in Müller et al. (2011). For this purpose the average number of cells in each compartment $Z_i(t_k)$ is normalized with the overall cell number of the flow cytometric sample. In addition, the number density function is scaled by the overall concentration of cells in the reactor $C(t_k)$ which can be measured independently:

$$\widetilde{c}_i(t_k) = \frac{Z_i(t_k)}{\sum\limits_i Z_i(t_k)} C(t_k)\,, \quad i = 1, \ldots, 128 \tag{6}$$

4.2 Translation of the inverse problem to a finite dimension

To adapt the presented model to experimental data an inverse problem has to be solved. This problem is set in an infinite dimension as the parameters characterizing the

kinetic processes of apoptosis, cell death, and virus replication/release depend on the internal coordinate φ. Two general solution approaches are described in literature. In the first one, the functional dependency of the parameter on the internal coordinate is assumed to be known and can be described by an analytic funtion with a few parameters, e.g. a gaussian distribution characterized by mean and variance (Sherer et al., 2008). If the shape of the function is not known a priori, it can be parametrized with a suitable approximation, e.g. piecewise constant, piecewise linear or spline approximation. The function is represented by a suitable number of nodes n_a. In this contribution, we follow the approach suggested by Luzyanina et al. (2009) where the unknown functional dependencies are approximated by piecewise cubic hermite - splines:

$$a(\varphi) = \sum_{j=1}^{n_a} a_j \phi_j(\varphi) \tag{7}$$

where ϕ_j is a piecewise cubic polynomial defined on the mesh $\Phi = [\varphi_0, \varphi_1, \ldots, \varphi_{n_a}]$. This approximation yields a function which is continuous up to its first derivative. Two types of parametrization may be considered. The first one assumes the nodes to be equidistantly distributed and only the values of the nodes are subject to the optimization. The second one considers node location and node value as parameters for the estimation procedure. Due to complexity issues, this contribution only deals with the first approach.

4.3 Overall Parameter Estimation Setup

The overall vector of unknown parameters is given by

$$\begin{aligned}\mathbf{p} = [\, & k_{vi},\ k_{rel},\ k_{deg},\ P_{eff},\\ & k_{net}(\varphi_1), \ldots, k_{net}(\varphi_{n_s}),\\ & k_{apo}(\varphi_1), \ldots, k_{apo}(\varphi_{n_s}),\\ & k_{net,apo}(\varphi_1), \ldots, k_{net,apo}(\varphi_{n_s}),\\ & k_{cd}(\varphi_1), \ldots, k_{cd}(\varphi_{n_s}) \quad] \end{aligned} \tag{8}$$

In addition to flow cytometric distribution data, the concentration of active virions $\widetilde{V}_{ac}$ and the total concentration of virus particles $\widetilde{V}$ are recorded yielding the measurement vector

$$\widetilde{y}(t_k) = \begin{pmatrix} \widetilde{c}_1(t_k) \\ \vdots \\ \widetilde{c}_{128}(t_k) \\ \widetilde{V}_{ac}(t_k) \\ \widetilde{V}_{inac}(t_k) \end{pmatrix} = \begin{pmatrix} \widetilde{c}_1(t_k) \\ \vdots \\ \widetilde{c}_{128}(t_k) \\ \widetilde{V}_{ac}(t_k) \\ \widetilde{V}(t_k) - \widetilde{V}_{ac}(t_k) \end{pmatrix} . \tag{9}$$

The parameters can now be estimated in a generalized least squares sense

$$\min_{\mathbf{p}} \sum_{k=1}^{n_t} e(t_k)^T \mathbf{W}(t_k) e(t_k) \tag{10}$$

with

$$e(t_k) = [\widetilde{y}(t_k) - y(t_k, \mathbf{p})] \ . \tag{11}$$

The weighting matrix $\mathbf{W}(t_k)$ is chosen as

$$\mathbf{W}(t_k) = diag\left[\max \widetilde{y}^{-2}\right] \ . \tag{12}$$

The reference values

$$y(t_k) = \begin{pmatrix} c_1(t_k) \\ \vdots \\ c_{128}(t_k) \\ V_{ac}(t_k) \\ V_{inac}(t_k) \end{pmatrix} \tag{13}$$

with

$$c_i = U_{c,i} + I_{c,i} + A_{c,i} \,, \quad i = 1, \ldots, 128 \tag{14}$$

are obtained from the simulation of the model equations as described previously.

