Supplementary Material
Parameter identifiability and model selection for PDE models of cell invasion

Yue Liu[1*], Kevin Suh[2], Philip K. Maini[1], Daniel J. Cohen[2,3], Ruth E. Baker[1]
[1]Mathematical Institute, University of Oxford
[2]Department of Chemical and Biological Engineering, Princeton University
[3]Department of Mechanical and Aerospace Engineering, Princeton University
* Corresponding author: yue.liu@maths.ox.ac.uk

1 Numerical methods

This section provides details of the numerical methods for model simulation and calculation of the profile likelihoods.

1.1 Numerical solutions of the PDE models

We use a finite difference method to simulate the general form of the model, given in Eq. (1). For simulations in two spatial dimensions, the size of the domain, corresponding to the size of the image, is $L_x = L_y = 4380$ µm. This domain is discretized into $n_x = 150$ by $n_y = 150$ squares, each with side length $\Delta x = \Delta y = 29.2$ µm. We used $\Delta t = 1/30$ h.

Let $C_{i,j,k}$ denote $C(x_i, y_j, t_k)$, where $x_i = (i-1)\Delta x$, $y_j = (j-1)\Delta y$, $t_k = (k-1)\Delta t$ are the mesh points. The scheme we used follows = [6], and can be written as follows:

$$\begin{aligned}\frac{\partial}{\partial x}\left[D(C_{i,j,k})\frac{\partial C_{i,j,k}}{\partial x}\right] &\approx \frac{1}{\Delta x}\left[D(C_{i+1/2,j,k})\frac{C_{i+1,j,k}-C_{i,j,k}}{\Delta x} - D(C_{i-1/2,j,k})\frac{C_{i,j,k}-C_{i-1,j,k}}{\Delta x}\right] \\ &= \frac{1}{(\Delta x)^2}\Big[D(C_{i+1/2,j,k})C_{i+1,j,k} - (D(C_{i+1/2,j,k}) + D(C_{i-1/2,j,k}))C_{i,j,k} \\ &\qquad + D(C_{i-1/2,j,k})C_{i-1,j,k}\Big], \end{aligned} \tag{SM.1}$$

where

$$D(C_{i+1/2,j,k}) = \frac{1}{2}\Big[D(C_{i,j,k}) + D(C_{i+1,j,k})\Big], \quad 1 \le i,j \le n_x = n_y = 150, \quad 1 \le k \le n_t = 77.$$

The discretization in the y direction is completely analogous. Zero flux boundary conditions are imposed at $x = 0, L_x$ and $y = 0, L_y$.

We use an implicit-explicit (IMEX) scheme [1, 4] for time-stepping, where the nonlinear diffusion coefficient, $D(C)$, and the proliferation term, $f(C)$, are treated with the explicit Euler method, and the diffusion term overall is treated with the implicit Crank-Nicolson method, which has second order convergence. The advantage of this scheme is that the explicit treatment of the nonlinear components of the equation allows us to avoid having to solve a nonlinear root-finding problem at every time step, which would be necessitated by a fully implicit scheme. The implicit treatment of the diffusion term improves the stability of the scheme, and [1] showed that this class of schemes has reasonably low relative errors when the diffusion term is not vanishingly small, which is the case in this work. The IMEX

Crank-Nicolson time stepping scheme can be written as

$$\frac{\partial C_{i,j,k}}{\partial t} \approx \frac{C_{i,j,k+1} - C_{i,j,k}}{\Delta t} = \frac{1}{2}\left[\nabla \cdot (D(C_{i,j,k})\nabla C_{i,j,k+1}) + \nabla \cdot (D(C_{i,j,k})\nabla C_{i,j,k})\right] + f(C_{i,j,k}).$$

We have verified that the scheme is convergent by successively halving Δx or Δt and recomputing the model solutions with the default parameter values in Eq. (SM.2), and check that the norm of the difference between successive model solutions decreases almost linearly on a log-log plot with respect to Δx or Δt.

To justify that the discretisation we have chosen is sufficiently fine, let C^1_{model} denote the model solution computed with $\Delta x = 29.2$ µm and $\Delta t = 1/30$ h, C^2_{model}, C^3_{model} be the model solution computed with Δx or Δt halved, respectively. Then the difference between C^1_{model} and C^2_{model}, averaged over all grid points, is 0.448, while that between C^1_{model} and C^3_{model} is 4.224, both much smaller than the averaged magnitude of the model solutions, which is on the order of 10^3, therefore we conclude that the numerical scheme is suitably accurate.

1.2 Optmisation procedure for MLE and profile likelihoods

To solve the optimisation problems for finding the MLEs and evaluating the profile likelihood functions, we use three algorithms, all implemented in MATLAB: the built-in *fmincon* and *globalsearch*, and Covariance Matrix Adaptation Evolution Strategy (CM-AES) [3], with the implementation obtained at [2].

The optimisation procedure is initialized with the following default parameter values:

$$D_0 = 1300 \ \mu\text{m}^2/\text{h}, \quad r = 0.3 \ \text{h}^{-1}, \quad K = 2600 \ \text{cell/mm}^2, \quad \alpha = \beta = \gamma = 1, \quad \eta = 0. \tag{SM.2}$$

We impose the following bounds for the parameters to guide the optimisation procedures:

$$\begin{gathered} 100 \ \mu\text{m}^2/\text{h} < D_0 < 10000 \ \mu\text{m}^2/\text{h}, \quad 0.01\text{h}^{-1} < r < 1\text{h}^{-1}, \\ 500 \ \text{cell/mm}^2 < K < 5000 \ \text{cell/mm}^2, \quad 0 < \alpha, \beta, \eta < 3, \quad 0 < \gamma < 9. \end{gathered} \tag{SM.3}$$

We use *globalsearch* to find the MLEs, and *fmincon* to evaluate points on the profile likelihood functions. In the case where *fmincon* struggles to find the true maximum, we use CM-AES instead.

2 Profile likelihoods for synthetic datasets

In this section, we present the profile likelihoods for each model for two sets of synthetic data. The main purpose of this exercise is to verify that the profile likelihoods behave as expected under ideal conditions. The synthetic data are generated by simulating the model, Eq. (1) of the main text, in one spatial dimension, using the parameter values in Eq. (SM.2), and perturbing by adding Gaussian noise to the model solution. The "low noise" dataset uses $\sigma = 20$, while the "high noise" dataset uses $\sigma = 400$. In comparison, the σ^* estimated from real data ranges between 380 – 460, depending on the dataset and the model.

The profile likelihoods for the high noise dataset are presented in Fig. 1, which shows that all profile likelihood curves are unimodal with a finite confidence interval, and the MLEs are close to the true parameter values. For the low noise dataset, the profile likelihood curves are very narrow, and centered almost exactly at the true parameter values. These results verify that the profile likelihoods can recover the true parameter values, at least in a highly idealized case, as the theories suggest.

The profile likelihood curves for the parameters of the Richards and Generalised Fisher models tend to be broader compared to those of the Standard Fisher model, which reflect the greater flexibility of the more complicated models to compensate for a change in one parameter value by shifting the other parameter values.

