Parameter Identifiability of RNA Dynamics in PDE Transport Models of Fluorescence Recovery After Photobleaching

Qinyu Xu[†]

Project advisor: Dr. Veronica Ciocanel[‡]

Abstract. The transport and localization of RNA molecules, crucial for cellular function and development, involve a combination of diffusion and active transport mechanisms. *Xenopus laevis* oocytes provide a well-established model system for studying RNA transport due to their large size, pronounced spatial organization, and well-characterized RNA localization patterns. Fluorescence Recovery After Photobleaching (FRAP) is an experimental technique that is widely used to investigate the dynamics of molecular movement within cells by observing the recovery of fluorescence intensity in a photobleached region over time. Previous studies have primarily modeled RNA dynamics in FRAP experiments using purely diffusion mechanisms without explicitly incorporating active transport. To advance the understanding of RNA dynamics, we develop a reaction-diffusion-advection partial differential equation (PDE) model integrating both transport and diffusion mechanisms. We propose a pipeline for identifiability analysis to assess the model's ability to uniquely determine parameter values from observed FRAP data. Based on profile likelihood analysis and reparametrization, we examine the relationship between non-identifiable parameters, which improves the robustness of parameter estimation. We find out that the identifiability of the four parameters of interest is not exactly the same in different regions of the cell. Specifically, transport velocity and diffusion coefficient are identifiable in all regions of the cell, and some combinations of binding rate and unbinding rate are found to be identifiable near the nucleus.

1. Introduction.

Studying RNA dynamics is central to understanding how gene expression is controlled in space and time during cellular processes, and it plays an especially important role in developmental biology where spatially regulated translation directs cell fate and patterning. A classical and experimentally tractable system for this is oocyte development in the frog *Xenopus laevis*. According to [10], *Xenopus laevis* undergo substantial growth during maturation, reaching a diameter of roughly 1.3 mm, and develop strong polarity characterized by the spatial distribution of numerous mRNAs and proteins. For instance, Vg1 and VegT mRNAs localize to the vegetal pole and are crucial for mesoderm and endoderm induction, respectively, thereby influencing embryonic axes specification [11, 15]. This subcellular mRNA patterning relies on microtubule-dependent transport, with RNA-binding proteins such as Vg1RBP/Vera mediating the recognition of localization elements and guiding transport [4]. Thus, quantitatively characterizing RNA transport and localization in *Xenopus laevis* oocytes provides mechanistic insight into how molecular-level dynamics drive large-scale developmental patterning, establishing a direct link between RNA movement and cell fate determination during early embryogenesis [6].

In living cells, RNA dynamics are commonly studied with fluorescence microscopy techniques, among which fluorescence recovery after photobleaching (FRAP) is a widely used, robust method

[†]Duke Kunshan University (qx44@duke.edu).

[‡]Departments of Mathematics and Biology, Duke University (veronica.ciocanel@duke.edu).

in biological analysis. It is a standard tool in cell biology for probing protein and RNA–protein complex dynamics in their native context [8, 13]. In a FRAP experiment, fluorescently tagged molecules, such as RNA–protein complexes labeled with a fluorescent protein, are visualized under a widefield microscope. A small, well-defined region of the cell is then exposed to a high-intensity laser pulse, which irreversibly photobleaches the fluorophores in that region, causing their fluorescence to drop sharply. Following bleaching, the fluorescence in the region is monitored over time as unbleached molecules from surrounding areas move in. The resulting fluorescence recovery curve reflects the mobility, transport, and exchange rates of the molecules [14]. This process is shown in Figure 1.

To interpret FRAP data and relate observed recovery curves to underlying physical processes, mathematical models have been developed in order to simulate the moving behaviors of RNA inside the cells, which involve several parameters such as diffusion coefficient and transport speed. Accurate parameter estimation allows researchers to quantitatively characterize RNA behavior, validate the model against experimental data, and gain insights into the underlying biological mechanisms. Crucially, however, the ability to infer parameters depends not only on the data quality but also on identifiability — whether the model structure and available data observations allow unique recovery of parameter values. Performing identifiability analysis is essential for the accurate parameter estimation, as it reveals which parameters can be uniquely determined based on the model and observed data, while other non-identifiable parameters cannot.

This paper focuses on the identifiability analysis of parameters in a reaction–diffusion–advection model of RNA localization during *Xenopus laevis* oocyte development. Building on prior studies that investigated the identifiability of the parameters in a reaction-diffusion model for RNA-binding protein PTBP3 in *Xenopus laevis* oocytes [1] and the reaction-diffusion-advection model exploration for RNA dynamics [2], we explicitly incorporate active transport into the inference framework and examine identifiability across spatially distinct oocyte regions. Our goal is to develop mathematical tools in order to fill in the gaps in the current understanding of RNA movement.

In the cellular environment, the nucleus is typically centrally located, while the vegetal cortex, situated opposite the animal pole, is a specialized region near the cell membrane. According to [2], it is believed that the moving behaviors of RNA are not the same in different regions of the cell, so we divide the cell into three regions as depicted in Figure 2, where Region 1 is 15 μm from the nucleus, Region 2 is 50 μm from the nucleus, and Region 3 is 25 μm from the vegetal cortex [2]. These three regions are selected to represent biologically distinct microenvironments within the oocyte, as previous FRAP experiments have shown that RNA transport mechanisms and cytoskeletal organization differ near the nucleus, in the intermediate cytoplasm, and close to the vegetal cortex. For each region, a 5-μm circular area of interest is bleached [2], so we can obtain three average FRAP data sets from experiments. As in previous work [2], after obtaining the baseline parameter sets that fit the real experimental FRAP data based on deterministic parameter estimation, we generate the corresponding synthetic FRAP data for the three regions. Then, the method of profile likelihood gives us the framework for determining the identifiability of parameters. We also derive the threshold for 95% confidence interval, which serves as the criterion to decide whether a certain

parameter is identifiable or not. Moreover, we investigate the parameter relationships between non-identifiable parameters, as well as apply this methodology to real and synthetic FRAP data.

By explicitly integrating active transport into the identifiability analysis and by examining the FRAP data across biologically meaningful oocyte regions, our work extends previous efforts and provides a practical framework for quantifying which transport model parameters are identifiable when analyzing mRNA localization in development using FRAP data.

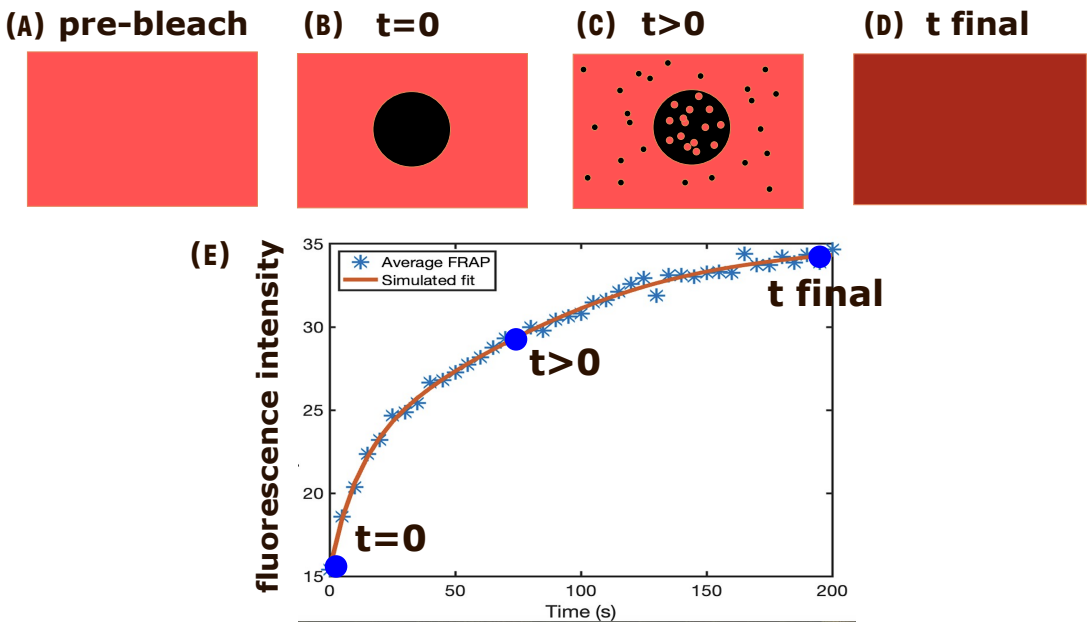

Figure 1. *FRAP analysis showing: (A) Initial uniform fluorescence, (B) Sharp intensity drop after photobleaching, (C) Gradual recovery as molecules diffuse into the bleached region, and (D) Final equilibrium fluorescence. The process of fluorescence recovery is depicted in (E). The recovery kinetics can reveal molecular mobility and binding dynamics.*

2. Mathematical modeling of the FRAP experiment. This section presents the reaction-diffusion-advection PDE model used to describe RNA dynamics in FRAP experiments, which details the experimental data acquisition and outlines the deterministic parameter estimation procedure to obtain baseline parameter values for subsequent identifiability analysis.

2.1. Reaction-diffusion-advection model for RNA dynamics. We model the dynamics of RNA based on two localization mechanisms. RNA can either diffuse with the diffusion coefficient D in the cytoplasm, or can be transported along microtubules by motor proteins with the speed c. Moreover, we assume that the RNA molecules can switch between these two mechanisms with certain rates. The RNA dynamics is therefore shown in Figure 3, which is given by this reaction-diffusion-advection model [2]:

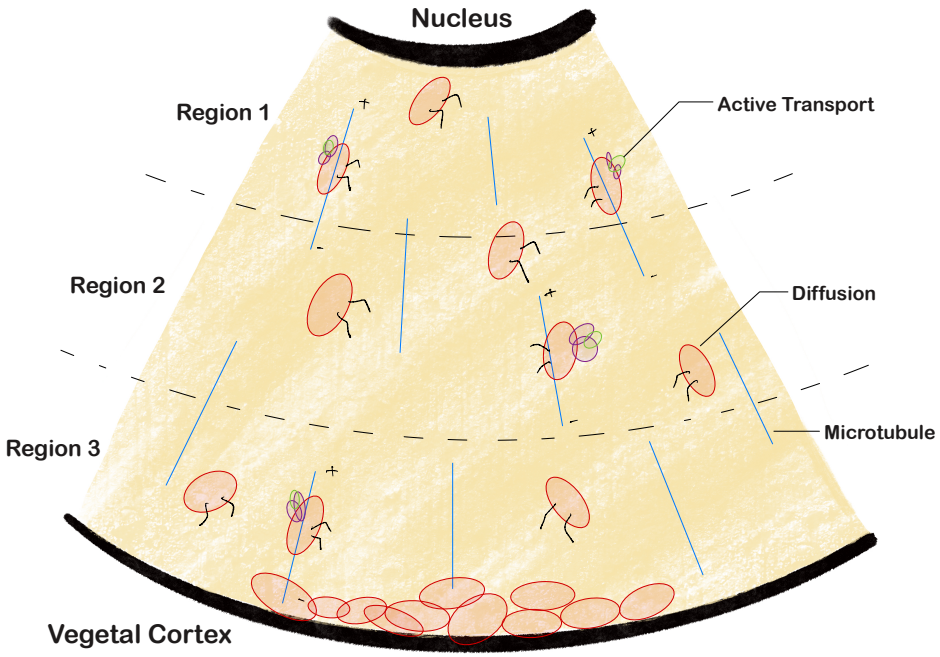

Figure 2. *This cartoon shows that experiments are done in three regions of the cell based on their distance from the nucleus (top) and vegetal cortex (bottom), following previous studies in [2] that partitioned the oocyte into three representative regions for FRAP measurements. Red particles represent RNAs, which are transported either by active transport, the directed movement along blue microtubule tracts driven by motor proteins (purple and green blobs), or by simple diffusion. Region 1 (top) is 15 μm from the nucleus, Region 2 (middle) is 50 μm from the nucleus, and Region 3 (bottom) is 25 μm from the vegetal cortex.*

$$\begin{aligned} \frac{\partial u}{\partial t} &= c\frac{\partial u}{\partial y} - \beta_1 u + \beta_2 v, \\ \frac{\partial v}{\partial t} &= D\Delta v + \beta_1 u - \beta_2 v. \end{aligned} \tag{2.1}$$