The emerging nonlinear least squares problem is solved with the gradient based optimization routine SNOPT (Gill et al., 2002) which can be called from MATLAB through an interface. SNOPT is a commercial software but a free student version is available. As this method can trap in local minima for non-convex objective functions, several runs with different initial guesses were always performed.

Due to the complexity in the parameter estimation procedure, in particular the high sensitivity of the delay on the cost function, a two step procedure was implemented. At first, τ was kept fixed and the other parameters were estimated. In a second step, all parameters exclusive of τ were set to the estimation results from the first step and τ was estimated.

Initial estimates for the unknown parameters without τ are based on the ones given in Müller et al. (2011). An initial value for the delay of $\tau_0 = 4\ h$ is chosen.

5. RESULTS

5.1 Linear Parameter Dependency

Table 1. Initial estimates and estimation results for constant model parameters

Parameter	k_{vi}	k_{rel}	k_{deg}	P_{eff}	τ
Unit	$ml \cdot h^{-1}$	h^{-1}	h^{-1}	–	h
Initial	$5 \cdot 10^{-8}$	600	0.1	0.02	4
Linear	$23.7616 \cdot 10^{-8}$	598.63	0.1211	0.0308	3.9952
3-nodes	$30.1773 \cdot 10^{-8}$	587.57	0.1124	0.0279	3.9976

In a first step, the functional shape is approximated by a hermite spline with two nodes located at φ_{min} and φ_{max}. Thereby, the unknown distributed parameters are assumed to depend linearly on the degree of fluorescence. In Figure 2 - 3 it can be seen that the model can be fitted adequately to the data. The value of the least squares cost function (10) is 4.5752. The resulting constant parameters are summarized in Table 1. In Figure 4, the functional dependencies of the distributed parameters are depicted. It can be seen that k_{net} increases with rising degree of fluorescence which is contrary to the assumptions presented in our previous work. The apoptosis rate is also increasing. One possible interpretation is that cells are more likely to become apoptotic if their degree of infection increases. To interpret the estimation results for the net rate $k_{net,apo}$ one has to keep in mind that the fluorescence degree of the apoptotic cells is generally decreasing.