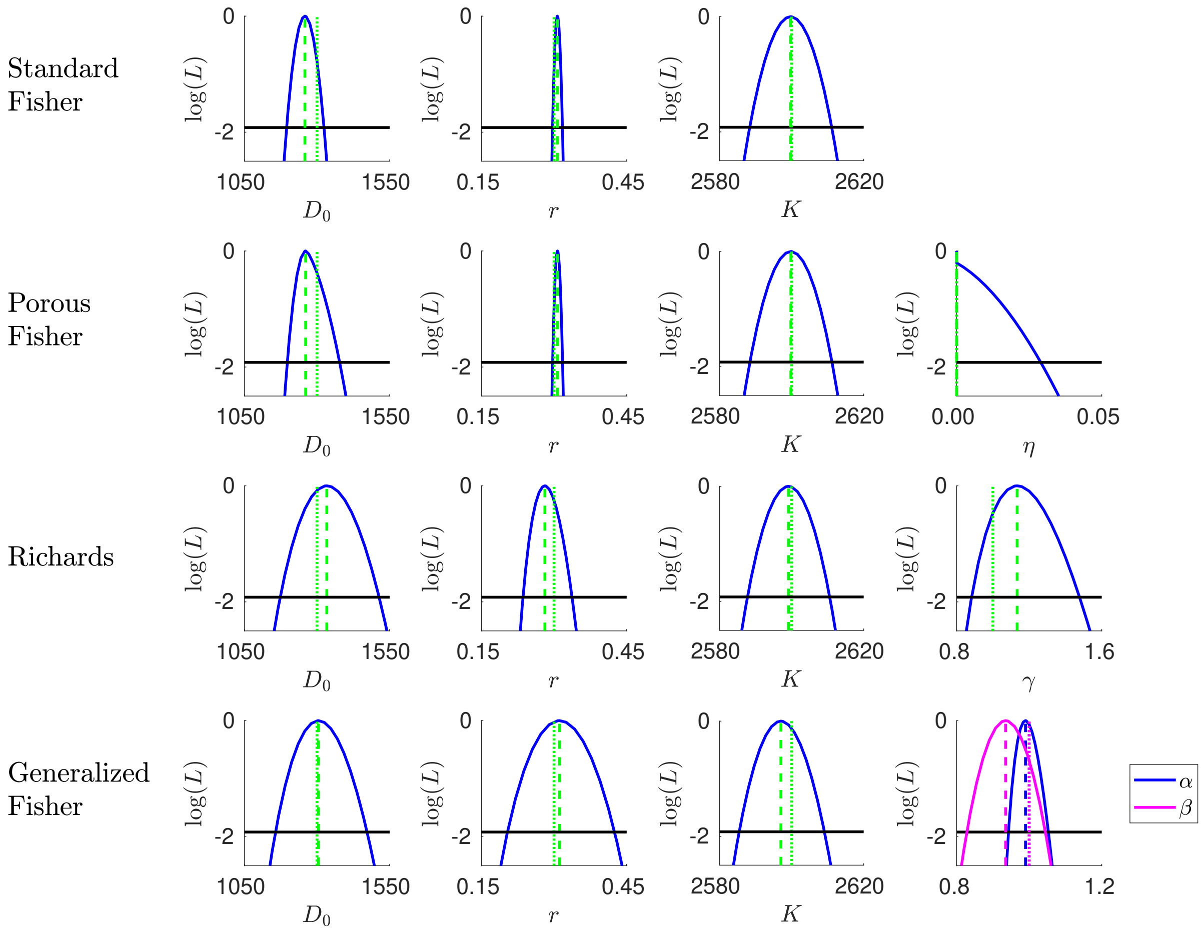

Figure 1: Profile likelihoods for the four models as described in Eq. (1) and Table 1 of the main text, for a synthetic dataset generated with Eq. (1) and parameter values in Eq. (SM.2), perturbed as in Eq. (2) of the main text with $\sigma = 400$. The dotted vertical lines mark the location of the true parameter values, while the dashed vertical lines mark the MLE for each parameter. The black horizontal line at -1.92 marks the threshold for the 95% confidence interval. The axis scale for the parameters shared between the models (D_0, r, K) is kept consistent.

3 Inference results for all datasets

In this section, we present the MLE and 95% confidence intervals of the parameter values calculated for all experimental datasets in table format, and σ^*, the MLE for the noise parameter, as well as the AIC and BIC. Recall that we have cell density data from eight experiments, which we refer to as the full datasets. Experiments 1–4 have circular initial conditions, while Experiments 5–8 have triangular initial conditions. For Experiments 1–4 we also consider the radially-averaged datasets. All results are given to four significant figures.

We also perform a χ^2-likelihood ratio test for nested models [5], and report the p-value. The Standard Fisher model is nested inside the Porous Fisher, Richards, and Generalised Fisher models, and this test provides a measure of whether the more complicated models have a significant improvement in maximum likelihood over the simpler model. Denote the MLE parameters of the Standard Fisher model as $\boldsymbol{\theta}_0^*$, and the MLE parameters of one of the more complicated models as $\boldsymbol{\theta}_1^*$. Let

$$\Lambda = -2\log\left[\frac{L(C_{\text{data}}|\boldsymbol{\theta}_0^*)}{L(C_{\text{data}}|\boldsymbol{\theta}_1^*)}\right] = 2\log L(C_{\text{data}}|\boldsymbol{\theta}_1^*) - 2\log L(C_{\text{data}}|\boldsymbol{\theta}_0^*),$$

be the test statistic based on the ratio of maximum likelihoods of the two nested models being compared. Then, under our assumption of normal i.i.d observation errors (Eq. (2)), by Wilks' theorem [7], $\Lambda \sim \chi^2_{(\text{df})}$, where the degrees of freedom, df, is equal to the number of additional parameters in the more complicated model compared to the simpler model. This allow us to compute a p-value, $p = 1 - \Phi(\lambda)$, where Φ is the cdf of $\Lambda \sim \chi^2_{(\text{df})}$. The $p-$value represents how likely the observed improvement in likelihood can happen, if the simpler model were the true model underlying the data. According to this metric, the more complicated model should be accepted if the p-value is sufficiently small.

For all datasets and all three more complicated models, we have $p < 0.05$, suggesting these models are a significant improvement upon the Standard Fisher model. However, this test is only accurate if the observation errors are indeed normal i.i.d. As discussed in the main text, the observation errors are likely correlated across space and time, hence the improvement in likelihoods are overestimated, and the p-values reported here are likely overestimates.

Model	D_0	r	K	Parameters unique to model	σ	AIC	BIC	$p-$value
Standard Fisher	1287 [1267, 1307]	0.2707 [0.2683, 0.2731]	2620 [2614, 2625]		39.698	162289	162310	-
Porous Fisher	1361 [1306, 1419]	0.2686 [0.2658, 0.2714]	2622 [2616, 2628]	$\eta : 0.0219[0.0069, 0.0372]$	39.689	162283	162311	4.07×10^{-3}
Richards	1467 [1400, 1535]	0.2272 [0.2146, 0.2410]	2612 [2606, 2618]	$\gamma : 1.3119[1.1950, 1.4416]$	39.663	162258	162287	1.07×10^{-8}
Generalised Fisher	1391 [1321, 1465]	0.1429 [0.1130, 0.1758]	2664 [2652, 2678]	$\alpha : 1.1086[1.0779, 1.1434]$ $\beta : 1.2034[1.1587, 1.2508]$	39.571	**162175**	**162210**	$< 10^{-20}$

Table 1: Experiment 1, radially-averaged dataset

Model	D_0	r	K	Parameters unique to model	σ	AIC	BIC	$p-$value
Standard Fisher	1287 [1282, 1293]	0.2775 [0.2767, 0.2782]	2621 [2619, 2622]		21.447	26161393	26161430	-
Porous Fisher	1545 [1529, 1564]	0.2702 [0.2694, 0.2710]	2628 [2627, 2630]	$\eta : 0.0719[0.0677, 0.0766]$	21.445	26160486	26160536	$< 10^{-20}$
Richards	1402 [1385, 1418]	0.2462 [0.2425, 0.2502]	2616 [2615, 2618]	$\gamma : 1.1972[1.1677, 1.2249]$	21.447	26161172	26161221	$< 10^{-20}$
Generalised Fisher	1423 [1403, 1443]	0.1013 [0.0945, 0.1085]	2701 [2698, 2704]	$\alpha : 1.1733[1.1688, 1.1818]$ $\beta : 1.3548[1.3437, 1.3663]$	21.437	**26158023**	**26158085**	$< 10^{-20}$

Table 2: Experiment 1, full dataset

Model	D_0	r	K	Parameters unique to model	σ	AIC	BIC	$p-$value
Standard Fisher	1211 [1192, 1231]	0.2780 [0.2755, 0.2805]	2551 [2546, 2556]		39.620	162216	162238	-
Porous Fisher	1650 [1578, 1725]	0.2673 [0.2645, 0.2701]	2564 [2558, 2570]	$\eta : 0.1270[0.1086, 0.1458]$	39.396	162009	162038	$< 10^{-20}$
Richards	981 [938, 1026]	0.3675 [0.3474, 0.3891]	2560 [2554, 2565]	$\gamma : 0.6808[0.6333, 0.7323]$	39.509	162115	162143	$< 10^{-20}$
Generalised Fisher	806 [758, 857]	0.2787 [0.2374, 0.3219]	2873 [2838, 2913]	$\alpha : 1.0744[1.0532, 1.0984]$ $\beta : 2.0151[1.9249, 2.1130]$	38.266	**160935**	**160970**	$< 10^{-20}$