In this PDE model, u refers to the concentration of RNA which is transported along microtubules, while v refers to the concentration of RNA which is diffusing. Since the bleached area in the FRAP process is very small, we assume that within this small area, all the directions of microtubules are vertical, which is the reason why we just include $\frac{\partial u}{\partial y}$ in Equation (2.1) (i.e., 1-dimensional transport). Furthermore, we define β_1 as the switching rate from transport RNA to diffusing RNA, which represents the unbinding rate between RNA and microtubules. Similarly, β_2 is defined as the binding rate, which is the switching rate from diffusing RNA to transport RNA. Based on this model, we are able to simulate the FRAP process using numerical PDE methods as in [5, 3]. The observed FRAP recovery signal $F(t)$ is modeled as the total fluorescence within the bleached region Ω, which depends on both the transport and diffusing RNA populations. Specifically, the fluorescence intensity is proportional to the sum of the concentrations u and v, integrated over the

bleach spot:

$$F(t) = \int_{\Omega} [u(\mathbf{x}, t) + v(\mathbf{x}, t)]\, d\mathbf{x} \tag{2.2}$$

where Ω is the circular bleached region of diameter 5 μm, and $\mathbf{x}$ represents the spatial dimensions (x, y). This expression captures the total amount of fluorescent molecules within the region of interest over time.

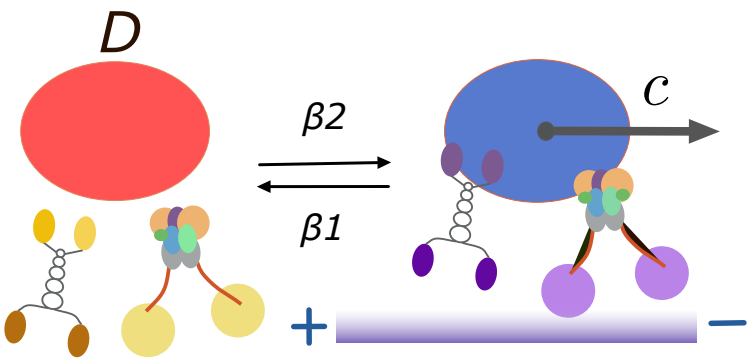

Figure 3. *Cartoon of the active transport of RNA, consisting of a population of diffusing particles with diffusion coefficient D, a population of moving particles with velocity c, as well as switching rates β_1 and β_2 between the two population. In the moving state, we assume that RNA molecules are attached to motor proteins and microtubules, while in the diffusion state, we assume they are detached from microtubules.*

2.2. Experimental FRAP data. The experimental FRAP data used in this study, shown as the blue points in Figure 4, were collected from *Xenopus laevis* oocytes following the procedure described in the FRAP acquisition section of [2]. In the experiments, a small circular region (5 μm in diameter) inside the oocyte was bleached with a laser, and the fluorescence recovery in this region was recorded over 200s at regular time intervals. The recovery curves reflect how labeled RNA molecules move back into the bleached area over time.

For each region of interest, data was collected from five oocytes and their recovery curves were averaged. The fluorescence values were corrected for background changes and normalized to a nearby unbleached region to reduce experimental noise, following the adjustment procedure outlined in [2]. The resulting averaged FRAP data captures the overall recovery trend and will serve as the input for our parameter estimation procedure. In Subsection 2.3, we fit this experimental data to Equation (2.1) in order to extract parameter values and generate synthetic curves, which clearly shows the comparison between the model-generated synthetic curves and the measured data.

2.3. Deterministic parameter estimation. For the real FRAP data observed from 5 oocytes in each region, deterministic parameter estimation was performed following the numerical fitting framework in [2]. In brief, the reaction–diffusion–advection PDE model (Equation (2.1)) was integrated numerically using an exponential time-differencing Runge–Kutta scheme with Fourier

spectral discretization. The observed FRAP signal was modeled as the sum of all particle states integrated over the bleach spot, and parameters c, D, β_1 and β_2 were estimated by minimizing the L^2-norm between model output and experimental curves using the MATLAB routine `lsqnonlin`. The resulting best-fit parameters for each region are summarized in Table 1.

Table 1

Table of parameter estimates for the three regions [2].

Region	c	D	β_1	β_2
1	0.049121	0.258205	2.35E-14	0.006331
2	0.094322	1.423721	0.003018	0.000762
3	0.067619	0.830449	4.05E-05	1.37E-06

We use the parameter estimates in Table 1 to inform some baseline parameter regimes for generating synthetic data, which are shown in Table 2. Table 2 highlights region-dependent differences in the baseline parameter regimes. While the transport velocity and diffusion coefficient vary across regions reflecting distinct transport environments, the binding and unbinding rates differ by orders of magnitude, suggesting that RNA–microtubule interactions may play different roles depending on the spatial location within the oocyte. Based on these baseline values (Table 2) and synthetic FRAP data (red curves in Figure 4), we are able to test our identifiability analysis framework on PDE-generated data before attempting it on real and noisy datasets [1]. It can be seen from Figure 4 that the red synthetic data curves are in qualitative agreement with the blue experimental FRAP data points.

Table 2

Table of baseline values for the three regions used for synthetic data generation in this work.

Region	c	D	β_1	β_2
1	0.05	0.25	10^{-6}	10^{-2}
2	0.1	1.5	10^{-3}	10^{-4}
3	0.1	0.8	10^{-5}	10^{-6}

3. Identifiability analysis for transport models of FRAP experiments.

In this section, we apply profile likelihood analysis to assess the identifiability of four key parameters, establish a statistical threshold for identifiability based on 95% confidence intervals, and extend the analysis to two-dimensional parameter spaces to validate pairwise identifiability of transport velocity and diffusion coefficient.

3.1. Profile likelihood analysis.

We then carry out identifiability analysis based on the method of profile likelihood, starting with the synthetic datasets in Figure 4 . Profile likelihood analysis can provide insights into the identifiability of parameters in the model by systematically setting one parameter constant and optimizing the rest of the parameters in fitting the data. In this work, we define $\boldsymbol{\theta}$ to be the vector of all the parameters of interest: $\boldsymbol{\theta} = (c, D, \beta_1, \beta_2)$. We first choose one interest parameter $\boldsymbol{\psi}$ (for instance, D). By changing the values of that interest parameter, we want to find the optimized values for other nuisance parameters $\boldsymbol{\lambda}$ (for instance, c, β_1, β_2), so that the profile likelihood function is maximized. For instance, if we set D to be the interest parameter,

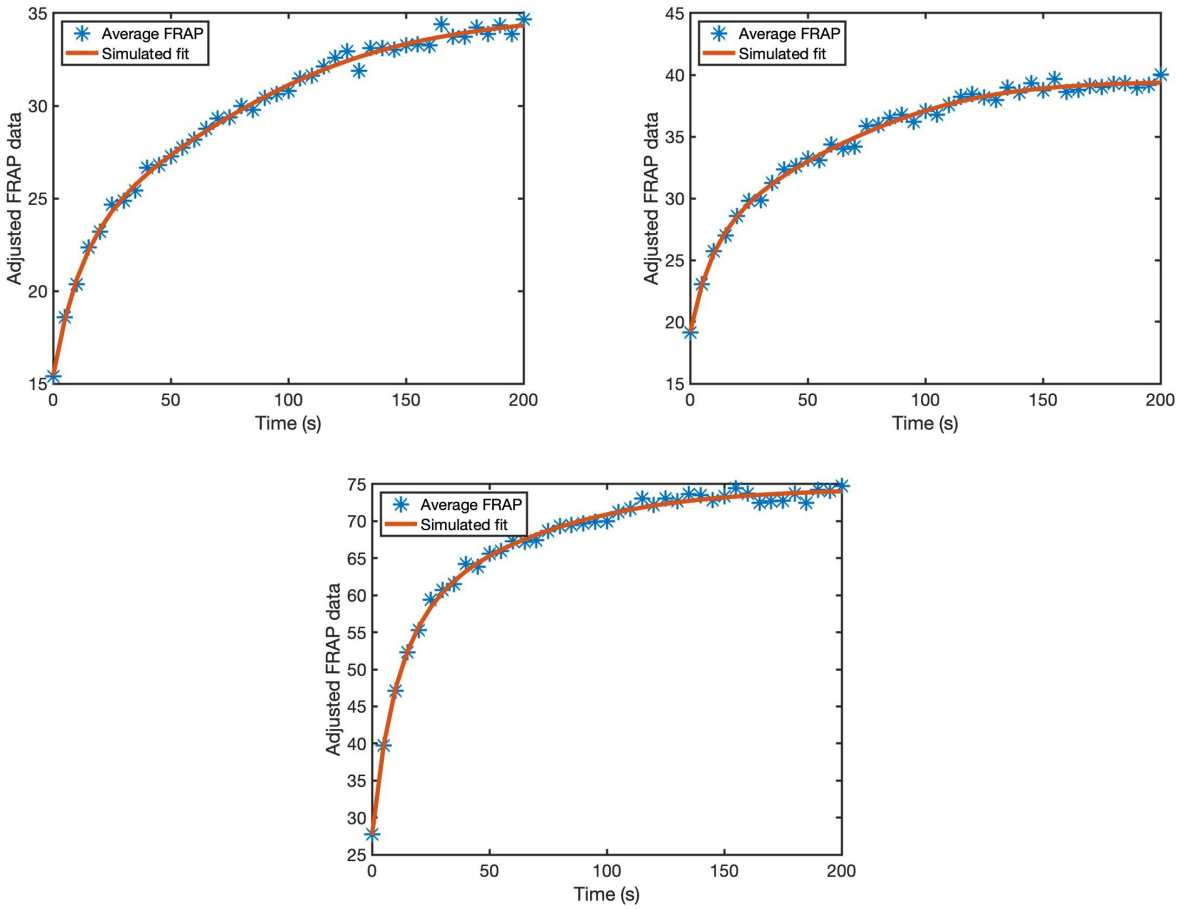

Figure 4. *Real FRAP data (blue points) and synthetic FRAP data (red lines) based on the baseline parameter values in Table 2 (top-left: Region 1, top-right: Region 2, bottom: Region 3).*

then by changing its value, we want to find the optimal $\boldsymbol{\lambda} = (c, \beta_1, \beta_2)$ so that the profile likelihood function is maximized. Moreover, if the measurement noise is assumed to be normally distributed, then the profile likelihood function can be expressed as:

$$p(y; \boldsymbol{\psi}, \boldsymbol{\lambda}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} exp\left(-\frac{1}{2\sigma^2}||y - y_{sim}(\boldsymbol{\psi}, \boldsymbol{\lambda})||_2^2\right) \tag{3.1}$$

where y refers to the noiseless synthetic data curve generated with baseline values in Table 2, $y_{sim}(\boldsymbol{\psi}, \boldsymbol{\lambda})$ refers to the noiseless synthetic data curve generated under each value of the interest parameter $\boldsymbol{\psi}$ together with the optimal values of the nuisance parameters $\boldsymbol{\lambda}$. Moreover, $||y - y_{sim}(\boldsymbol{\psi}, \boldsymbol{\lambda})||_2^2$ represents the least square error, and σ is the standard deviation of the noise in the FRAP data. We estimate σ by calculating the difference between each point on the synthetic data curve and the FRAP data curve generated by parameters in Table 1:

$$\hat{\sigma} = \sqrt{\frac{\sum(y - y_{sim}(\boldsymbol{\psi}, \boldsymbol{\lambda}))^2}{N}} \tag{3.2}$$

where N is the number of data points.