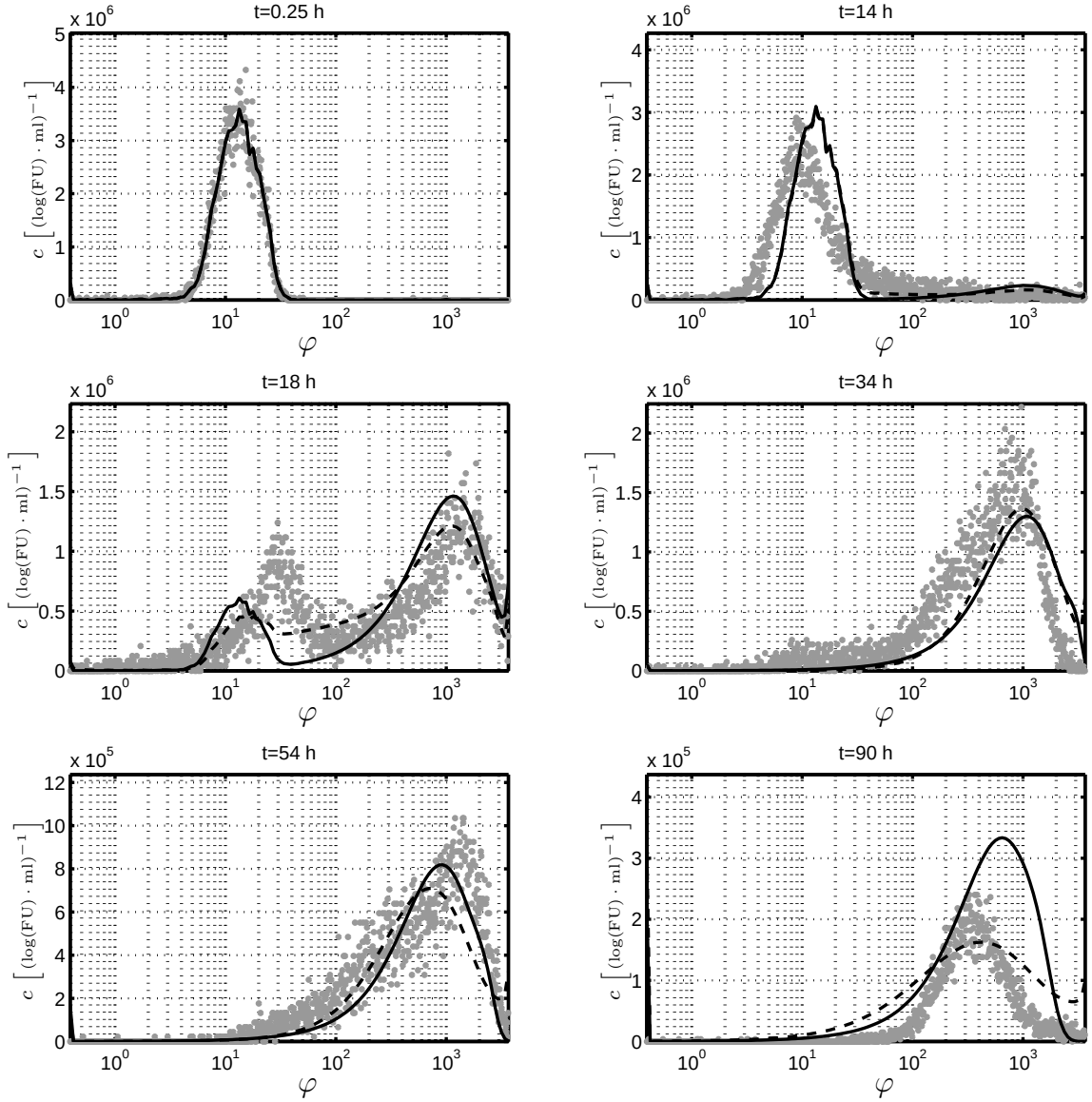

Fig. 2. Flow cytometric measurements (dotted) and model predictions for linear (solid) and nonlinear approach (dashed)

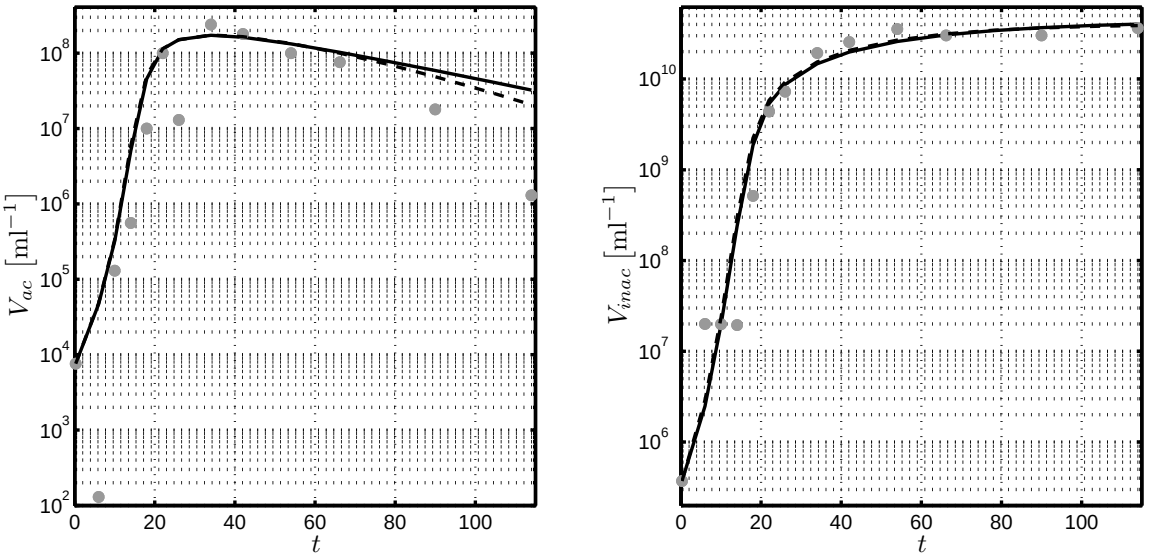

Fig. 3. Active and inactive virions: measurements (dotted) and model predictions for linear (solid) and nonlinear approach (dashed)