Table 3: Experiment 2, radially-averaged dataset

Model	D_0	r	K	Parameters unique to model	σ	AIC	BIC	$p-$value
Standard Fisher	1157 [1152, 1163]	0.2896 [0.2888, 0.2904]	2550 [2548, 2552]		21.258	26099801	26099838	-
Porous Fisher	1527 [1509, 1545]	0.2788 [0.2779, 0.2795]	2561 [2560, 2563]	$\eta : 0.1099[0.1071, 0.1129]$	21.250	26097458	26097507	$< 10^{-20}$
Richards	916 [907, 926]	0.4061 [0.4004, 0.4119]	2561 [2559, 2562]	$\gamma : 0.6307[0.6196, 0.6419]$	21.250	26097356	26097406	$< 10^{-20}$
Generalised Fisher	826 [816, 837]	0.2383 [0.2283, 0.2484]	2893 [2881, 2905]	$\alpha : 1.1091[1.1024, 1.1160]$ $\beta : 2.1168[2.0860, 2.1485]$	21.204	**26082314**	**26082376**	$< 10^{-20}$

Table 4: Experiment 2, full dataset

Model	D_0	r	K	Parameters unique to model	σ	AIC	BIC	$p-$value
Standard Fisher	1107 [1083, 1131]	0.3172 [0.3133, 0.3212]	2518 [2511, 2525]		47.887	169220	169242	-
Porous Fisher	1228 [1160, 1301]	0.3136 [0.3093, 0.3179]	2520 [2513, 2527]	$\eta : 0.0394[0.0191, 0.0604]$	47.868	**169207**	**169236**	1.11×10^{-4}
Richards	1200 [1117, 1288]	0.2860 [0.2622, 0.3122]	2514.7859 [2507, 2522]	$\gamma : 1.1669[1.0249, 1.3367]$	47.880	169217	169245	1.94×10^{-2}
Generalised Fisher	1146 [1064, 1237]	0.2391 [0.1941, 0.2866]	2534 [2516, 2553]	$\alpha : 1.0501[1.0231, 1.0811]$ $\beta : 1.1005[1.0093, 1.1850]$	47.867	169209	169245	4.77×10^{-4}

Table 5: Experiment 3, radially-averaged dataset

Model	D_0	r	K	Parameters unique to model	σ	AIC	BIC	$p-$value
Standard Fisher	1118 [1113,1122]	0.3198 [0.3191,0.3206]	2521 [2520,2523]		21.418	26152075	26152112	-
Porous Fisher	1300 [1286,1314]	0.3144 [0.3136,0.3153]	2524 [2523, 2526]	$\eta : 0.0569[0.0529, 0.0609]$	21.416	26151214	26151263	$< 10^{-20}$
Richards	1293 [1275,1307]	0.2633 [0.2597,0.2678]	2514 [2513,2516]	$\gamma : 1.3493[1.3131, 1.3801]$	21.417	26151457	26151506	$< 10^{-20}$
Generalised Fisher	1377 [1353,1398]	0.1999 [0.1919,0.2080]	2500 [2498,2504]	$\alpha : 1.0506[1.0439, 1.0576]$ $\beta : 0.8671[0.8443, 0.8943]$	21.416	**26151117**	**26151179**	$< 10^{-20}$

Table 6: Experiment 3, full dataset

Model	D_0	r	K	Parameters unique to model	σ	AIC	BIC	$p-$value
Standard Fisher	1239 [1221,1257]	0.2849 [0.2825,0.2873]	2784 [2779,2790]		39.387	161998	162019	-
Porous Fisher	1406 [1356,1458]	0.2800 [0.2773,0.2827]	2789 [2784,2795]	$\eta : 0.0499[0.0363, 0.0637]$	39.329	161945	161974	1.56×10^{-13}
Richards	1466 [1410,1523]	0.2273 [0.2168,0.2387]	2775 [2770,2781]	$\gamma : 1.4221[1.3136, 1.5407]$	39.304	161922	161951	1.14×10^{-18}
Generalised Fisher	1416 [1358,1476]	0.1464 [0.1216,0.1732]	2799 [2789,2810]	$\alpha : 1.1009[1.0761, 1.1284]$ $\beta : 1.0869[1.0457, 1.1293]$	39.292	**161913**	**161948**	4.32×10^{-20}

Table 7: Experiment 4, radially-averaged dataset

Model	D_0	r	K	Parameters unique to model	σ	AIC	BIC	$p-$value
Standard Fisher	1252 [1247,1258]	0.2865 [0.2858,0.2872]	2788 [2787,2790]		21.436	26157597	26157634	-
Porous Fisher	1440 [1425,1454]	0.2809 [0.2801,0.2817]	2794 [2792,2795]	$\eta : 0.0539[0.0500, 0.0575]$	21.433	26156816	26156866	$< 10^{-20}$
Richards	1565 [1548,1580]	0.2108 [0.2084,0.2137]	2775 [2773,2776]	$\gamma : 1.6398[1.6013, 1.6736]$	21.429	**26155625**	**26155675**	$< 10^{-20}$
Generalised Fisher	1552 [1533,1570]	0.1028 [0.0965,0.1093]	2797 [2794,2800]	$\alpha : 1.1489[1.1398, 1.1583]$ $\beta : 1.0701[1.0572, 1.0831]$	21.429	26155629	26155691	$< 10^{-20}$

Table 8: Experiment 4, full dataset

Model	D_0	r	K	Parameters unique to model	σ	AIC	BIC	$p-$value
Standard Fisher	1416 [1410,1422]	0.3085 [0.3077,0.3093]	2344 [2343,2346]		20.048	25693899	25693936	-
Porous Fisher	4102 [4059,4145]	0.2739 [0.2731,0.2747]	2377 [2375,2379]	$\eta : 0.5677[0.5609, 0.5746]$	19.920	25649571	25649620	$< 10^{-20}$
Richards	2697 [2688,2706]	0.1364 [0.1359,0.1367]	2336 [2334,2337]	$\gamma : 7.8055[7.7527, 7.9284]$	19.749	**25589713**	**25589762**	$< 10^{-20}$
Generalised Fisher	-	-	-	$\alpha : -$ $\beta : -$	-	-	-	-

Table 9: Experiment 5, full dataset

Model	D_0	r	K	Parameters unique to model	σ	AIC	BIC	$p-$value
Standard Fisher	1161 [1156,1166]	0.3244 [0.3235,0.3253]	2307 [2305,2308]		19.820	25614660	25614697	-
Porous Fisher	2713 [2685,2743]	0.2931 [0.2922,0.2940]	2333 [2331,2334]	$\eta : 0.4097[0.4041, 0.4157]$	19.743	25587515	25587565	$< 10^{-20}$
Richards	2290 [2282,2299]	0.1413 [0.1409,0.1416]	2288 [2286,2289]	$\gamma : 8.1042[8.0335, 8.1723]$	19.593	**25534767**	**25534816**	$< 10^{-20}$
Generalised Fisher	-	-	-	$\alpha : -$ $\beta : -$	-	-	-	-

Table 10: Experiment 6, full dataset

Model	D_0	r	K	Parameters unique to model	σ	AIC	BIC	$p-$value
Standard Fisher	1845 [1837,1853]	0.2260 [0.2254,0.2266]	2419 [2417,2421]		19.945	25658251	25658288	-
Porous Fisher	10061 [9894,10232]	0.1637 [0.1628,0.1646]	2627 [2622,2631]	$\eta : 1.0443[1.0313, 1.0573]$	19.750	25590211	25590260	$< 10^{-20}$
Richards	3180 [3170,3190]	0.1057 [0.1055,0.1059]	2353 [2352,2354]	$\gamma : \infty*$	19.556	**25521509**	**25521559**	$< 10^{-20}$
Generalised Fisher	-	-	-	$\alpha : -$ $\beta : -$	-	-	-	-

Table 11: Experiment 7, full dataset. Note that for γ in the Richards model, the profile likelihood seems to be monotonically increasing up to the upper bound of $\gamma = 9$ which we have imposed for numerical stability, therefore the true MLE is likely to be very large or infinite.