For each region, we have computed the corresponding estimate values for σ: $\hat{\sigma_1} = 0.275, \hat{\sigma_2} = 0.365, \hat{\sigma_3} = 0.614$.

To visualize the results of the profile likelihood analysis, we plot the likelihood value for each value of the interest parameter. The possible results are shown in Figure 5 [1]. In the first plot where the curve is just a flat line, this parameter is considered as structurally non-identifiable, since the profile likelihood is not sensitive to the change of that parameter. In the second plot, the curve does not approach zero at the right end, indicating that this parameter is practically non-identifiable. Unlike structural non-identifiability which arises from inherent structure of the model, practical non-identifiability occurs primarily because the available data provide insufficient information to precisely estimate the parameter. This means that although the model could theoretically determine the parameter uniquely, the available data are not sufficient to estimate it reliably. As for the third case, we consider this parameter to be both structurally and practically identifiable.

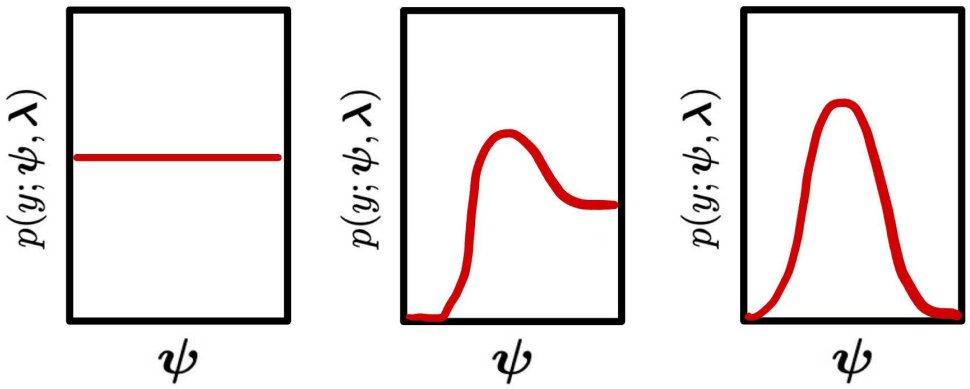

Figure 5. *Interpretation of profile likelihoods. A flat likelihood (left) corresponds to structural non-identifiability, a profile that does not decrease to 0 on one or both sides of the maximum (center) indicates practical non-identifiability, and a profile with a fast decrease to 0 on both sides of the maximum (right) shows both structural and practical identifiability.*

3.2. Threshold for identifiability using profile likelihood analysis. In order to establish a clear criterion to determine whether a parameter is identifiable or not using profile likelihood analysis, we calculate the threshold for the 95% confidence interval for each profile likelihood plot. Specifically, confidence intervals are determined by applying a threshold to the likelihood ratio statistic, corresponding to a χ^2 quantile. Parameters are considered identifiable when likelihood plots cross this threshold on both sides [12]. The basic formula for computing the value of this confidence threshold is given by [12, 7]:

$$CI_{j,\alpha}(y) = \{\alpha_p | 2(LL(y|\hat{\boldsymbol{\theta}}) - PL_j(p)) < \Delta\alpha\} \tag{3.3}$$

where the subscript j denotes the index of the interest parameter being analyzed for identifiability, α is the confidence level (e.g., $\alpha = 0.05$ for a 95% confidence interval). $LL(y|\hat{\boldsymbol{\theta}})$ and $PL_j(p)$ refer to the log likelihood with the interest parameter set as value p:

$$LL(y|\boldsymbol{\theta}) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{\|y - y_{sim}(\boldsymbol{\psi}, \boldsymbol{\lambda})\|^2}{2\sigma^2} \tag{3.4}$$

and $\Delta\alpha$ comes from the chi-square value ($\Delta\alpha = 3.841$ for the 95% confidence interval). Moreover, $\hat{\boldsymbol{\theta}}$ refers to the set of parameter values that gives the maximum value of the profile likelihood. Therefore, $\hat{\boldsymbol{\theta}}$ should be the set of baseline values, and we obtain $LL(y|\hat{\boldsymbol{\theta}}) = -\frac{1}{2}\ln(2\pi\sigma^2)$ with the least square error equal to 0. Then, we plug the log likelihood into Equation (3.3) to get:

$$2(LL(y|\hat{\boldsymbol{\theta}}) - PL_j(p)) < \Delta\alpha \tag{3.5}$$

$$\Rightarrow 2(-\frac{1}{2}ln(2\pi\sigma^2) + \frac{1}{2}ln(2\pi\sigma^2) + \frac{\|y - y_{sim}(\boldsymbol{\psi}, \boldsymbol{\lambda})\|^2}{2\sigma^2}) < \Delta(\alpha) \tag{3.6}$$

$$\Rightarrow \frac{\|y - y_{sim}(\boldsymbol{\psi}, \boldsymbol{\lambda})\|^2}{\sigma^2} < \Delta(\alpha) \tag{3.7}$$

Based on Equation (3.1) for profile likelihood, the value for the threshold is given by:

$$p(y; \boldsymbol{\psi}, \boldsymbol{\lambda}) > \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\Delta\alpha}{2}} \tag{3.8}$$

3.3. Outcomes of profile likelihood analysis. After computing the profile likelihood for all of the four parameters of interest in the three regions, and calculating the threshold for the likelihood in each plot, the results are given in Figure 6.

Overall, all three regions show similar identifiability properties for the four parameters. Notably, c and D are identifiable, while β_1 and β_2 are non-identifiable. Moreover, we notice that β_2 in Region 1 and β_1 in Region 2 are practically non-identifiable. As for region 3, the plots for β_1 and β_2 are just flat lines, suggesting structural non-identifiability for both kinetic rates.

3.4. Two-dimensional profile likelihood. In systems biology, models are becoming increasingly complex as the field strives to provide holistic descriptions of biological systems that capture both their static properties and dynamic interactions. The two-dimensional profile likelihood approach has emerged as a powerful tool by offering enhanced insights into parameter uncertainty and improving model identifiability [9].

To extend the previous one-dimensional profile likelihood to a two-dimensional case, we set c and D to be our two interest parameters. Then, given each possible pair of c and D chosen from a grid around their baseline values, we want to find the optimized values for β_1 and β_2 so that the profile likelihood function is maximized.

$$PL_{c,D}(p_1, p_2) = \max_{\theta \in \{\theta | c=p_1, D=p_2\}} LL(y|\hat{\boldsymbol{\theta}}) \tag{3.9}$$

We visualize the two-dimensional profile likelihood results for the three regions in our model and data by drawing the corresponding 3D plots in Figure 7. For each point on the (c, D)-plane, we obtain one value for the least square error, which is computed by calculating the difference between

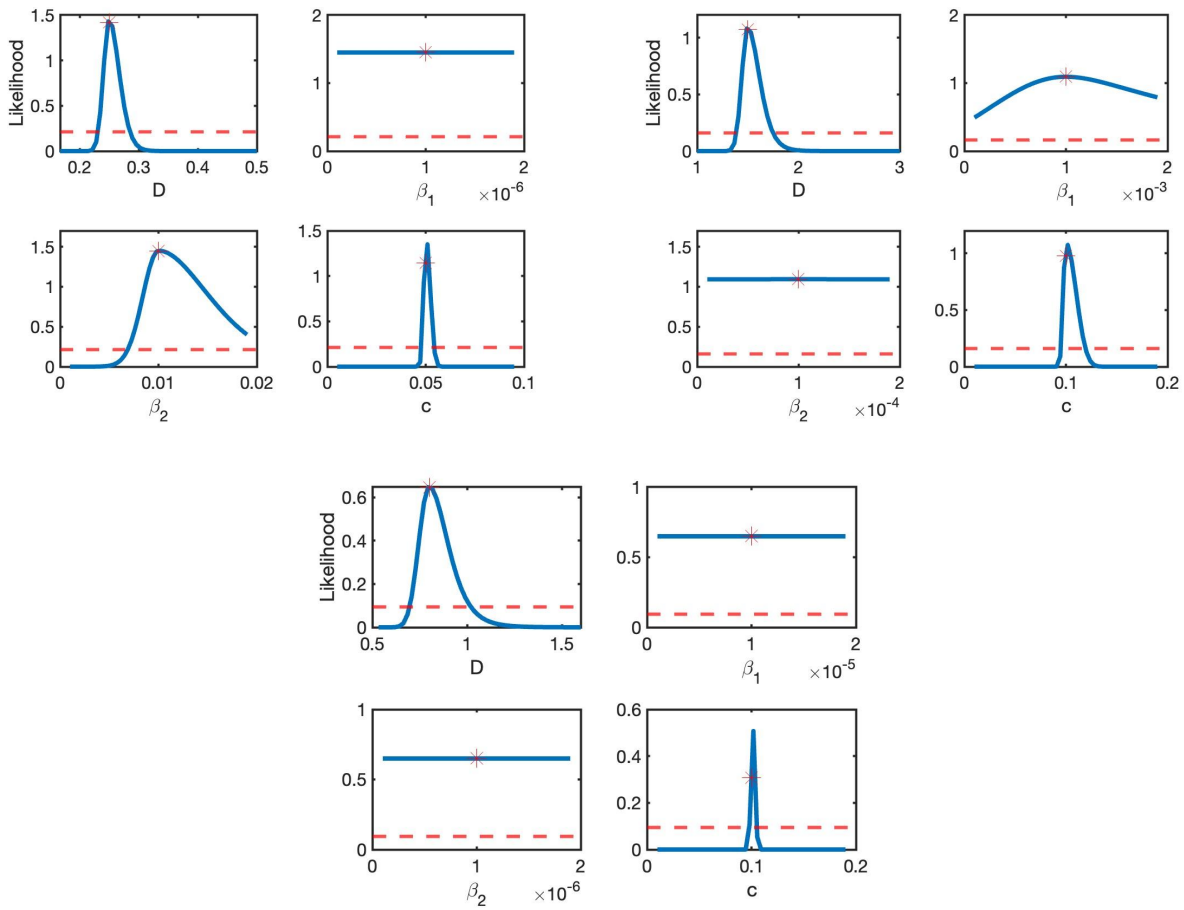

Figure 6. *Profile likelihood results for the three regions given noiseless FRAP data synthetically generated using Equation (2.1) and the parameter sets in Table 2. Baseline values are indicated as red stars in each plot (top-left: Region 1, top-right: Region 2, bottom: Region 3).*

the synthetic data curve generated under optimized values of β_1 and β_2 and the baseline value curve (red curves shown in Figure 4). The likelihood value is then derived by Equation (3.1). It can be seen from Figure 7 that the profile likelihood value achieves its peak around the baseline values for c and D as shown in Table 2. A transparent plane is added in each plot, which represents the threshold calculated using Equation (3.8).