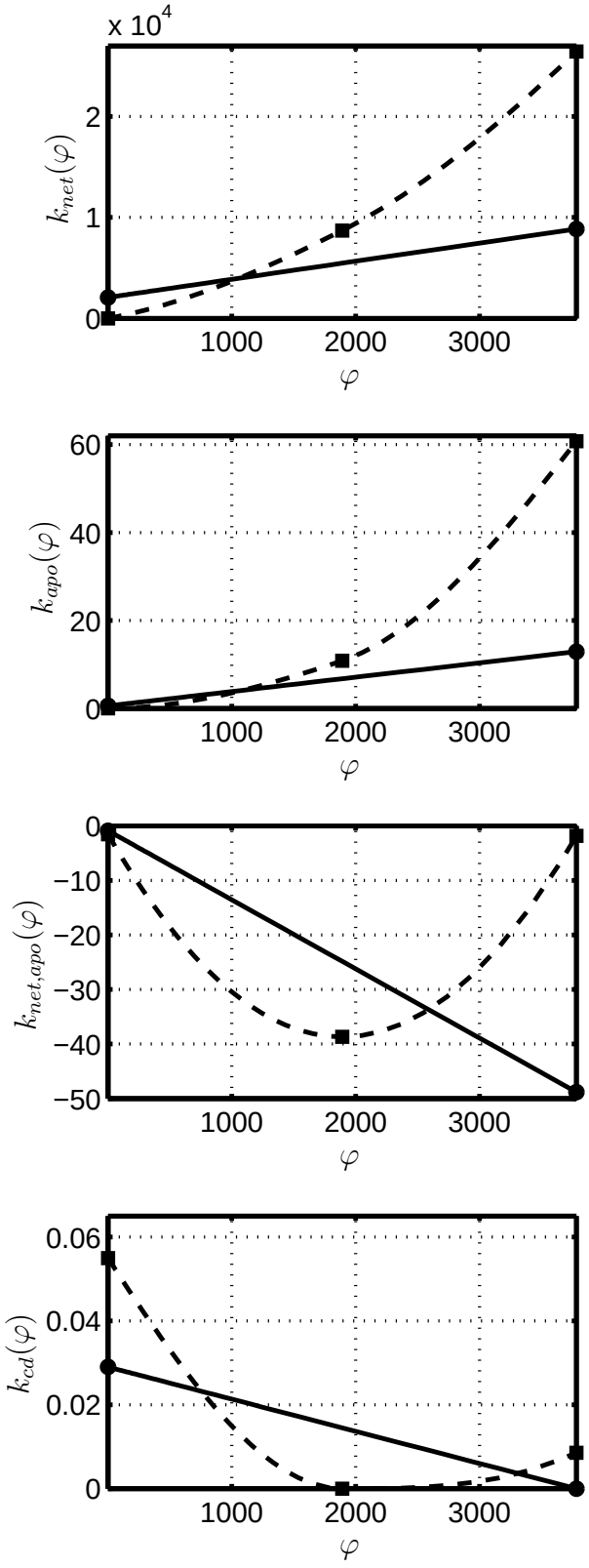

Fig. 4. Estimation results for fluorescence dependent parameters, linear approach (solid), nonlinear approach (dashed); corresponding spline nodes are displayed by dots and squares

It can be seen in Figure 4 that the absolut value of $k_{net,apo}$ is decreasing into the direction of lower fluorescence levels. Hence the loss of fluorescence intensity is slowed down the more the apoptotic cells move to lower fluorescences. The estimated cell lysis rate k_{cd} implies that apoptotic cells in lower fluorescence regions are more probable to lyse. In combination with the overall negative net rate $k_{net,apo}$ the following interpretation comes to mind: the longer a cell is apoptotic the more probable it is to lyse.