Model	D_0	r	K	Parameters unique to model	σ	AIC	BIC	$p-$value
Standard Fisher	1448 [1442,1454]	0.2669 [0.2662,0.2676]	2294 [2292,2296]		19.598	25536527	25536564	-
Porous Fisher	3504 [3467,3543]	0.2374 [0.2367,0.2382]	2337 [2335,2339]	$\eta : 0.4475[0.4414, 0.4539]$	19.518	25508165	25508214	$< 10^{-20}$
Richards	2666 [2658,2675]	0.1199 [0.1196,0.1201]	2241 [2239,2242]	$\gamma : \infty*$	19.341	**25444985**	**25445034**	$< 10^{-20}$
Generalised Fisher	-	-	-	$\alpha : -$ $\beta : -$	-	-	-	-

Table 12: Experiment 8, full dataset. Similar observations for γ as in Experiment 7.

We also present the profile likelihoods for Experiments 2–8 (those for Experiment 1 are presented in Fig. 2 and Fig. 3 of the main text).

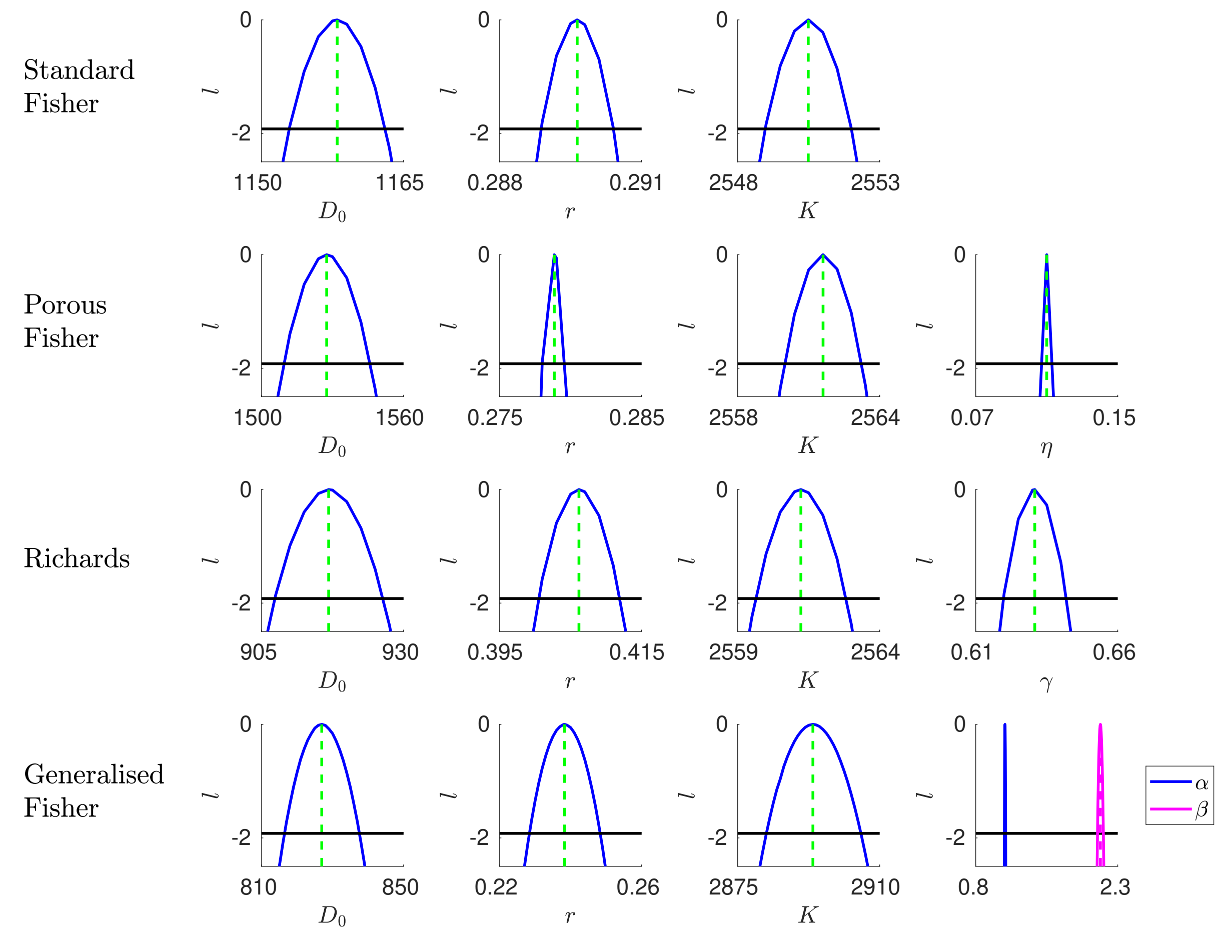

Figure 2: Profile likelihoods for Experiment 2, full dataset.

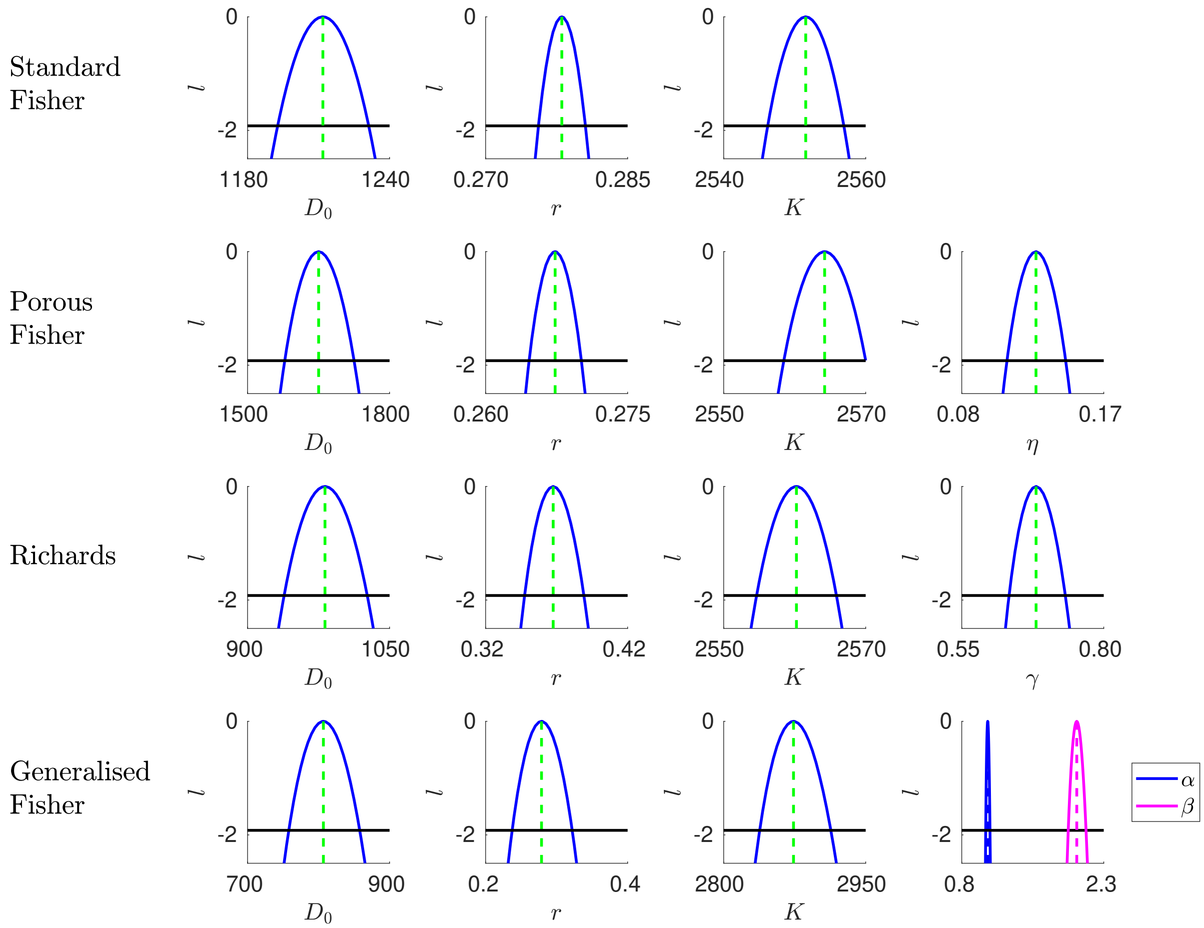

Figure 3: Profile likelihoods for Experiment 2, radially averaged dataset.