Overall, we find that both c and D are identifiable, which is consistent with the one-dimensional profile likelihood results. The estimated transport velocity c and diffusion coefficient D are in line with biologically plausible values reported in FRAP studies of RNA in *Xenopus laevis* oocytes [2], supporting the biological relevance of our parameter estimates and indicating that our model captures key features of intracellular RNA transport dynamics. Two-dimensional profile likelihood thus offers a valuable tool for analyzing experimental data by providing both the range of plausible measurement outcomes for a given experiment and their impact on the parameter likelihood profile [9, 7]. However, a limitation of this approach is that it is computationally more intensive compared to one-dimensional analyses, especially when applied to models with many parameters. Nevertheless, our results demonstrate that for the FRAP data and cellular regions considered here,

the transport speed c and diffusion coefficient D can be assumed to be identifiable.

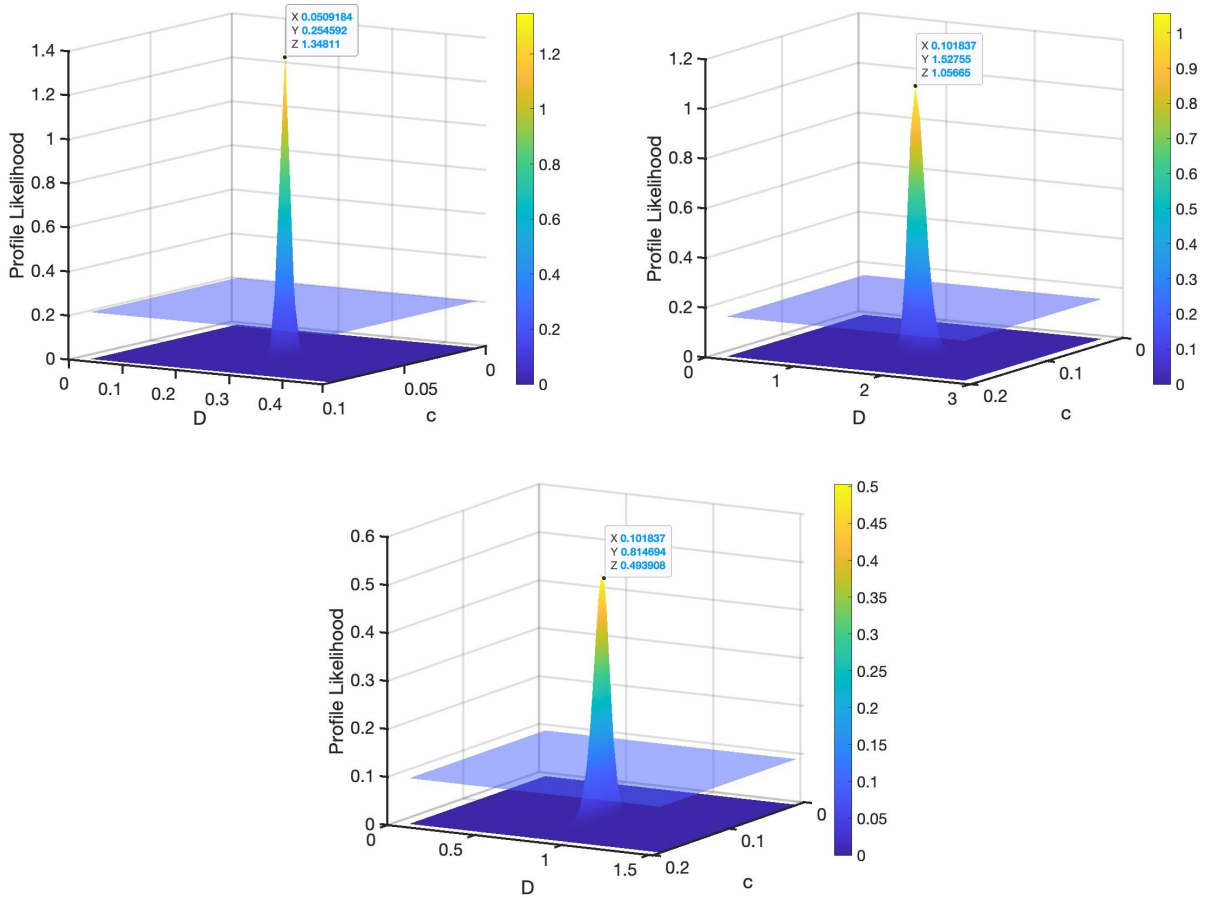

Figure 7. *Two-dimensional profile likelihood given each possible pair of (c, D) chosen around their baseline values in each region, confirming the identifiability of transport velocity and diffusion coefficient. The z-axis shows the corresponding profile likelihood values calculated using Equation (3.1), together with a threshold plane based on Equation (3.8) (top-left: Region 1, top-right: Region 2, bottom: Region 3).*

4. Investigating parameter relationships in transport models of FRAP experiments.

Here, we explore compensatory relationships between non-identifiable parameters using subset profiles, contour plots, and slope vector field visualization. Moreover, we introduce a reparametrization approach to derive identifiable parameter combinations in regions where binding and unbinding rates are correlated.

4.1. Subset profiles and 3D contour plots.

According to the results derived from the one-dimensional profile likelihood, we find that c and D are identifiable, while β_1 and β_2 are non-identifiable. Then, we expand the methodology in [1] in order to investigate potential parameter relationships between β_1 and β_2 that may be identifiable.

We first begin by drawing the subset profiles based on the data from profile likelihood for the three regions. Following the approach in previous studies [1], we treat β_1 as the interest parameter in each subset profile, where fixing a single parameter while re-optimizing the remaining parameters allows

for a clear visualization of compensatory relationships and aids in assessing parameter identifiability. For each fixed value of β_1, we record the corresponding optimal value of β_2 that maximize the likelihood value. This allows us to visualize the compensatory relationship between these two kinetic parameters [1]. As shown in Figure 8, the x-axis refers to the values for interest parameter β_1, and the y-axis denotes the corresponding optimal values for β_2.

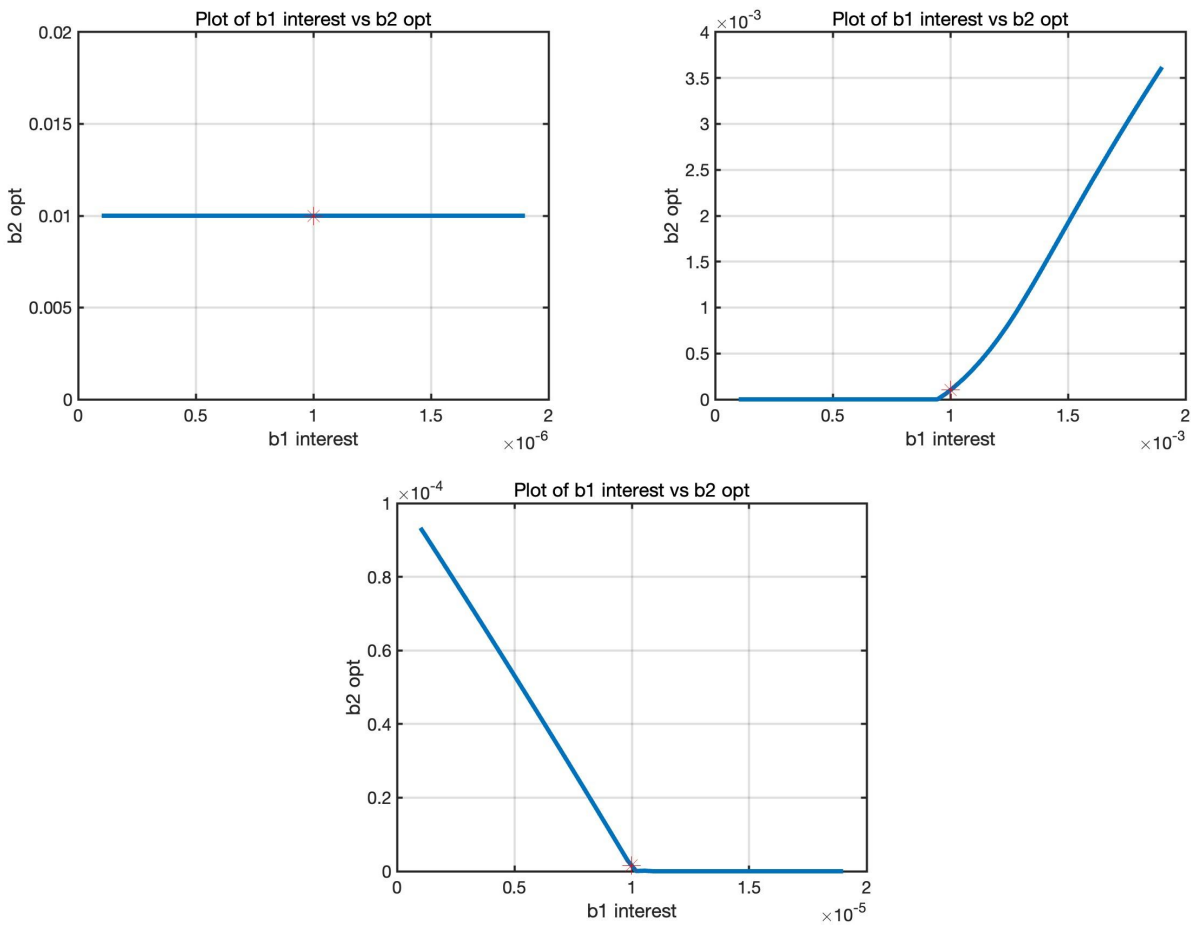

Figure 8. *Subset profiles for interest parameter β_1 on the x-axis and the corresponding optimized β_2 on the y-axis for each region given noiseless FRAP data synthetically generated using Equation (2.1) and the parameter set in Table 2. Baseline values are indicated as red stars in each plot (top-left: Region 1, top-right: Region 2, bottom: Region 3).*

Figure 8 shows that there could be some linear relationships between β_1 and β_2 in some regions of parameter space. The slopes of these lines, which quantify the compensatory relationship between the two parameters, will be further utilized in Figure 10.

To better interpret the relationships between β_1 and β_2, our goal is to explore the likelihood landscape by varying β_1 and β_2 in a grid around their baseline values in each region under fixed values for c and D (set to their baseline values in Table 2). We then carry out the forward PDE computation by plugging the four chosen parameter values into PDE model (Equation (2.1)) to obtain another synthetic data curve. For each point in the β_1-β_2 plane, there is one corresponding least square error by calculating the difference between the synthetic data curve and the baseline

curve (red curves shown in Figure 4). Figure 9 shows the contour plot of this error as a function of (β_1, β_2) for the three regions. We notice that the least square error is minimized along a line (highlighted in red). Therefore, we cannot distinguish parameters (β_1, β_2) on this red line, as they all give the minimized error function [1].

