Online Resource 1

Testing spatial Markovianity of breakthrough curves to separate velocity heterogeneity from mass transfer

Thi Thanh Hoa Nguyen, Thi My Hong Tran, Song Giang Le

Transport in Porous Media. Corresponding author: ntthoa@hcmunre.edu.vn

This document gives details of the analyses summarised in Sects. 4.1, 4.7, 4.8 and 5 of the article. The code and the summary results are provided in the accompanying archive (Online Resource 2); Sect. S8 lists the scripts.

S1 Sensitivity of the Markov and flow-rate tests (Sect. 4.7)

Table S1 compares single-realization and ensemble sampling, the Péclet number and the test statistic for the self-convolution test. The Péclet runs used the first 20 of the 30 fields of A ($\sigma^2_{\ln K} = 1$); the growth factor of $\kappa_2(d)/d$ between λ and 64λ was 2.7, 4.1 and 4.4 for Pe = 10, 100 and 1000. Table S2 gives the flow-rate test with three quantiles. In the homogeneous map cells ($\sigma^2_{\ln K} = 0$), whose BTCs are very narrow at Pe = 100, the random-walk step factor was reduced from $c_a = 0.25$ to 0.05: with $c_a = 0.25$ the false-alarm rate of the Markov test at the shortest plane pair was 9.3%, with $c_a = 0.05$ it was 4.8% (400 replicates, $N = 1000$), which identifies a time-stepping artefact absent for the broader BTCs of heterogeneous fields.

Table S1: Rejection rates of the self-convolution test ($N = 1000$, 200 replicates, $\alpha = 0.05$). Single: samples from one realization; ensemble: each arrival time from an independently drawn realization; KS: Kolmogorov–Smirnov; AD: Anderson–Darling; A1, A4: mechanism A with $\sigma^2_{\ln K} = 1, 4$

Case	λ–2λ	8λ–16λ	32λ–64λ
A1, single / ensemble	0.94 / 0.995	0.77 / 0.26	0.81 / 0.20
A4, single / ensemble	0.91 / 0.99	0.85 / 0.35	0.67 / 0.11
C, single / ensemble	0.97 / 0.98	0.50 / 0.08	0.16 / 0.03
B, single / ensemble	0.06 / 0.03	0.06 / 0.05	0.04 / 0.03
A1, Pe = 10 / 100 / 1000	0.73 / 0.97 / 1.00	0.61 / 0.78 / 0.89	0.53 / 0.82 / 0.86
A4, KS / AD	–	0.85 / 0.91	–
A1, KS / AD	–	0.77 / 0.85	–
B, KS / AD	–	0.03 / 0.05	–

Table S2: Rejection rate of the flow-rate test with different quantiles ($N = 3000$, plane 16λ, 100 replicates)

Cell ($\sigma^2_{\ln K}$; R; $k = 0.1$)	q_{90}	q_{95}	q_{99}
0; 0	0.00	0.00	0.00
0; 0.1	0.00	1.00	1.00
0; 0.3	1.00	1.00	1.00
1; 0	0.20	0.13	0.08
1; 0.1	0.22	0.92	1.00
1; 0.3	1.00	1.00	1.00
2; 0	0.13	0.25	0.33

S2 Parameter drift of the MIM model (Sect. 4.1)

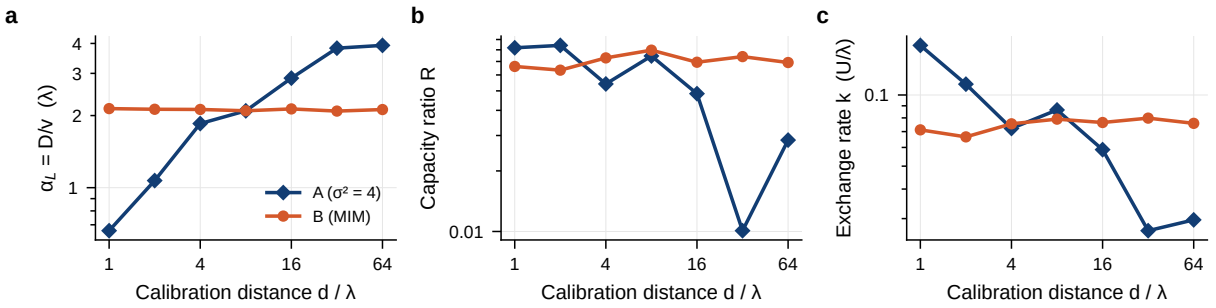


Figure S1: MIM parameters fitted separately at each control plane for mechanisms A ($\sigma^2_{\ln K} = 4$, ensemble) and B: **a** longitudinal dispersivity D/v, **b** capacity ratio R, **c** release rate k. For B the parameters are constant within estimation noise ($\alpha_L = 2.09$–2.14, $k = 0.066$–0.080)