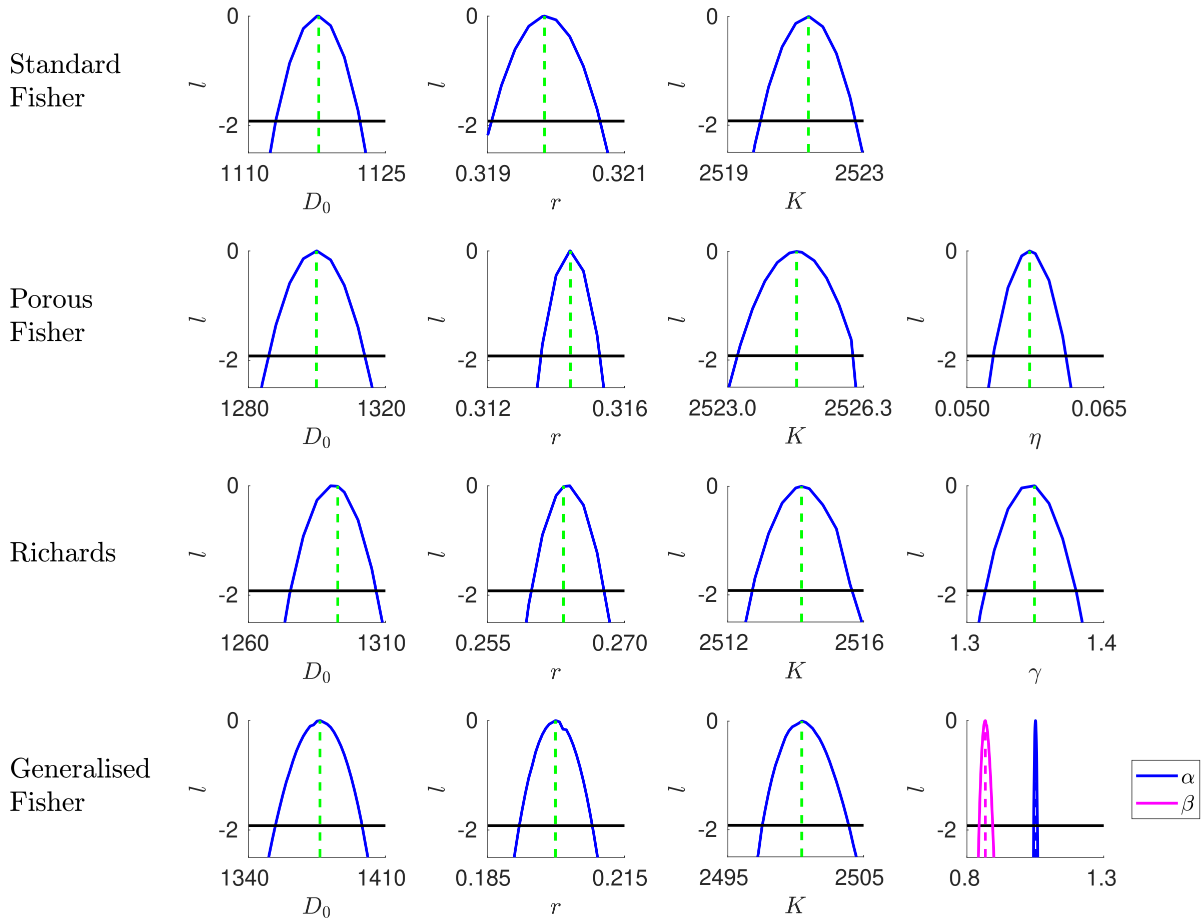

Figure 4: Profile likelihoods for Experiment 3, full dataset.

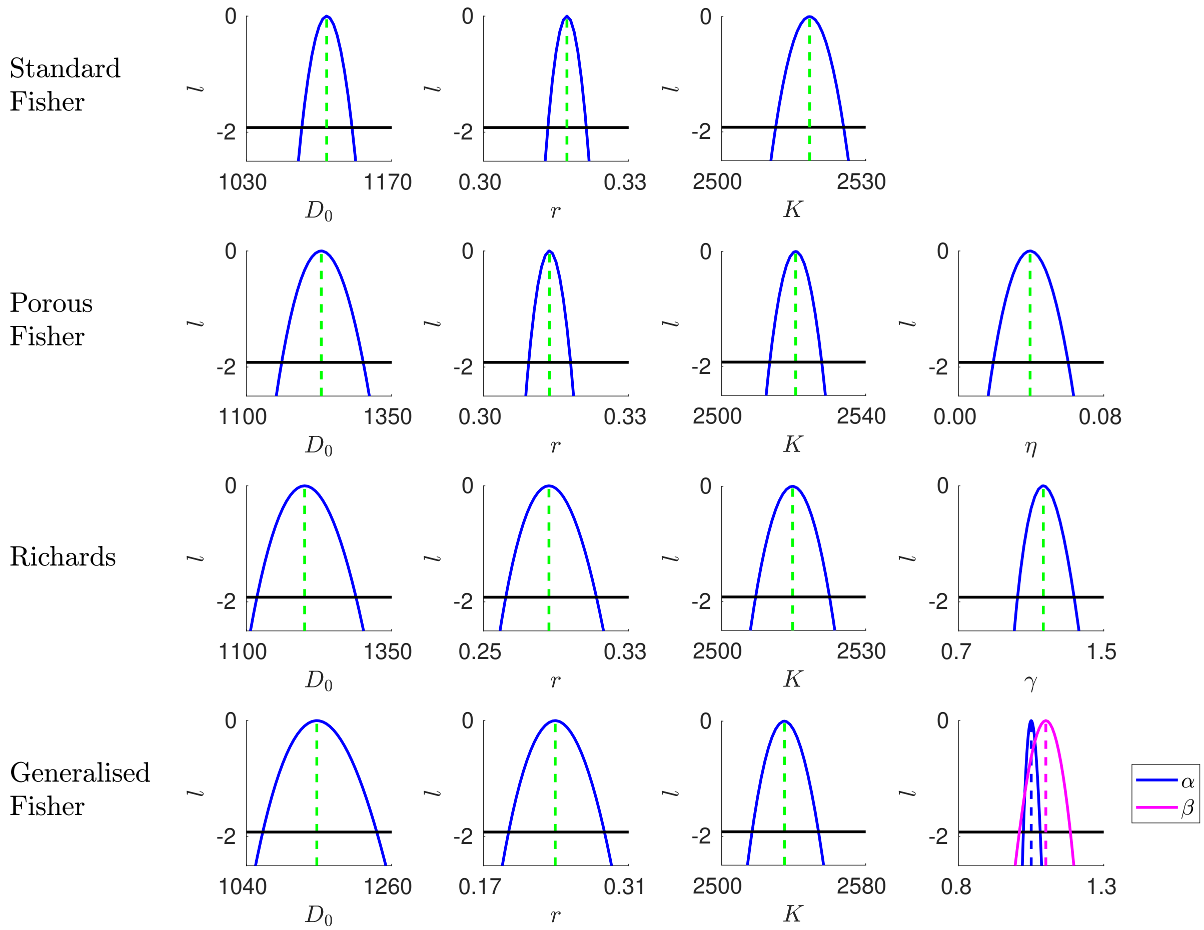

Figure 5: Profile likelihoods for Experiment 3, radially averaged dataset.