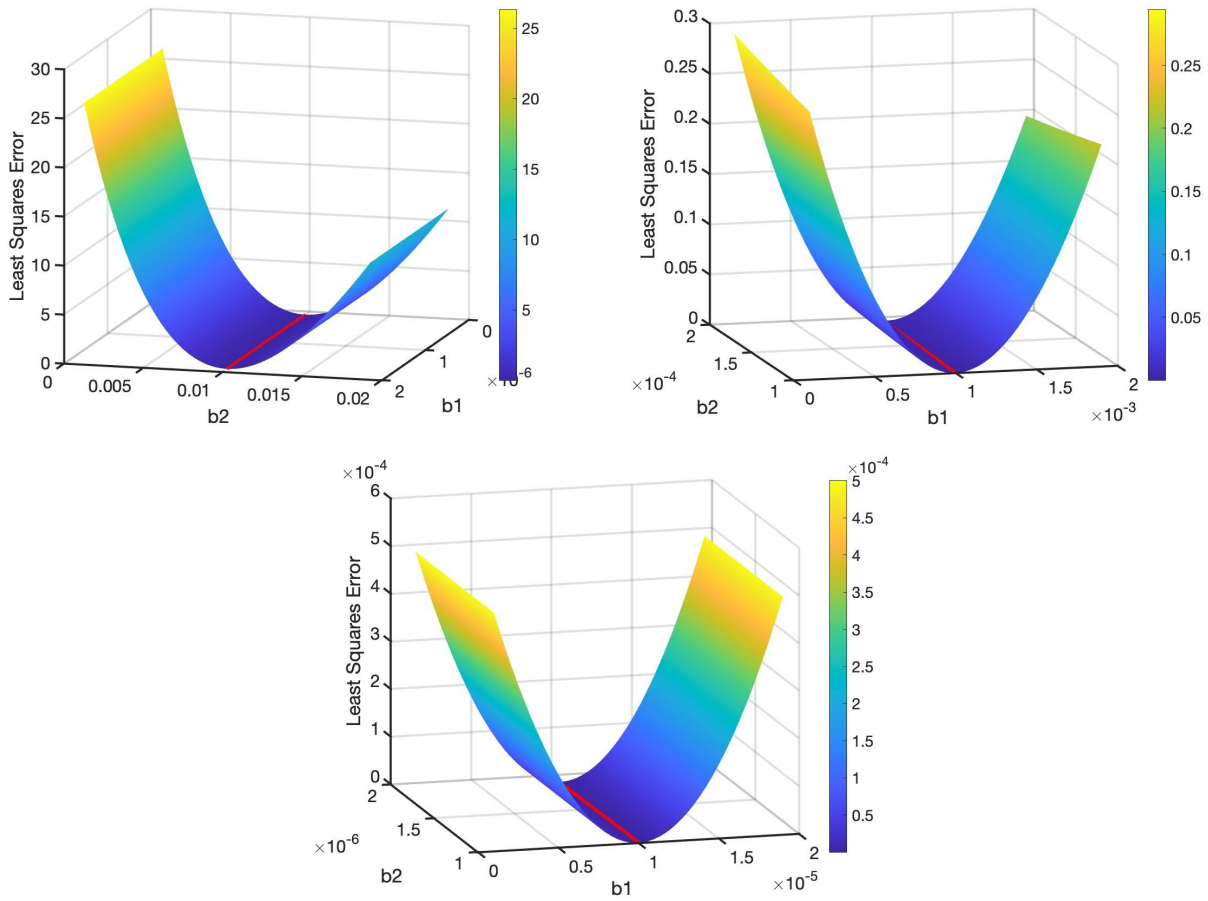

Figure 9. *Contour plots of the least square error between synthetic FRAP data generated using baseline values of c and D in Table 2, as well as β_1 and β_2 values from a grid around their baseline values for each region (top-left: Region 1, top-right: Region 2, bottom: Region 3). The least square error is minimized along a line (highlighted in red), on which we cannot distinguish parameters (β_1, β_2), as they all give the minimized error function.*

4.2. Slope vector field generation. In order to find the corresponding relationships between β_1 and β_2 in Region 1, Region 2 and Region 3, we generate a slope vector field as in [1]. In the slope vector field, each vector represents the slope in the subset profiles showing the relationship between β_1 and β_2 (Figure 8). The goal is to draw a contour curve in the vector field for each region, representing the combination of β_1 and β_2 that yields the maximum profile likelihood value.

Since the diffusion coefficient D and transport speed c are identifiable, we set them to be their baseline values in Table 2. Then, we select a grid in $(log_{10}\beta_1, log_{10}\beta_2)$-plane. For each point in the plane, we generate a synthetic FRAP data set based on Equation (2.1) and the chosen parameter values. For each synthetic data set, we compute the profile likelihood for β_1 and β_2, then set the tangent vector at each point to be the slope of the subset profile shown in Figure 8. The pattern

of the slope vector field for each region is shown in Figure 10.

We then choose a curve τ going transversely across the slope vector field, meaning that for each point on this curve, its tangent vector is not aligned with the direction of the vector field. We can either choose the curve τ in explicit analytical form or else using linear interpolation and a forward Euler scheme applied to gradients of the vector to compute such a curve numerically ([1]). Here, we use an explicit analytical parametrization similar to the one in [1]:

$$\begin{aligned} \log_{10}\beta_1 &= s + \sqrt{s^2+1} - 6 \\ \log_{10}\beta_2 &= -s + \sqrt{s^2+1} - 6 \end{aligned} \tag{4.1}$$

This reparametrization yields the yellow curve for each region shown in Figure 10. Moreover, in each plot, the point P corresponds to the baseline values of β_1 and β_2 in each region.

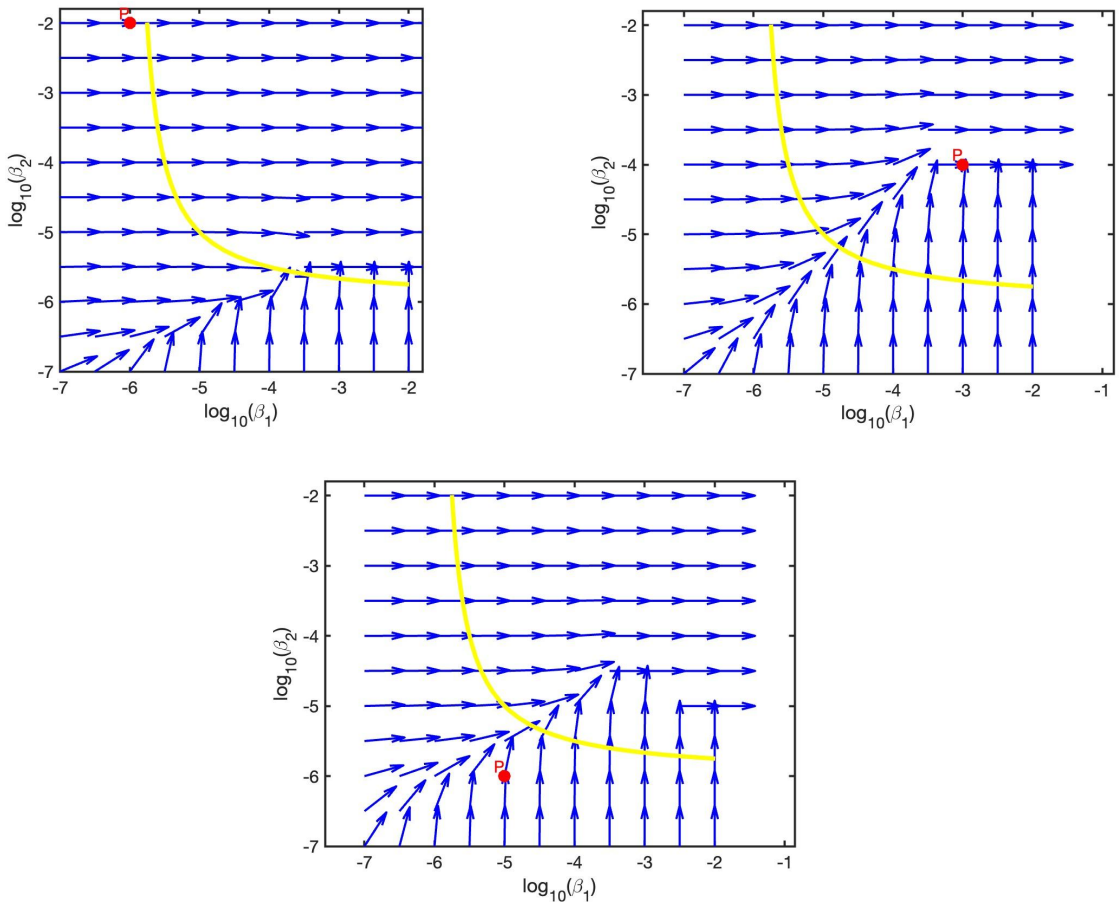

Figure 10. *Slope vector field g eneration: the slope of each arrow corresponds to the slope in the subset profile for each profile l ikelihood. A yellow curve τ crosses transversely the slope vector fi eld, with po int $P = (\beta_1^*, \beta_2^*)$ indicating the baseline values for β_1 and β_2 (top-left: Region 1, top-right: Region 2, bottom: Region 3).*

Then, we apply the profile likelihood analysis for the three parameters (c, D, s). We especially want to see whether the new parameter s is identifiable or not. First, both c and D are identifiable,

which is consistent with the results shown in Figure 6. As expected, in Figure 11 and Figure 12, the clear peaks of the s-profile likelihood for Region 1 and Region 2 demonstrate that the re-parametrized parameter s is indeed identifiable. Next, we plug the optimized s value at the peak into Equation (4.1) to get the values for β_1 and β_2, and that will correspond to a point Q on the curve τ. By the method of linear interpolation, we then generate a green contour curve which begins from Q. For each point on this curve, the profile likelihood function is maximized. Therefore, this green contour curve represents the relationship between β_1 and β_2 that is identifiable in both regions. Notably, the ground-truth point p is very close to this contour curve.

For Region 3, the likelihood value is maximized along a line instead of one peak as shown in Figure 13. Therefore, we plug in some s values on that line into the reparametrization formula, and that corresponds to points Q_1, Q_2, Q_3 on the curve τ. Based on each Q, we then generate the corresponding contour curves. Since we know that the ground-truth point P must be very close to the contour curve, the green and the purple curve beginning from Q_1 and Q_2 are more likely to bound the relationship between β_1 and β_2. However, we cannot determine a contour curve that can best represent the relationship between β_1 and β_2 in this case, since there is not a single peak as depicted in the s-profile likelihood plot. This means that, for region 3, we are not able to determine an identifiable correlation for β_1 and β_2 without additional information.

4.3. Pipeline for identifiability analysis and parameter relationship investigation. Here, we outline the systematic workflow employed to analyze parameter identifiability and explore relationships among non-identifiable parameters. The pipeline consists of four key steps that progressively refine our understanding of the model parameters. We will apply this pipeline to real experimental data in Section 5.