S3 Model-based comparison (Sect. 4.8)

For each of 25 replicates, one realization was drawn and $N = 1000$ arrival times were taken from disjoint particle sets at 8λ and 16λ. Arrival times were binned on 20 logarithmic bins spanning the sample, and models were fitted by binned Poisson maximum likelihood (Nelder–Mead in log-parameter space, several starting points, exact MIM density of Eq. (7) with 120 Gauss–Legendre nodes). *AIC:* ADE (2 parameters) and MIM (4 parameters) fitted at 8λ; MIM selected if $\mathrm{AIC}_{\mathrm{MIM}} < \mathrm{AIC}_{\mathrm{ADE}}$. *Drift test:* $\mathrm{LR} = 2(\ell_{\mathrm{joint}} - \ell_8 - \ell_{16})$, with ℓ the negative log-likelihood of a common MIM parameter set fitted to both planes or of separate sets; rejection if the χ^2_4 p-value is below 0.05. The homogeneous MRMT medium had $v = 1$, $D = 0.2$ and five rate classes with $R_i = 0.06$ and k_i log-spaced between 0.01 and 1, sampled exactly (inverse-Gaussian mobile time plus independent compound-Poisson immobile times); its Markov test used 200 replicates. The Markov-test column for A, B and C is the single-realization KS rejection rate of Table S1 (pair 8λ–16λ).

Table S3: Model-based approaches and the self-convolution (Markov) test on the same kind of data ($N = 1000$, planes 8λ and 16λ)

Data	AIC selects MIM	Drift test rejects	Median LR	Markov test rejects
B (MIM only)	0.72	0.04	2.3	0.06
A, $\sigma^2_{\ln K} = 1$	0.64	0.84	26.6	0.77
A, $\sigma^2_{\ln K} = 4$	0.88	0.80	46.1	0.85
C ($\sigma^2_{\ln K} = 1$ + MIM)	0.96	0.80	18.7	0.50
Homogeneous MRMT (5 rates)	–	0.44	8.1	0.06

S4 Flow-rate test, Damköhler number and multirate exchange (Sect. 4.7)

Homogeneous mobile domain ($v = U$, $D = D_m = 0.01$), plane 16λ, $N = 3000$ per flow rate, $U = 1$ and 0.25, 100 replicates, 199 permutations. One-sided: rejection for tail contraction (the test used in the article); two-sided: rejection for any change of the 95th percentile of scaled time. T is the mean of $\log(q_{95}^{U=1}/q_{95}^{U=0.25})$; positive values indicate contraction at the lower flow rate. The two-sided test also rejects the homogeneous case without exchange because the Péclet number changes with U, so it cannot be used without a model of local dispersion.

Table S4: Rejection rates of the flow-rate test for exchange of different kinetics

Exchange	One-sided	Two-sided	Mean T
No exchange	0.00	1.00	-0.056
Single rate, $R = 0.3$, $k = 0.003$	0.00	1.00	-0.383
Single rate, $R = 0.3$, $k = 0.01$	0.00	1.00	-1.024
Single rate, $R = 0.3$, $k = 0.03$	0.98	0.98	+0.245
Single rate, $R = 0.3$, $k = 0.1$	1.00	1.00	+0.299
Single rate, $R = 0.3$, $k = 0.3$	1.00	1.00	+0.192
Single rate, $R = 0.3$, $k = 1$	1.00	1.00	+0.087
Single rate, $R = 0.3$, $k = 3$	1.00	0.99	+0.021
Two rates, $R = 0.15 + 0.15$, $k = 1$, 0.01	0.00	1.00	-0.369
Five rates, total $R = 0.3$, $k = 0.003$–3	0.03	0.07	-0.012
Five rates, total $R = 0.3$, $k = 0.03$–3	0.95	0.92	+0.102
Five rates, total $R = 0.1$, $k = 0.003$–3	0.00	0.32	-0.029