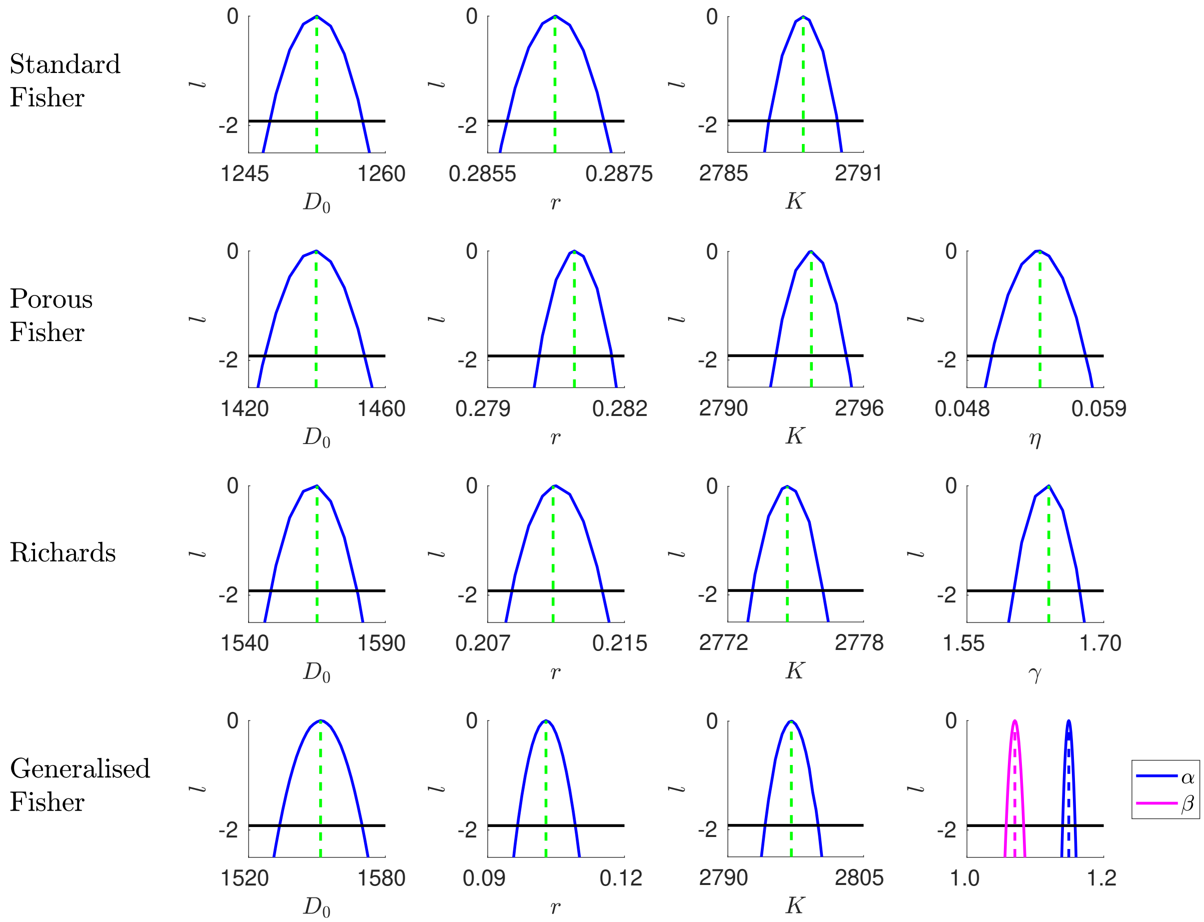

Figure 6: Profile likelihoods for Experiment 4, full dataset.

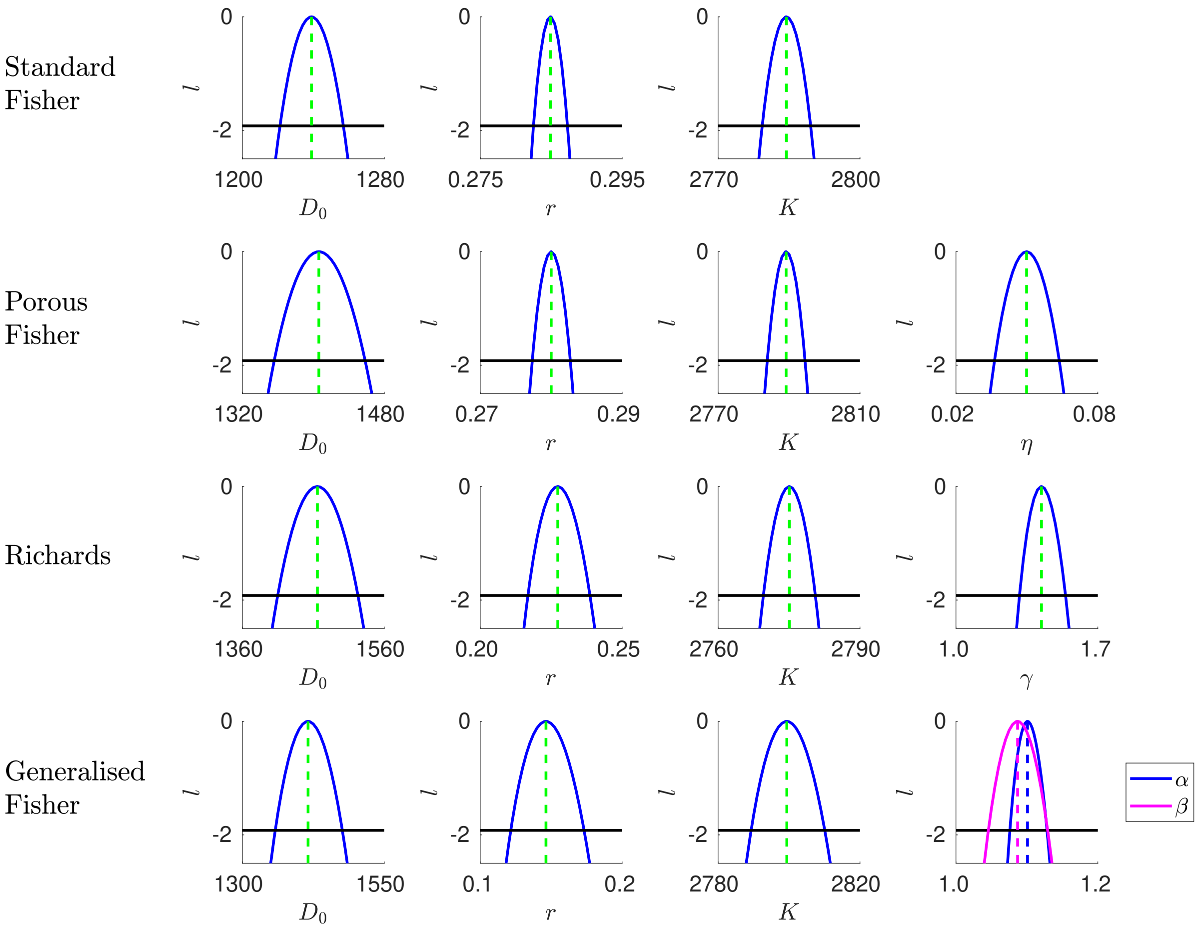

Figure 7: Profile likelihoods for Experiment 4, radially averaged dataset.

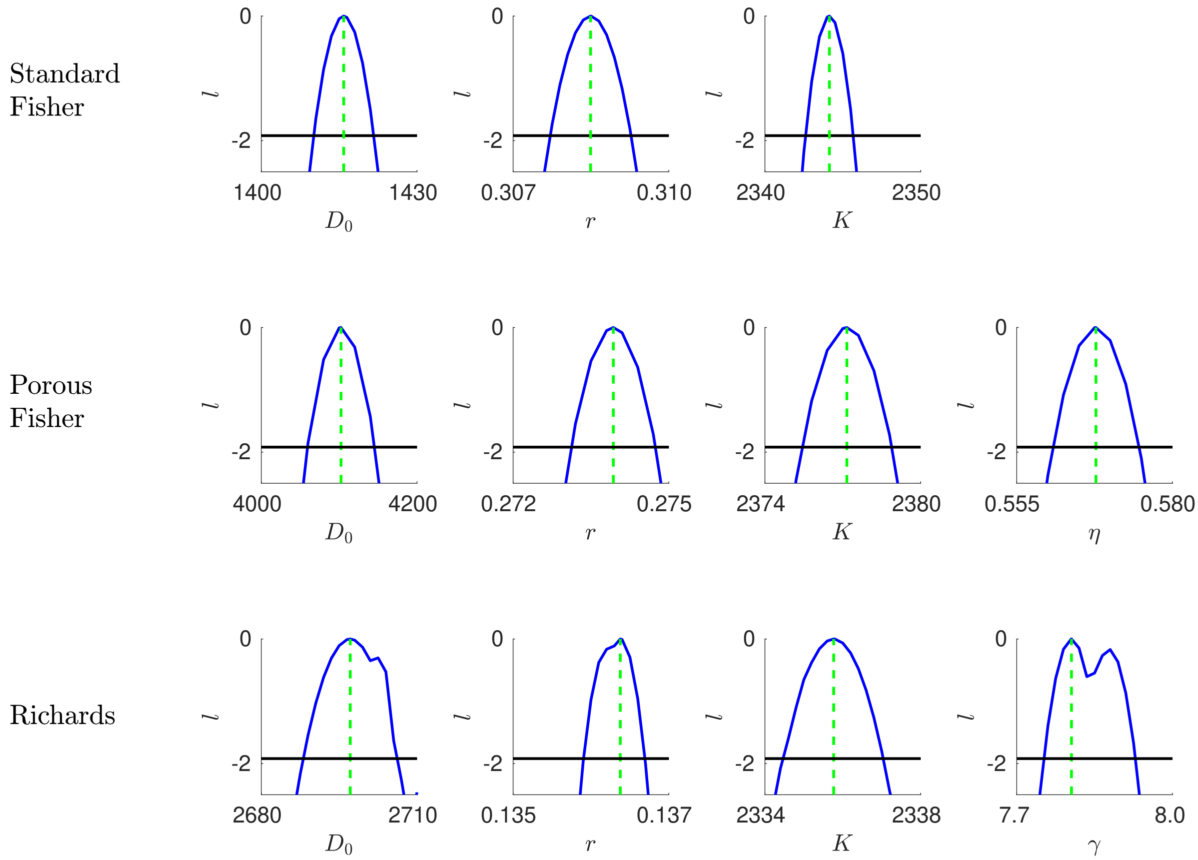

Figure 8: Profile likelihoods for Experiment 5, full dataset.