5.2 Nonlinear Parameter Dependency

When the distributed parameters are approximated by hermite splines with three nodes, a nonlinear functional dependency is assumed. As a result, the fit in the fluorescence distribution can be improved significantly particularly for later sample points (see Fig. 2). The fit of the active virions is improved in later time samples, too. Comparing the least squares cost function (10) shows that the value drops to 3.8614. The estimation results for k_{net} and k_{apo} show the same trend as in the case of linear parameter dependencies though the slopes are larger. The cell death rate is large for apoptotic cells with a low degree of fluorescence and decreases with increasing fluorescence intensities. For $\varphi \to \varphi_{max}$ the rate is increasing which can be explained in combination with the results for the net rate of the apoptotic cells. The rate has a parabolic shape. This means that apoptotic cells with a low or a very high intracellular amount of NP protein do only move very slowly into regions of lower φ. One possible explanation may be that apoptotic cells within regions of very high fluorescence intensity tend to die instead of staying apoptotic and moving to regions of lower fluorescence levels.

For more complex parameter functions with more than three nodes the parameter estimation problem became ill posed resulting in multiple local minima.

6. CONCLUSION

A deterministic population balance model for influenza virus replication in MDCK cell cultures has been presented. The model comprises a system of hyperbolic partial differential equations using a fluorescence degree as internal coordinate and ordinary differential equations. Unknown kinetic parameters were estimated from experimental data. Therefor, hermite spline approximations of different complexity for the distributed parameters were considered to translate the inverse problem to a finite dimension. In a first step, a linear dependency on the degree of fluorescence was assumed. This allows to reproduce the virus dynamics and the distribution dynamics with a reasonable agreement to the data. The fit could be improved further using nonlinear functional dependencies of the parameters on the internal coordinate. The model gives insight into the general mechanisms during vaccine production. Future efforts will be made to include structured information of intracellular processes like the virus replication mechanism or apoptosis induction into the model formulation.

REFERENCES

Banks, H.T., Sutton, K.L., Thompson, W.C., Bocharov, G., Roose, D., Schenkel, T., and Meyerhans, A. (2011).

Estimation of cell proliferation dynamics using CFSE data. *Bull. of Math. Biol.*, 73(1), 116–150.

Gill, P.E., Murray, W., and Saunders, M.A. (2002). SNOPT: An sqp algorithm for large-scale constrained optimization. *SIAM J. on Optim.*, 12(4), 979–1006.

Luzyanina, T., Roose, D., and Bocharov, G. (2009). Distributed parameter identification for a label-structured cell population dynamics model using CFSE histogram time-series data. *J. of Math. Biol.*, 59(5), 581–603.

Müller, T., Dürr, R., Isken, B., Schulze-Horsel, J., Reichl, U., and Kienle, A. (2011). Population balance modelling of influenza virus replication during vaccine production - influence of apoptosis. In *ESCAPE 21*, 1336–1340.

Müller, T., Schulze-Horsel, J., Sidorenko, Y., Reichl, U., and Kienle, A. (2008). Population balance modelling of influenza virus replication during vaccine production. In *ESCAPE 18*, 133–138.

Schulze-Horsel, J., Schulze, M., Agalaridis, G., Genzel, Y., and Reichl, U. (2009). Infection dynamics and virus-induced apoptosis in cell culture-based influenza vaccine production–flow cytometry and mathematical modeling. *Vaccine*, 27(20), 2712 – 2722.

Sherer, E., Tocce, E., Hannemann, R.E., Rundell, A.E., and Ramkrishna, D. (2008). Identification of age-structured models: Cell cycle phase transitions. *Biotechn. and Bioeng.*, 99(4), 960–974.

Sidorenko, Y., Schulze-Horsel, J., Voigt, A., Reichl, U., and Kienle, A. (2008a). Stochastic population balance modeling of influenza virus replication in vaccine production processes. *Chem. Eng. Sc.*, 63(1), 157–169.

Sidorenko, Y., Voigt, A., Schulze-Horsel, J., Reichl, U., and Kienle, A. (2008b). Stochastic population balance modeling of influenza virus replication in vaccine production processes. ii. detailed description of the replication mechanism. *Chem. Eng. Sc.*, 63(8), 2299–2304.

Walter, E. and Pronzato, L. (1997). *Identification of Parametric Models from Experimental Data*. Communications and Control Engineering. Springer, 1 edition.