1. **Profile likelihood computation**
   The first step assesses identifiability of individual parameters using synthetic FRAP data. For each parameter of interest (e.g., c, D, β_1, β_2), we fix its value across a grid around the baseline values (Table 2) while optimizing the remaining nuisance parameters to maximize the profile likelihood (Equation (3.1)). We establish the 95% confidence threshold (Equation (3.8)) to judge identifiability. This reveals that identifiable parameters like c and D exhibit sharp peaks in their likelihood profiles, while non-identifiable parameters (e.g., β_1, β_2) show flat or plateaued profiles (Figure 6).
2. **Two-dimensional profile likelihood exploration**
   The second step validates pairwise identifiability of parameters c and D. We define a grid of (c, D) values around baseline estimates (Table 2) and for each pair, optimize β_1 and β_2 to maximize likelihood (Equation (3.9)). The resulting likelihood landscape visualized as a 3D surface (Figure 7) shows a clear peak near baseline values, confirming that c and D are jointly identifiable.
3. **Contour interpretation for non-identifiable parameters**
   The third step investigates relationships between β_1 and β_2 when c and D are fixed at baseline values. We compute the least square error (LSE) for a (β_1, β_2) grid and identify the contour line where the LSE is minimized (Figure 9). This reveals a correlation between

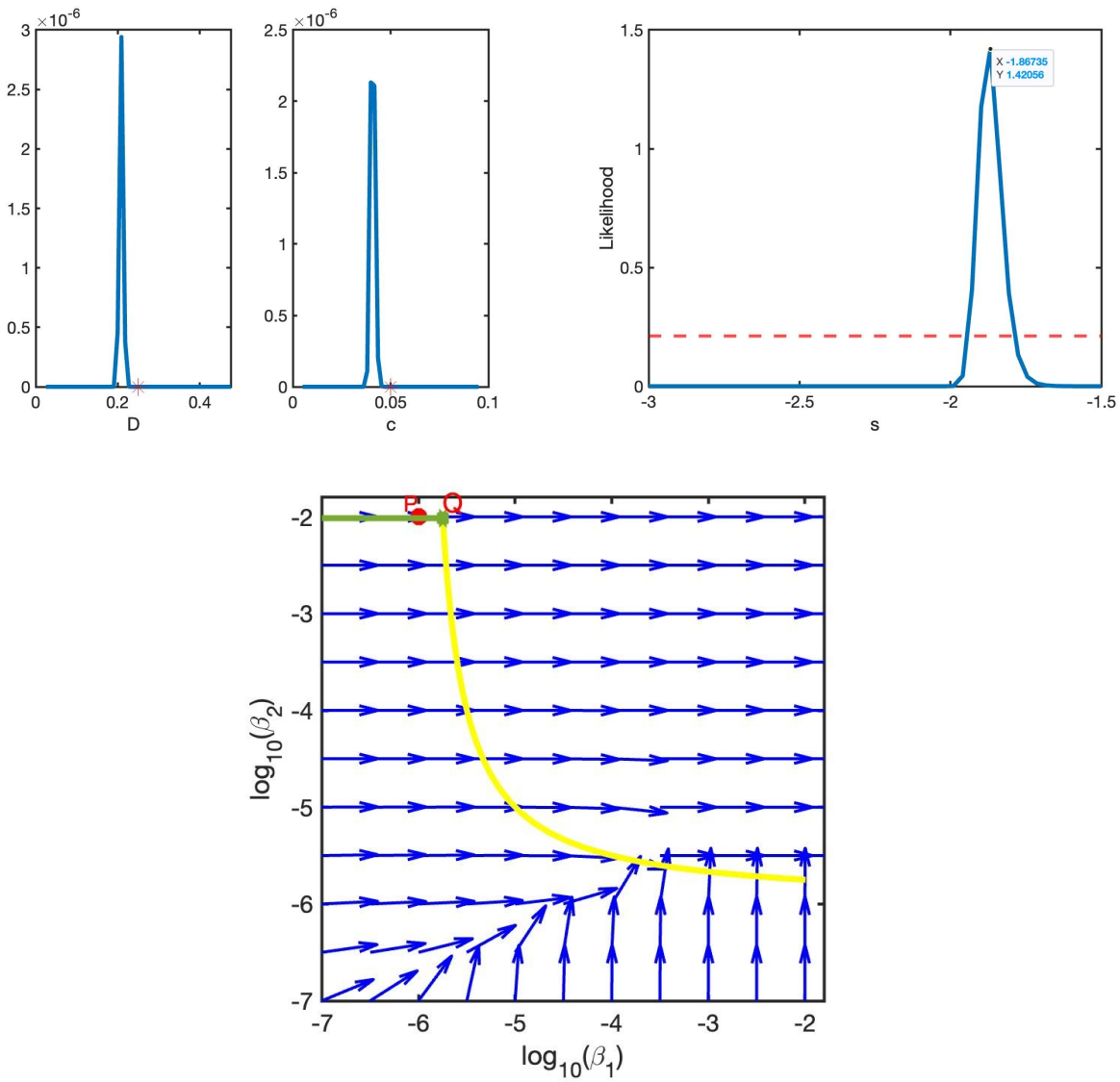

Figure 11. *Profile likelihoods for each interest parameter c, D, and s in Region 1 given noiseless FRAP data generated using the parameters in Table 2 (top-left and top-right). s achieves the maximum at $s^* = -1.86735$, corresponding to the point $Q = (\beta_1^{**}, \beta_2^{**})$ on the yellow curve τ, as well as the trace of the error-minimizing green contour curve (bottom).*

β_1 and β_2 along the minimized LSE line, indicating structural non-identifiability due to compensatory effects between these parameters.

4. **Slope vector field and reparametrization**

   The final step derives identifiable combinations of β_1 and β_2:

   - For subset profile construction, for each β_1 value on a logarithmic grid, we compute the optimal β_2 that maximizes the profile likelihood while fixing c and D at baseline values, generating subset profiles (Figure 8).
   - We construct a slope vector field (Figure 10) where arrows indicate the direction of local β_1-β_2 correlations.
   - For transverse curve parametrization, we introduce a new parameter s and define

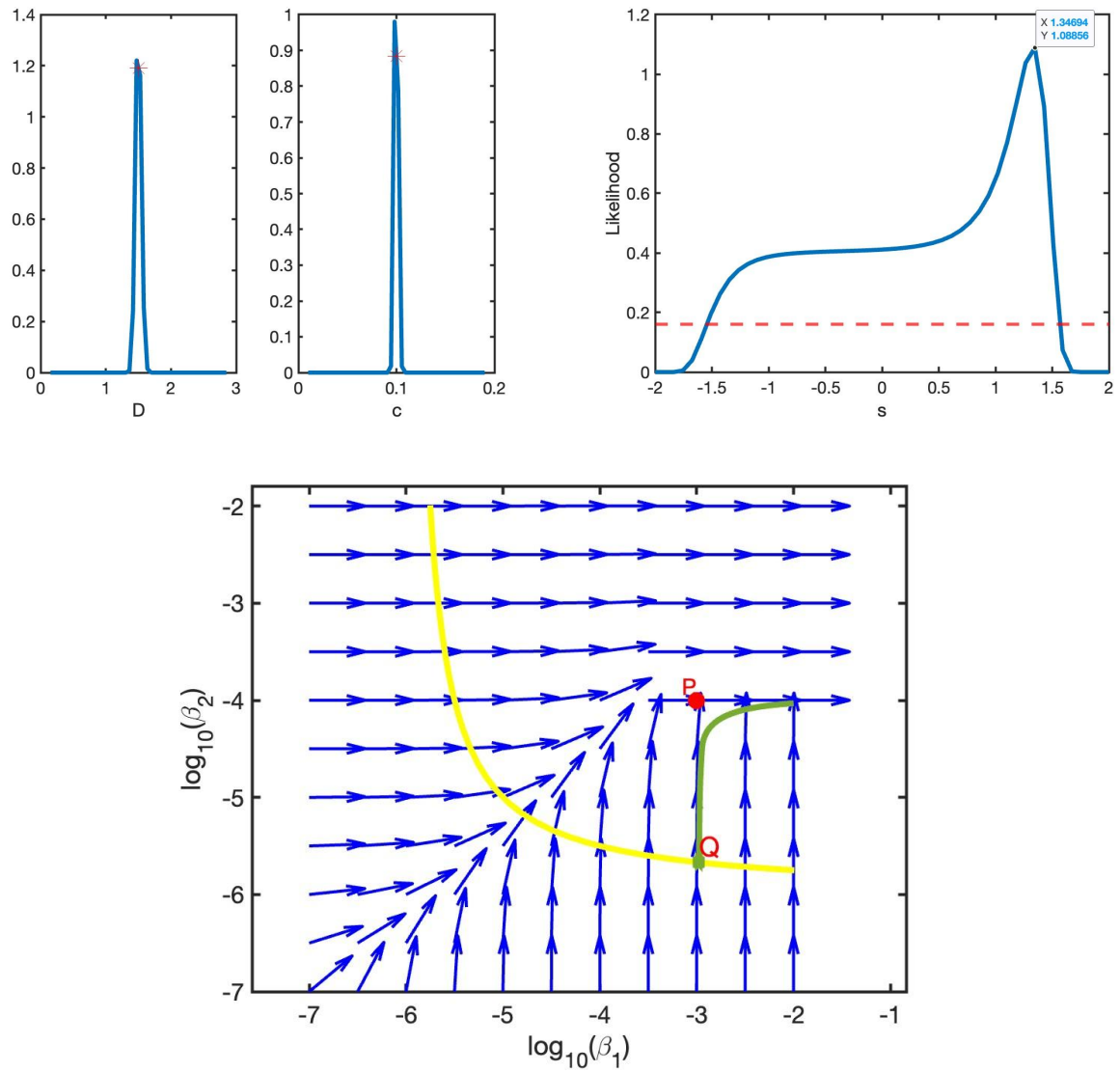

Figure 12. *Profile likelihoods for each interest parameter c, D and s in Region 2 given noiseless FRAP data generated using the parameters in Table 2 (top-left and top-right). s achieves the maximum at $s^* = 1.34694$, corresponding to the point $Q = (\beta_1^{**}, \beta_2^{**})$ on the yellow curve τ, as well as the trace of error-minimizing green contour curve (bottom).*

an analytic curve τ transverse to the vector field (Equation (4.1)), marking baseline points $P = (\beta_1^*, \beta_2^*)$ from Table 2.

- We compute the profile likelihood for c, D, s, identify the peak s^* yielding maximum likelihood, and obtain the corresponding point $Q = (\beta_1^{**}, \beta_2^{**})$ through Equation (4.1). Linear interpolation is then used to draw the curves beginning from Q, which illustrate the relationship between β_1 and β_2 (Figure 11, Figure 12, Figure 13).

The key insights for our data are that for Regions 1 and 2, the unique peak at s^* generates a contour curve through Q representing all optimal (β_1, β_2) pairs. In Region 3, the flat likelihood profile for s indicates no unique optimum.