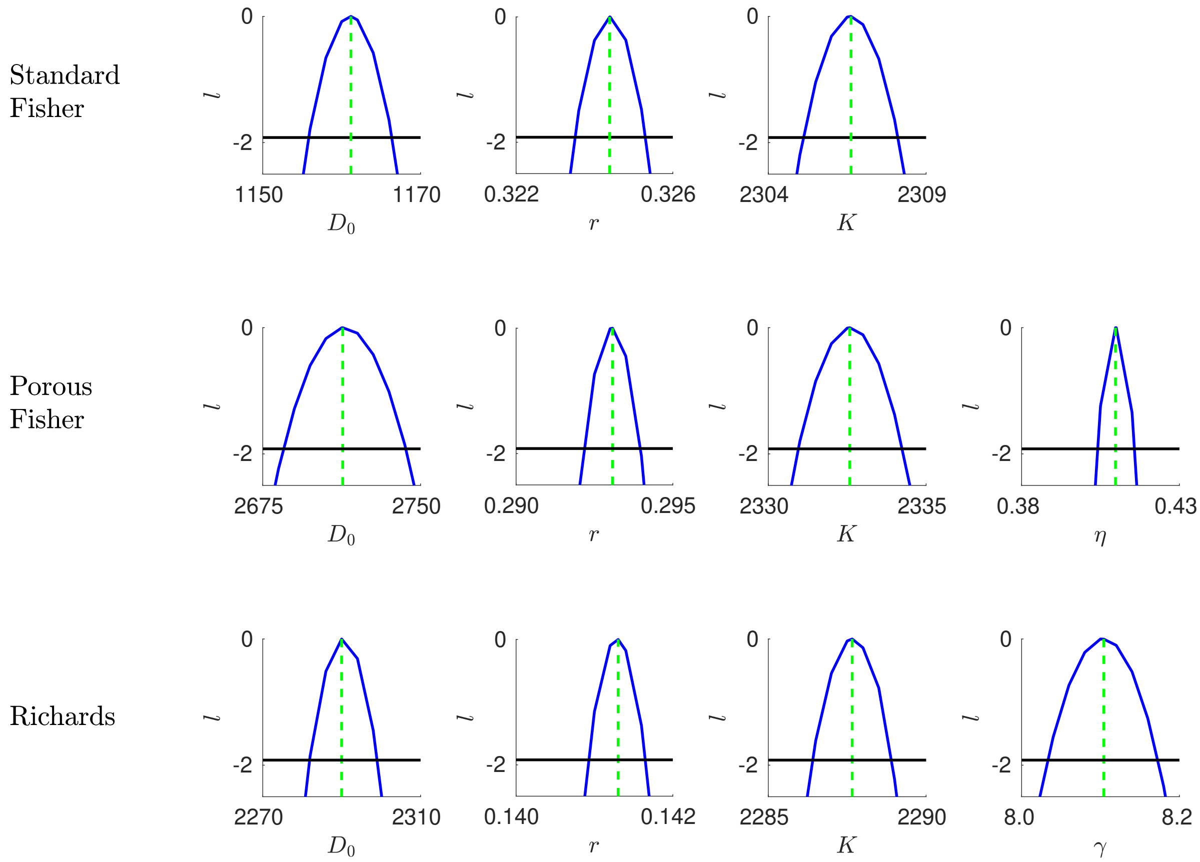

Figure 9: Profile likelihoods for Experiment 6, full dataset.

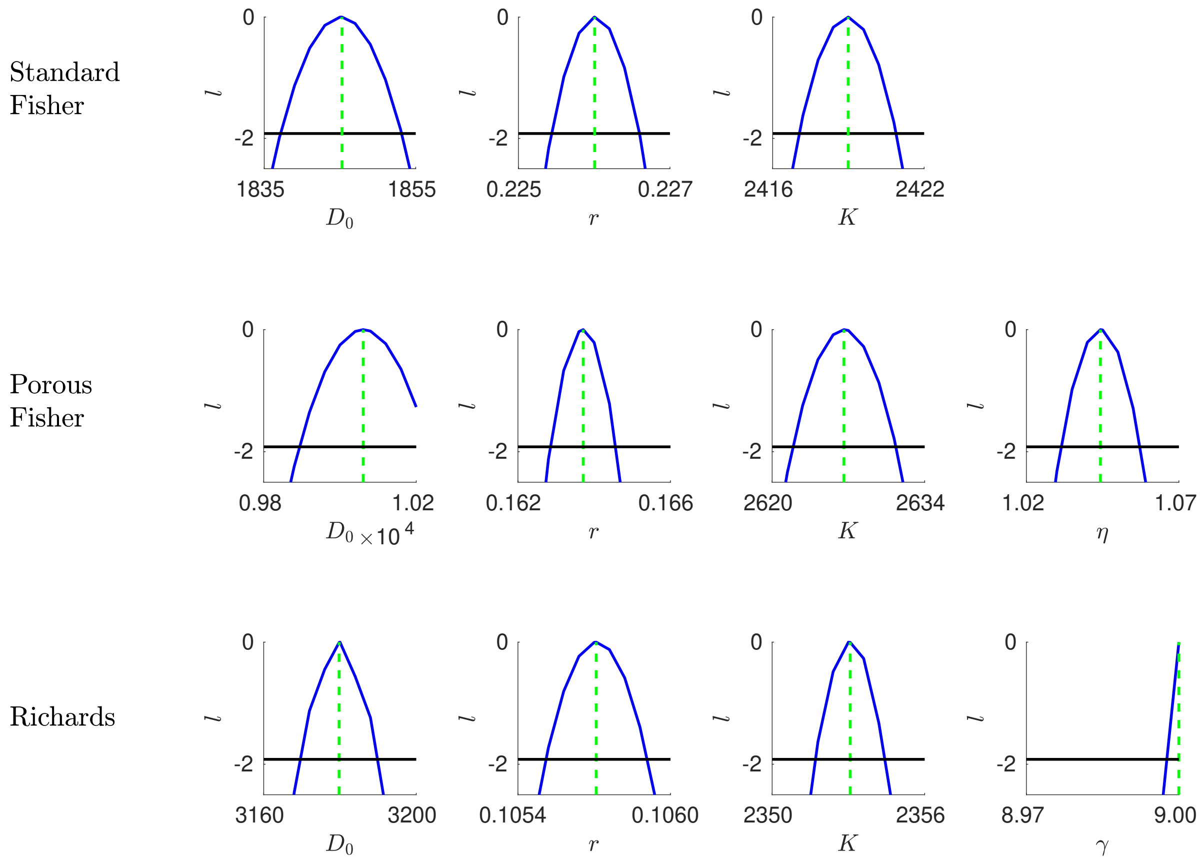

Figure 10: Profile likelihoods for Experiment 7, full dataset.

