Parameter Identification Problem Based on FRAP Images: From Data Processing to Design of Photobleaching Experiments

Ctirad Matonoha¹, Jiří Jablonský², Štěpán Papáček²

¹Institute of Computer Science, AS CR, Prague

²Institute of Complex Systems, CENAKVA, FFPW, University of South Bohemia in České Budějovice

> *joint work with* S. Kindermann (JKU LINZ)

High Performance Computing in Science and Engineering, May 25 - 28, 2015, Hotel Soláň IT4Innovations National Supercomputing Center,

VSB-Technical University of Ostrava

FRAP data: Spatio-temporal images



Experimental data from an experiment with red algae *Porphyridium cruentum*. FRAP (**Fluorescence Recovery After Photobleaching**) technique helps us to reveal the mobility of photosynthetic proteins (phycobilisomes) due to the DIFFUSION (only!?) on thylakoid membrane.

State-of-the-art FRAP data (pre)processing



A - **Upper left:** Representative image taken 8 s after bleaching. The regions of interest (ROI) are labeled as yellow rectangles.

A - **Lower left:** The recovery dynamics in time-sequence of one-dimensional bleach profiles perpendicular to the bleach stripe. The fluorescence in central zone is recovered with the growing time.

B - On the right: So-called FRAP recovery curves for 4 different ROI, i.e., time series of space-averaged fluorescence signal.

Initial Boundary Value Problem (IBVP): Reaction-diffusion PDE for a previously not fixed number (q_c) of interacting components $(u_i)_{i=1}^{q_c}$,

$$\frac{\partial u(x,t)}{\partial t} = D\Delta u(x,t) - Ku(x,t)$$
(1)

with initial conditions

$$u_i(x,0) = u_{i0}\phi(x), \quad i \in \{1,\ldots,q_c\},$$
 (2)

 $\phi(x)$ is some given initial shape, D is a diagonal matrix of diffusion coefficients $(D_i)_{i=1}^{q_c}$, and K is a (singular) matrix of (q_r) reaction rates. The total number of parameters to identify is $q = q_c + q_r$, i.e., $p \in \mathbb{R}^q$. The measured fluorescent signal is proportional to the sum of concentration profiles $\sum_{i=1}^{n} u_i(x, t)$.

Boundary conditions could be, e.g.,

$$u(x,t) = g_D(x,t)$$
 or $\frac{\partial u(x,t)}{\partial n} = g_N(x,t)$ on $\partial \Omega \times [0,T].$ (3)

Q.#1. Is it possible to identify both model structure and parameters?

Parameter identification problem

We define the forward map (also called the parameter-to-data map)

$$F: \mathcal{R}^q \to \mathcal{R}^{N_{\text{data}}}, \quad F(p) = (u(x_i, t_i))_{i=1}^{N_{\text{data}}}$$

Our regression model is now

$$F(p) = \text{data} \tag{4}$$

where the data are modeled as contaminated with additive white noise

$$\text{data} = F(p_T) + e = (u(x_i, t_i))_{i=1}^{N_{\text{data}}} + (e_i)_{i=1}^{N_{\text{data}}}$$

Here $p_T \in \mathcal{R}^q$ denote the true coefficients and $e \in \mathcal{R}^{N_{\text{data}}}$ is a data error vector which we assume to be normally distributed with variance σ^2 :

$$e_i = \mathcal{N}(0, \sigma^2), \quad i = 1, \dots, N_{\text{data}}$$

The aim is to find a vector of parameters $p \in \mathcal{R}^q$ such that (4) is satisfied in some appropriate sense:

$$\|F(p_c) - \text{data}\|^2 = \min_{p} \|F(p) - \text{data}\|^2.$$
 (5)

This problem is usually ill-posed thus regularization has to be employed.

Sensitivity analysis: Key concept of our approach

For the sensitivity analysis we require the Fréchet-derivative $F'[p] \in \mathcal{R}^{N_{\text{data}} \times q}$ of the forward map F, that is

$$F'[p] = \frac{\partial}{\partial p} F(p) = \left(\frac{\partial}{\partial p} u(x_1, t_1), \dots, \frac{\partial}{\partial p} u(x_{N_{\text{data}}}, t_{N_{\text{data}}})\right)^T \quad (6)$$

A corresponding quantity is the Fisher information matrix (FIM)

$$M_q = F'[p]^T F'[p] \in \mathcal{R}^{q \times q}$$

which collapses into the scalar quantity

$$M_1 = \sum_{i=1}^{N_{\text{data}}} \left(\frac{\partial}{\partial p} u(x_i, t_i) \right)^2$$

for the one single parameter case ($p \in \mathcal{R}$).

Q.#2. How to determine the Fréchet-derivatives having real data only?

Suppose we have computed $p_c \in \mathcal{R}$ as a least-squares solution to (5). Let us define the residual as

$$\mathrm{res}^2(p_c) = \|F(p_c) - \mathrm{data}\|^2 = \sum_{i=1}^{N_{\mathrm{data}}} \left[u_{p_c}(x_i, t_i) - \mathrm{data}_i\right]^2,$$

where u_{p_c} is a solution to (1-3) for the computed parameter value p_c .