S5 Temporally autocorrelated noise (Sect. 4.7)

Concentration curves as in Sect. 3.3 ($N = 3000$, $\sigma_r = 5\%$, DL $= 1\%$, planes 8λ–16λ, 30 realizations), with the multiplicative noise generated by an AR(1) process of coefficient ϕ. In the bootstrap the noise was either white (autocorrelation ignored) or AR(1) with $\hat{\phi}$ estimated as the lag-one autocorrelation of the relative residuals about a five-point running mean of the measured curves. The estimator is biased by the curvature of the BTC and by counting noise ($\hat{\phi} \approx 0.7$ for both values of ϕ); the test nevertheless kept its size in both variants. For comparison, the rejection rate for A ($\sigma^2_{\ln K} = 4$) with white noise is 0.60.

Table S5: Concentration test with AR(1) noise

ϕ	Noise in bootstrap	Data	Rejection rate	Mean $\hat{\phi}$
0.5	ignored	A, $\sigma^2_{\ln K} = 4$	0.57	0.00
0.5	ignored	B (MIM only)	0.00	0.00
0.5	AR(1), estimated	A, $\sigma^2_{\ln K} = 4$	0.50	0.67
0.5	AR(1), estimated	B (MIM only)	0.00	0.71
0.8	ignored	A, $\sigma^2_{\ln K} = 4$	0.63	0.00
0.8	ignored	B (MIM only)	0.03	0.00
0.8	AR(1), estimated	A, $\sigma^2_{\ln K} = 4$	0.57	0.68
0.8	AR(1), estimated	B (MIM only)	0.03	0.71

S6 Diffusion-dominated regions (Sect. 5)

Fraction of cells with local Péclet number $|\mathbf{v}|\lambda/D_m < 1$ (cell-centred velocity, $D_m = 0.01$), averaged over 10 fields per variance.

Table S6: Area fraction of diffusion-dominated cells

$\sigma^2_{\ln K}$	$U = 1$	$U = 0.25$
0.5	0.0e+00	0.0e+00
1	0.0e+00	0.0e+00
2	6.4e-06	1.1e-03
4	1.0e-03	1.5e-02

S7 Confidence intervals of the increment correlation (Fig. 3c)

95% bootstrap intervals of the Spearman correlation between $T(d)$ and $T(2d) - T(d)$, resampling realizations with replacement (100 bootstrap samples, 4000 particles per realization).

Table S7: 95% intervals of ρ_S for the pairs $(d, 2d)$

Data	λ	2λ	4λ	8λ	16λ	32λ
A, $\sigma^2_{\ln K} = 0.5$	0.61–0.69	0.47–0.59	0.29–0.45	0.14–0.30	-0.00–0.12	-0.06–0.09
A, $\sigma^2_{\ln K} = 1$	0.65–0.73	0.54–0.63	0.33–0.46	0.17–0.33	-0.05–0.11	0.01–0.18
A, $\sigma^2_{\ln K} = 2$	0.67–0.73	0.49–0.61	0.25–0.40	0.15–0.28	0.05–0.19	-0.08–0.03
A, $\sigma^2_{\ln K} = 4$	0.64–0.76	0.50–0.67	0.30–0.44	0.12–0.31	0.07–0.24	0.05–0.17
B	-0.01–0.01	-0.01–0.00	-0.01–0.01	-0.01–0.01	-0.01–0.01	-0.01–0.01

S8 Code

`solver.py` random fields, finite-volume flow, particle tracking with sub-grid exchange.

`models.py` inverse-Gaussian and exact MIM densities, binned likelihood fitting.

`run_pilot.py`, `exact_B.py`, `run_C.py`, `run_map.py`, `run_pe.py` simulations of mechanisms A, B, C, the classification map and the Péclet runs.

`analyze*.py`, `power_fix.py`, `gof_power.py`, `conc_test.py`, `robustness.py` tests and power analyses of the article.

`model_based.py`, `mrmt_lr.py`, `multirate2.py`, `round2_extra.py` analyses of this document.

`paper_figs.py` figures (written as Fig1–Fig8 in the order of the preprint; Fig3 is Fig. S1 here and Fig4–Fig8 are Figs. 3–7 of the article).