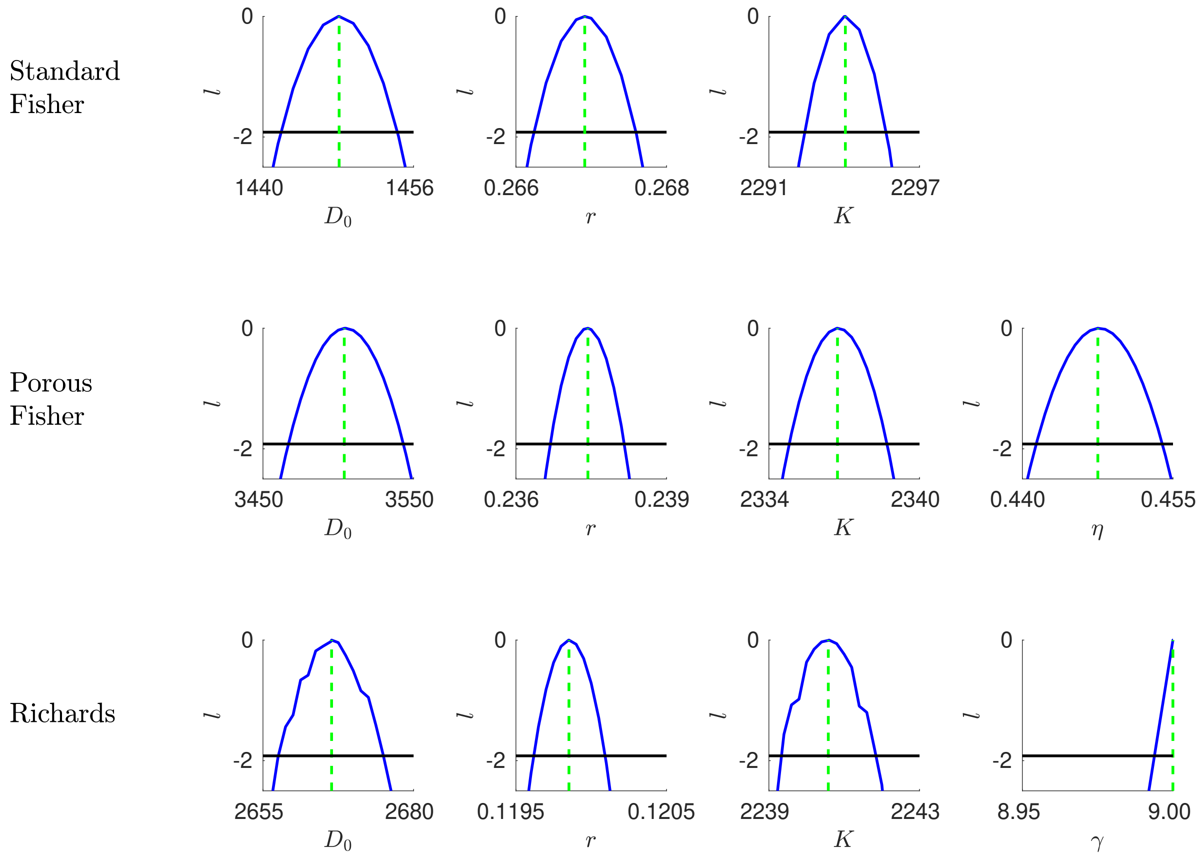

Figure 11: Profile likelihoods for Experiment 8, full dataset.

4 Profile likelihoods for down-sampled data

In Fig. 12 we present the profile likelihoods for the down-sampled datasets.

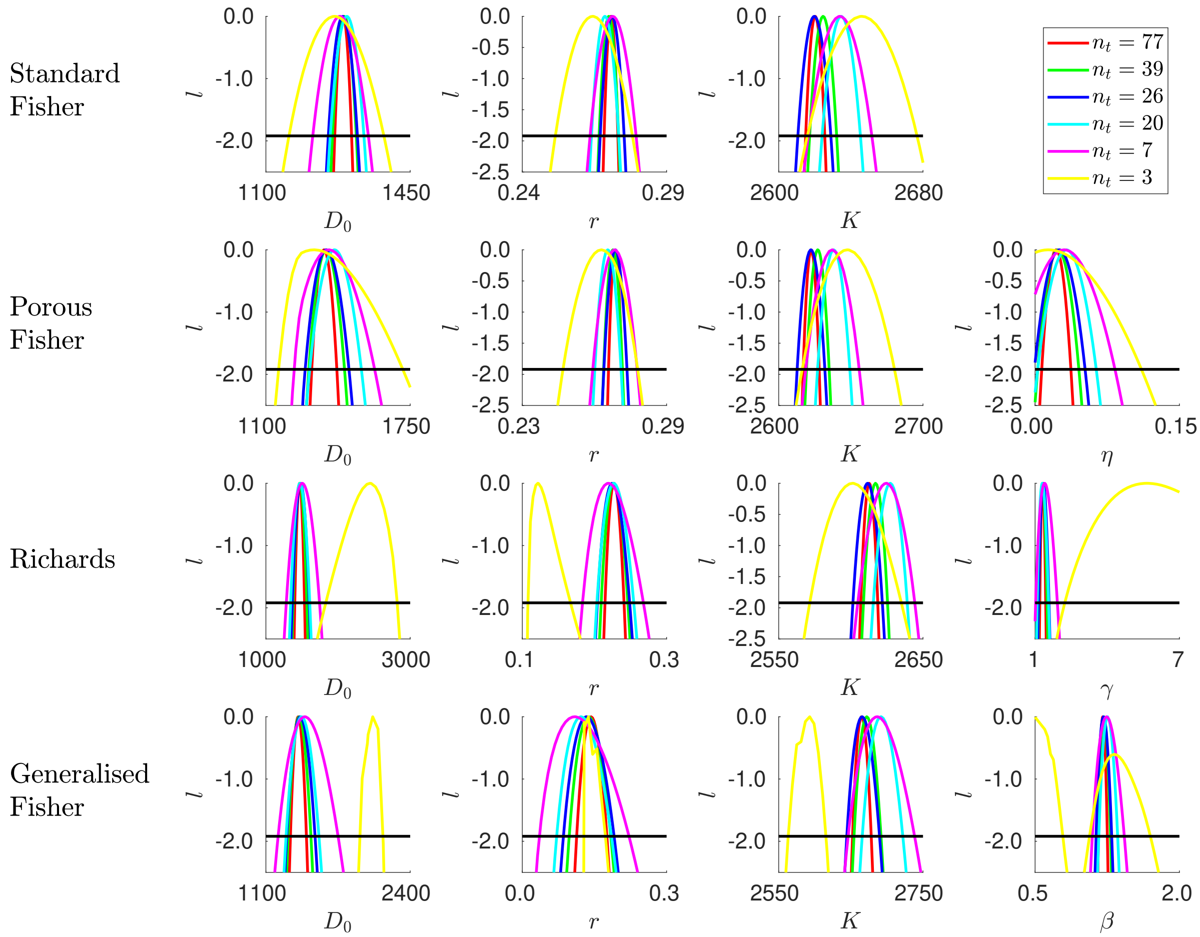

Figure 12: Profile likelihoods for the down-sampled datasets. A subset of these were presented in Fig. 7 of the main text.

References

[1] Uri M. Ascher, Steven J. Ruuth, and Brian T. R. Wetton. "Implicit-explicit methods for time-dependent partial differential equations". In: *SIAM Journal on Numerical Analysis* 32 (1995), pp. 797–823. ISSN: 0036-1429. DOI: 10.1137/0732037.

[2] Nikolaus Hansen. *The CMA evolution strategy*. 2014. URL: https://cma-es.github.io/ (visited on 11/21/2022).

[3] Nikolaus Hansen and Andreas Ostermeier. "Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation". In: *Proceedings of IEEE International Conference on Evolutionary Computation*. 1996, pp. 312–317. DOI: 10.1109/ICEC.1996.542381.

[4] Steven J. Ruuth. "Implicit-explicit methods for reaction-diffusion problems in pattern formation". en. In: *Journal of Mathematical Biology* 34 (1995), pp. 148–176. ISSN: 1432-1416. DOI: 10.1007/BF00178771.

[5] George A. F. Seber and Christopher J. Wild. *Nonlinear Regression*. Wiley-Interscience, 2003.

[6] David J. Warne, Ruth E. Baker, and Matthew J. Simpson. "Using experimental data and information criteria to guide model selection for reaction–diffusion problems in mathematical biology". In: *Bulletin of Mathematical Biology* 81.6 (2019), pp. 1760–1804. ISSN: 0092-8240, 1522-9602. DOI: 10.1007/s11538-019-00589-x.

[7] Samuel S. Wilks. "The large-sample distribution of the likelihood ratio for testing composite hypotheses". In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62. ISSN: 0003-4851. JSTOR: 2957648. (Visited on 11/30/2023).