Then it is possible to quantify an error between p_c and p_T . In fact, we bound the product of two quantities: $1 - \alpha$ **confidence interval** (squared) and the sum of squared sensitivities

$$(p_c - p_T)^2 \sum_{i=1}^{N_{\text{data}}} \left[\frac{\partial}{\partial p} u(x_i, t_i) \right]^2 \le \frac{\text{res}^2(p_c)}{N_{\text{data}} - 1} f_{1, N_{\text{data}} - 1}(\alpha) \approx \sigma^2, \quad (7)$$

where $f_{1,N_{data}-1}(\alpha)$ corresponds to the upper α quantile of the Fisher distribution with 1 ($\rho_c \in \mathcal{R}$) and $N_{data} - 1$ degrees of freedom.

Assume the Fick diffusion equation for one component (concentration y) in 1D domain, i.e.,

$$\frac{\partial}{\partial t}y(x,t)=p\Delta y(x,t),$$

 $x \in \mathcal{R}$ and $y(x, 0) = y_0(x) = u_{0,0} exp(-2x^2/r_0^2)$. Further we compare two data (pre)processings: Integrated vs. Full data approach. We compute the parameter values p_c corresponding to both approaches and 4 cases of data regions with the growing size (ratios $\frac{L}{2r_0} = 1, 5, 10, 20$). The resulting visualization of all 8 results is presented in the next slide – boxplots of the squared error $|p_T - p_c|^2$ corresponding to 4 cases of data regions with the growing size (ratios).

In each pair, the left boxplot corresponds to using full data, while the right boxplot corresponds to using spatially integrated data on the same ROI. According to [4], the **data set represented by the FRAP** recovery curves (the integrated data approach) leads to a larger confidence interval compared to the spatio-temporal data (confirmed by this case study, indeed).



Note: While the value $|p_T - p_c|^2$ for the full data case is decreasing for the growing $\frac{L}{2r_0}$, it exists an optimal L for the integrated data case.

Optimizing an experimental design variable – bleach radius r₀,

By maximizing the sensitivity (\Rightarrow minimal confidence interval) we can optimize an experimental design variable, e.g., bleach radius r_0 .

Consider the case of a dense set of observations on a space-time cylinder $Q = \left[-\frac{L}{2}, \frac{L}{2}\right] \times [0, T]$ in 1D case and try to infer about the optimal bleach radius r_{opt} yielding maximal sensitivity.

We approximate the sensitivity by integrals and forget the grid factor $\frac{1}{\Delta x \Delta t}$ which is assumed to be fixed. We introduce the function

$$S(r_0) = \int_0^T \int_{-\frac{t}{2}}^{\frac{t}{2}} \left| \frac{\partial}{\partial p} u(x, t) \right|^2 \mathrm{d}x \mathrm{d}t,$$

and we try to find out its maximal value

$$S(r_{opt}) = \max_{r_0>0} S(r_0).$$

In the setting according to [4], $S(r_0)$ has a unique maximum

$$r_{opt} \approx 1.728 \sqrt{T \rho_c}.$$
 (8)

Q.#3. Does some explanation exist for (8)?

- Our method of FRAP data processing, based on full spatio-temporal data, IBVP formulation and the sensitivity analysis, outperforms the state-of-the-art methods because it does not use any simplified assumptions (e.g., homogeneous B.C.).
- Using our software CA-FRAP we determine both the diffusion coefficients **mean value** and its confidence interval, as well.
- The analysis of sensitivity (FIM) leads us to the **optimal experimental design**. Experimental factors, e.g., size of the bleach spot, time interval between measurements, size of the monitored area (ROI), are set in order to **MAXIMIZE SENSITIVITY**.

Thanks & (Our) References



🗣 Papáček. Š., Kaňa, R., and Matonoha, C.

Estimation of diffusivity of phycobilisomes on thylakoid membrane based on spatio-temporal FRAP images.

Mathematical and Computer Modelling, Vol. 57, p. 1907-1912, 2013.



Kaňa R., Kotabová E., Lukeš M., Papáček Š., Matonoha C., Liu L.N., Prášil O., Mullineaux C.W.

Phycobilisome Mobility and Its Role in the Regulation of Light Harvesting in Red Algae

Plant Physiology, Vol. 165, p. 1618-1631, 2014.

🔈 Matonoha, C., and Papáček, Š.

On the connection and equivalence of two methods for solving an ill-posed inverse problem based on FRAP data.

Journal of Computational and Applied Mathematics, submitted.



📎 Papáček, Š., Jablonský, J., Kaňa, R., Matonoha, C., Kindermann, S.

From data processing to experimental design and back again: A parameter identification problem based on FRAP images.

Proceedings of 17th ICDIP 2015, Dubai, UAE.

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic – projects CENAKVA (No. CZ.1.05/2.1.00/01.0024) and CENAKVA II (No. LO1205 under the NPU I program), by Postdok JU (CZ.1.07/2.3.00/30.0006), and by the long-term strategic development financing of the Institute of Computer Science (RVO:67985807).