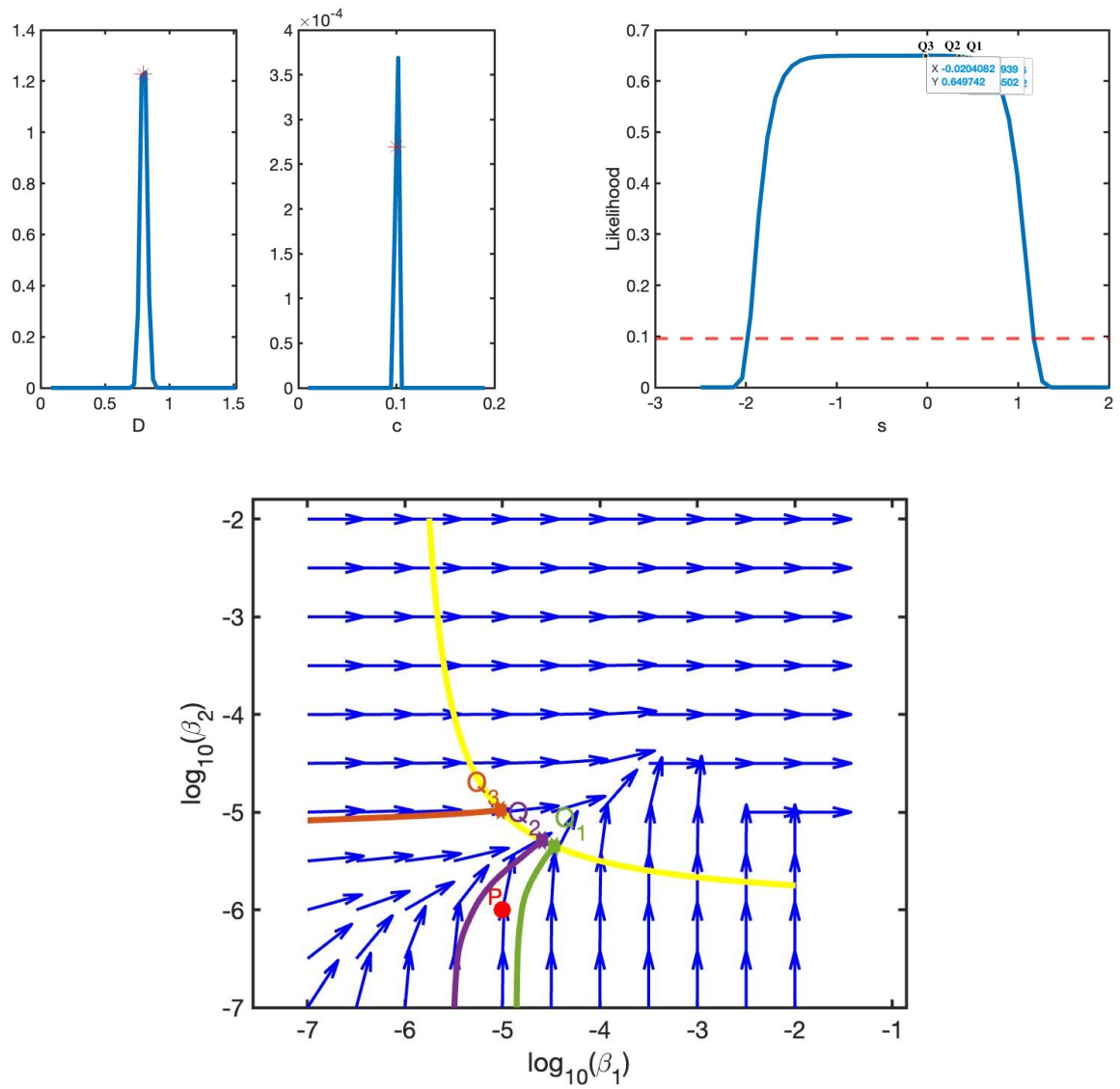

Figure 13. *Profile likelihoods for each interest parameter c, D, and s in Region 3 given noiseless FRAP data generated using the parameters in Table 2 (top-left and top-right). Choosing three values for s: $s^* = 0.438776, 0.346939, -0.0204082$ corresponds to points $Q_1 = (\beta_{1,1}^{**}, \beta_{2,1}^{**})$, $Q_2 = (\beta_{1,2}^{**}, \beta_{2,2}^{**})$, $Q_3 = (\beta_{1,3}^{**}, \beta_{2,3}^{**})$ on the yellow curve τ, as well as the traces of error-minimizing contour curves (bottom).*

5. Application to experimental FRAP datasets. After validating this approach for synthetic FRAP datasets, we apply the outlined pipeline to the real average experimental FRAP data for each region. In particular, we generate the vector field and the corresponding contour curves using linear interpolation, but this time we use the real experimental FRAP data (blue points as shown in Figure 4), and we set our baseline values as shown in Table 1. The pattern of the slope vector field and the contour curves in this case are given in Figure 14, Figure 15 and Figure 16, which look very similar to those under our chosen baseline values, since the parameter values in Table 1 are very similar to those in Table 2.

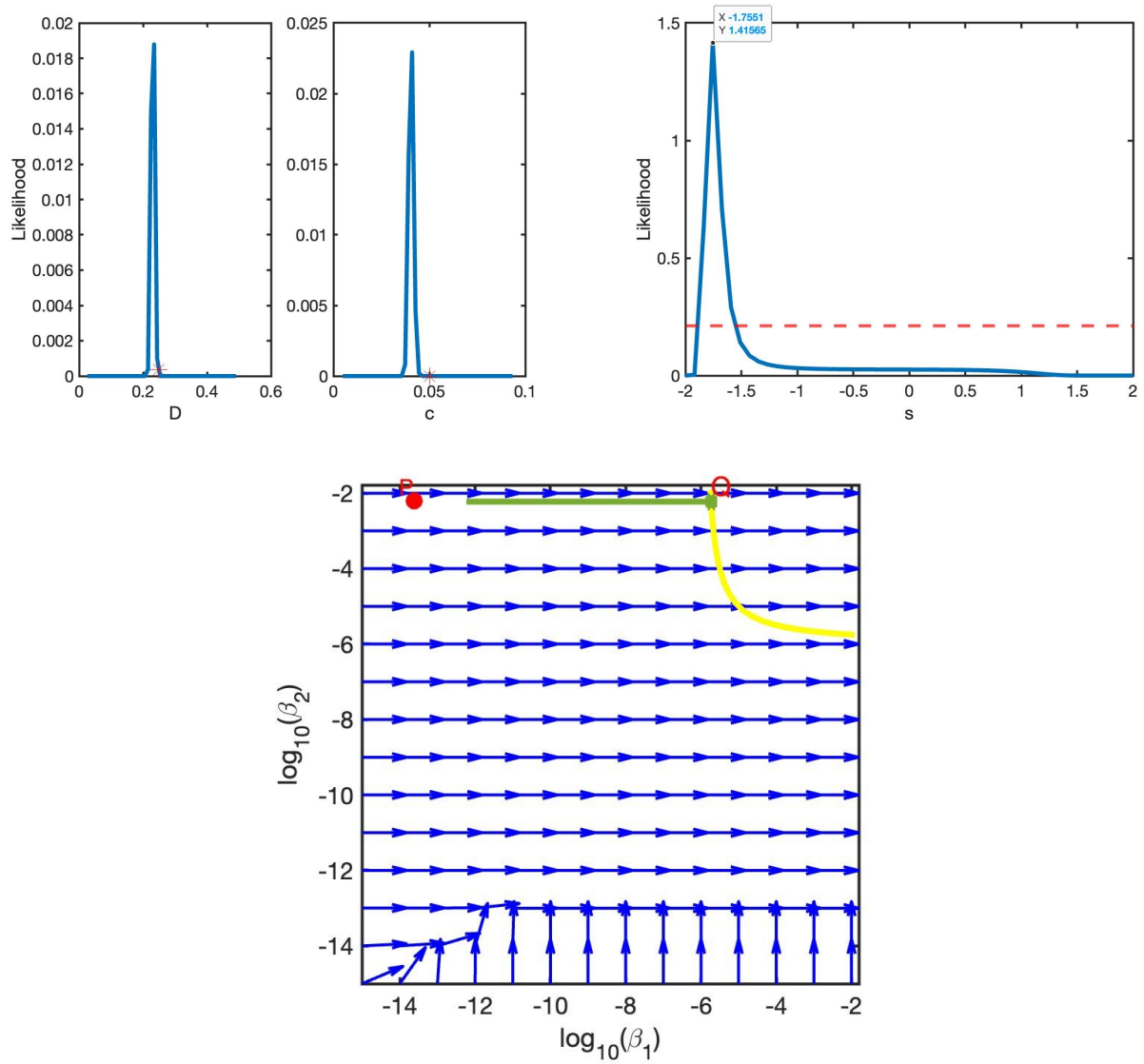

Figure 14. *Profile likelihoods for each interest parameter c, D and s in Region 1 given real experimental FRAP data in Figure 4 (top-left and top-right). s achieves the maximum at $s^* = -1.7551$, corresponding to the point $Q = (\beta_1^{**}, \beta_2^{**})$ on the yellow curve τ, as well as the trace of error-minimizing green contour curve (bottom).*

6. Discussion. Comparing the identifiability results that we have obtained from profile likelihood (Figure 6) for β_1 and β_2 and the contour curves representing the relationship between β_1 and β_2 in slope vector fields (Figure 11, Figure 12 and Figure 13), we observe the difference between practical non-identifiability and structural non-identifiability. For Region 1, it can be seen from the flat contour curve that although the value of β_1 is not unique, β_2 is fixed at 10^{-2}. This is consistent with the likelihood profile graphs for β_1 and β_2, where β_2 is almost identifiable except that the right end of the curve does not go below the threshold. Similar conclusions also hold for Region 2, where β_1 is fixed at 10^{-3} for the main part of the green contour curve, coinciding with β_1 being more identifiable than β_2 as shown in the profile likelihood. On the other hand, for Region 3, we see from the profile likelihood curve that both β_1 and β_2 are structurally non-identifiable.

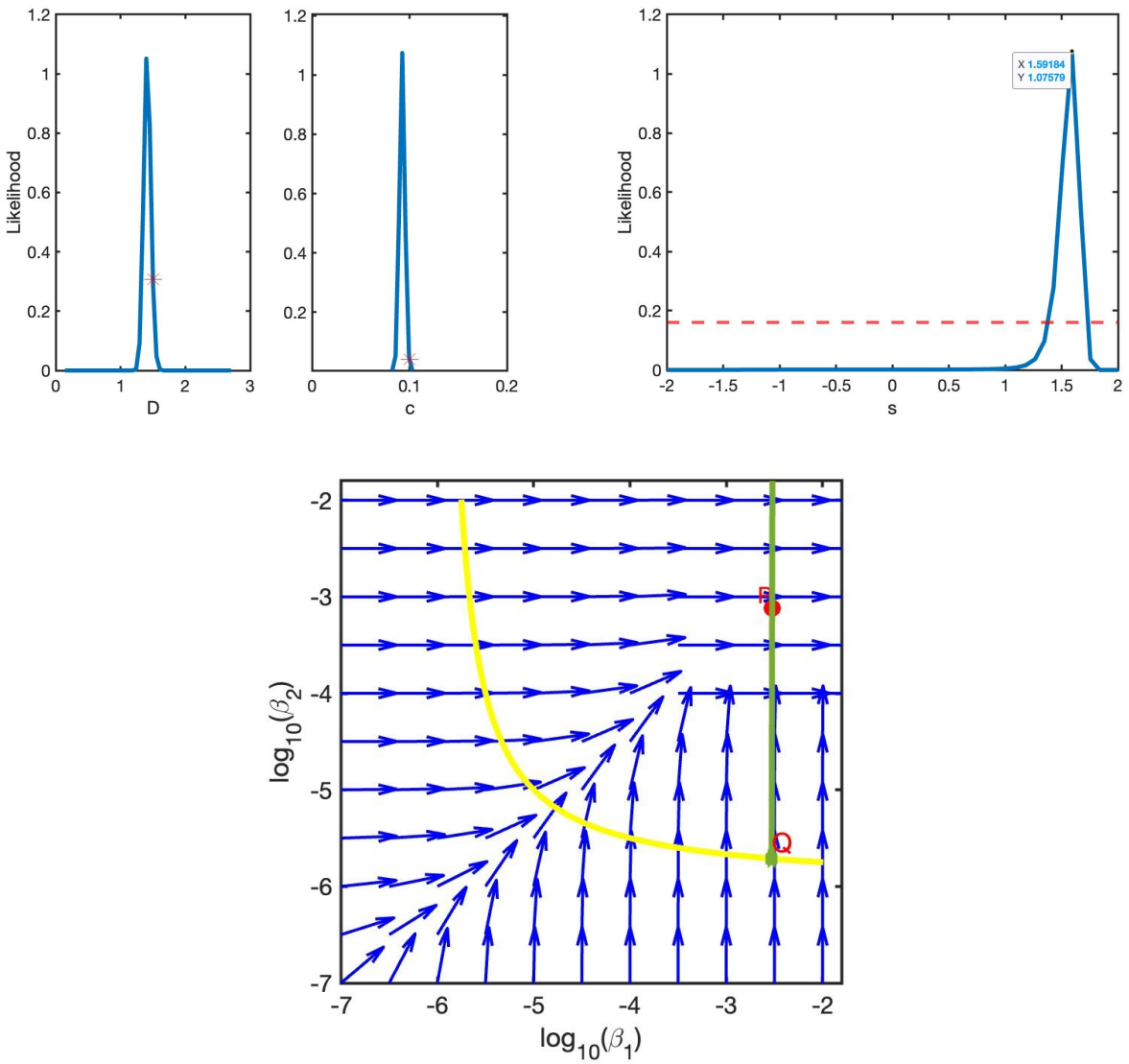

Figure 15. *Profile likelihoods for each interest parameter c, D and s in Region 2 given real experimental FRAP data in Figure 4 (top-left and top-right). s achieves the maximum at $s^* = 1.59184$, corresponding to the point $Q = (\beta_1^{**}, \beta_2^{**})$ on the yellow curve τ, as well as the trace of error-minimizing green contour curve (bottom).*

Similarly, we find that the re-parameterized s parameter is not identifiable in Region 3, making it difficult to determine a best contour curve representing the relationship between β_1 and β_2.

Our findings build upon and extend the identifiability analysis conducted in prior work on a simpler reaction-diffusion model for RNA-binding protein dynamics [1]. A core similarity between the studies is the non-identifiability of individual binding and unbinding rates (β_1, β_2) when inferred from experimental FRAP data, highlighting a fundamental limitation of FRAP for disentangling specific on and off rates within complex transport systems. The primary advancement of this work lies in the explicit integration of active transport (advection) into the PDE framework, creating a more comprehensive reaction-diffusion-advection model. This expansion yields a significant difference: the transport velocity c emerges as a clearly identifiable parameter alongside the diffusion

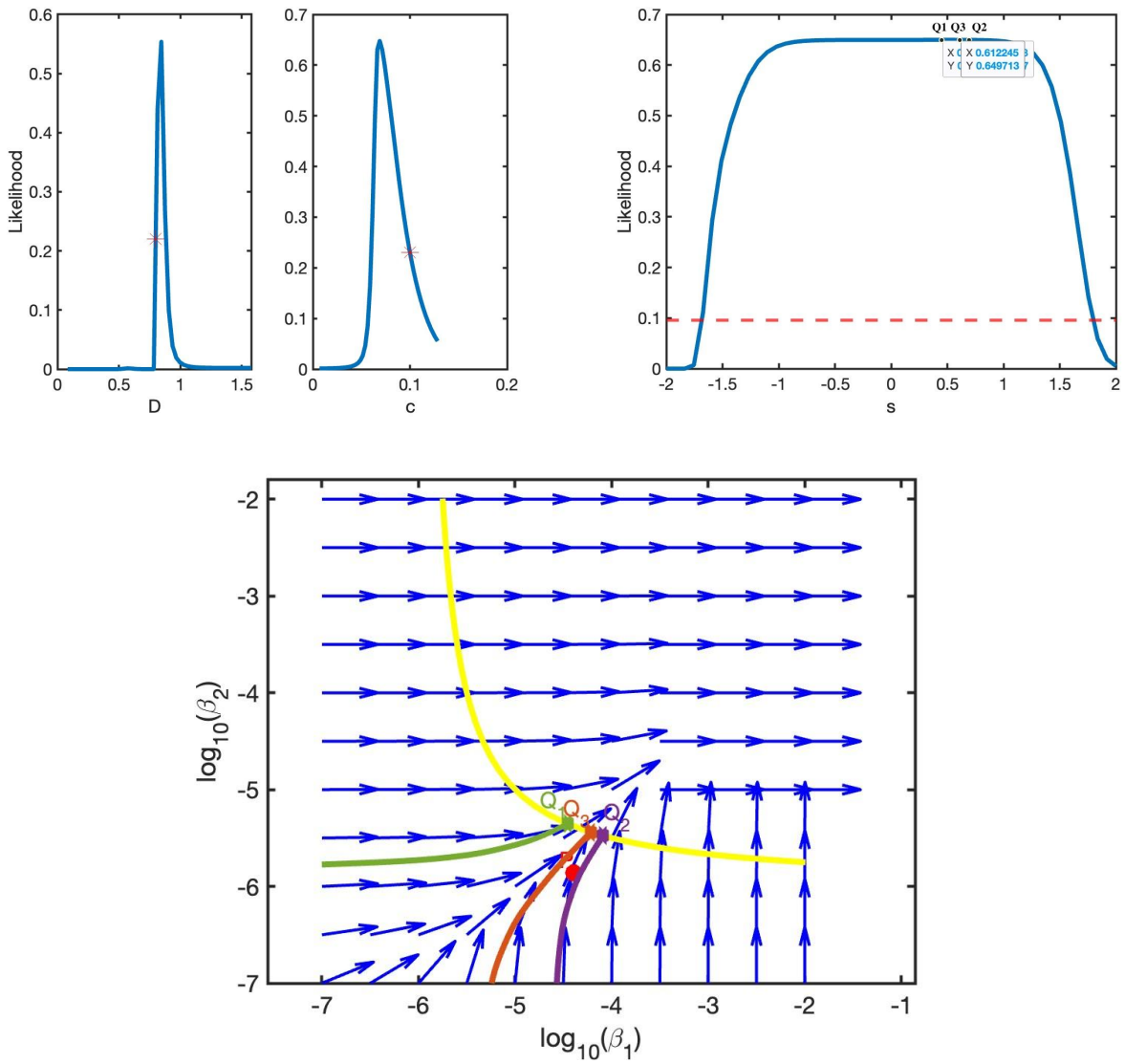

Figure 16. *Profile likelihoods for each interest parameter c, D and s in Region 3 given real experimental FRAP data in Figure 4 (top-left and top-right). Choosing three values for s: $s^* = 0.44898, 0.693878, 0.612245$ corresponds to points $Q_1 = (\beta^{**}_{1,1}, \beta^{**}_{2,1})$, $Q_2 = (\beta^{**}_{1,2}, \beta^{**}_{2,2})$, $Q_3 = (\beta^{**}_{1,3}, \beta^{**}_{2,3})$ on the yellow curve τ, as well as the traces of error-minimizing contour curves (bottom).*

coefficient D, in contrast to the purely diffusive model where transport mechanisms were not represented. Furthermore, while both studies reveal compensatory relationships between β_1 and β_2, the spatial analysis across oocyte regions in this work shows that identifiable combinations of these rates are context-dependent, being resolvable near the nucleus but not near the vegetal cortex.

This spatial heterogeneity in identifiability carries direct implications for the biological interpretation of RNA dynamics within the *Xenopus laevis* oocyte. Near the nucleus, where RNA trafficking activity is more pronounced, combinations of binding and unbinding rates become identifiable, suggesting stronger coupling between RNAs and the transport machinery. In contrast, farther from the nucleus and near the vegetal cortex, RNA dynamics are dominated by transport and diffusion,

making binding-related parameters more difficult to resolve from FRAP data alone.

In the context of FRAP experiments and PDE-based transport models, identifying which parameters can be uniquely estimated is crucial for linking quantitative model outputs to molecular mechanisms such as diffusion, active transport, and binding interactions. Being able to uniquely estimate parameters such as diffusion coefficient D and transport velocity c means that these values can be reliably connected to underlying RNA mobility and localization mechanisms. In contrast, the non-identifiability of binding and unbinding rates (β_1, β_2) indicates limitations of FRAP-based inference in estimating these parameters and suggests that additional experimental strategies may be needed. While our three-region framework offers an initial spatial mapping of identifiability, future work could adopt a finer discretization with more regions to further clarify whether observed identifiability transitions reflect biological variation. Overall, our analysis provides a rigorous foundation for using PDE models not only to fit experimental data but also to generate mechanistic insight into how RNAs are localized and regulated within cells.

Acknowledgments. I would like to extend my heartfelt gratitude to my mentor, Professor Veronica Ciocanel, for her invaluable guidance, insightful feedback, and unwavering support throughout the course of my research.

REFERENCES

[1] Maria-Veronica Ciocanel, Lee Ding, Lucas Mastromatteo, Sarah Reichheld, Sarah Cabral, Kimberly Mowry, and Björn Sandstede. Parameter identifiability in pde models of fluorescence recovery after photobleaching. *Bulletin of Mathematical Biology*, 86(4):36, 2024.

[2] Maria-Veronica Ciocanel, Jill A Kreiling, James A Gagnon, Kimberly L Mowry, and Björn Sandstede. Analysis of active transport by fluorescence recovery after photobleaching. *Biophysical journal*, 112(8):1714–1725, 2017.

[3] Steven M Cox and Paul C Matthews. Exponential time differencing for stiff systems. *Journal of Computational Physics*, 176(2):430–455, 2002.

[4] Anna Git and Nancy Standart. The kh domains of xenopus vg1rbp mediate rna binding and self-association. *Rna*, 8(10):1319–1333, 2002.

[5] Aly-Khan Kassam and Lloyd N Trefethen. Fourth-order time-stepping for stiff pdes. *SIAM Journal on Scientific Computing*, 26(4):1214–1233, 2005.

[6] Mary Lou King, Timothy J Messitt, and Kimberly L Mowry. Putting rnas in the right place at the right time: Rna localization in the frog oocyte. *Biology of the Cell*, 97(1):19–33, 2005.

[7] Clemens Kreutz, Andreas Raue, Daniel Kaschek, and Jens Timmer. Profile likelihood in systems biology. *The FEBS journal*, 280(11):2564–2571, 2013.

[8] Jennifer Lippincott-Schwartz, Erik Snapp, and Anne Kenworthy. Studying protein dynamics in living cells. *Nature reviews Molecular cell biology*, 2(6):444–456, 2001.

[9] Tim Litwin, Jens Timmer, and Clemens Kreutz. Optimal experimental design based on two-dimensional likelihood profiles. *Frontiers in Molecular Biosciences*, 9:800856, 2022.

[10] Kimberly L Mowry. Using the xenopus oocyte toolbox. *Cold Spring Harbor Protocols*, 2020(4):pdb–top095844, 2020.

[11] Kimberly L Mowry and Douglas A Melton. Vegetal messenger rna localization directed by a 340-nt rna sequence element in xenopus oocytes. *Science*, 255(5047):991–994, 1992.

[12] Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

[13] Eric AJ Reits and Jacques J Neefjes. From fixed to frap: measuring protein mobility and activity in living cells. *Nature cell biology*, 3(6):E145, 2001.

[14] Brian L Sprague and James G McNally. Frap analysis of binding: proper and fitting. *Trends in cell biology*, 15(2):84–91, 2005.

[15] Jian Zhang and Mary Lou King. Xenopus vegt rna is localized to the vegetal cortex during oogenesis and encodes a novel t-box transcription factor involved in mesodermal patterning. *Development*, 122(12):4119–4129, 1996.