SNAP-DIFFUSION: FROM UNPAIRED SNAPSHOTS TO ANISOTROPIC DIFFUSION OPERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Unpaired population snapshots reveal marginal evolution but do not determine an unrestricted stochastic generator. We introduce SNAP-Diffusion, which recovers a structured anisotropic diffusion field by converting snapshot statistics and occupation quadrature into a finite weak Fokker–Planck system over a uniformly elliptic coefficient class. Restricted coercivity of the weak design yields identifiability within this class and deterministic stability to perturbations in both responses and design. Under fixed Fick drift, coefficient-recovery error further controls held-out density error and associated distributional and observable errors. In controlled two-dimensional transport, SNAP-Diffusion estimates one operator from short-horizon training releases and freezes it for downstream prediction. The recovered anisotropic operator improves prediction from unseen releases, remains accurate beyond the identification horizon, and transfers to first-passage functionals without refitting. A controlled FX experiment applies the same principle to state-dependent local correlation recovered from joint snapshots. On held-out tenors and contracts, SNAP-Diffusion reduces mean pricing error by 68% and 71% relative to constant- and time-dependent-correlation models and by 19% relative to same-class forward calibration. Together, the results connect weak snapshot identification of structured diffusion fields to reusable operators for downstream prediction.

1 INTRODUCTION

Population studies often observe independent samples at a few time points without linking individuals into trajectories, motivating the recovery of stochastic dynamics from aggregate snapshots (Ma et al., 2021; Guan et al., 2024). Such marginals do not determine an unrestricted drift–diffusion pair (Weinreb et al., 2018); identifiability therefore requires structure on the unknown dynamics. We focus on anisotropic diffusion, where the coefficient is a positive-definite matrix field rather than a scalar. As illustrated in Fig. 1, scalar diffusion controls only the local magnitude of spreading, whereas an anisotropic tensor also captures its direction and orientation. The resulting question is whether this structured field can be recovered from unpaired snapshots well enough to predict dynamics beyond those used for identification.

We introduce **SNAP-Diffusion**, which converts snapshot statistics and occupation quadrature into a finite weak Fokker–Planck system for a uniformly elliptic diffusion field. Multiple releases in the same medium provide complementary weak constraints on a shared coefficient field. SNAP-Diffusion solves the resulting constrained inverse problem directly, without reconstructing trajectories or repeatedly simulating candidate dynamics, and freezes the recovered operator for downstream prediction. This train-once, freeze-and-evaluate design separates coefficient identification from the tasks used to assess the recovered dynamics.

Our analysis characterizes when the finite snapshot design contains enough information to identify the declared coefficient class. Restricted coercivity yields identifiability and deterministic stability to perturbations in the weak responses and design. Under fixed Fick drift, coefficient-recovery error further controls prediction error for an unseen release, linking the inverse problem to held-out density and observable accuracy. The theory therefore connects the information contained in the weak snapshot system to the predictive accuracy of the recovered operator.

Contributions. Our contributions are threefold.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

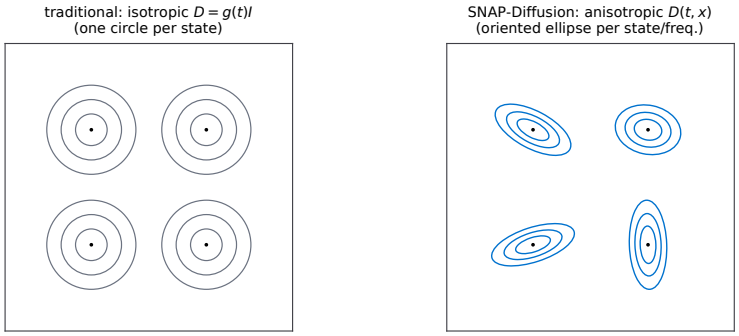


Figure 1: **Isotropic versus anisotropic diffusion.** A scalar coefficient assigns the same local variance in every direction (left), whereas a matrix-valued coefficient captures direction-dependent spreading and its local orientation (right).

Method. We introduce SNAP-Diffusion, a finite weak Fokker–Planck estimator for structured matrix-valued diffusion fields from unpaired temporal snapshots. The estimator requires neither trajectory reconstruction nor simulation-in-the-loop fitting and produces a single operator that can be reused without downstream refitting.

Theory. We characterize identifiability through restricted coercivity of the weak design and establish deterministic stability to response and design perturbations. We further propagate coefficient error to held-out density and observable errors, providing a direct theoretical link between snapshot identification and downstream prediction.

Evidence. Controlled transport experiments show that recovering spatially varying anisotropy improves prediction for unseen releases, longer horizons, and first-passage functionals; the advantage disappears under isotropic truth and deteriorates under basis misspecification, isolating recoverable directional structure as the source of the gain. On a common constant-diffusion benchmark, SNAP-Diffusion closely matches simulation-based forward calibration in downstream accuracy. A joint-snapshot FX experiment further shows that recovering state-dependent local correlation improves held-out pricing relative to simpler correlation models and same-class forward calibration.

2 Related Work

Snapshot inverse problems. Temporal marginals do not generally identify their generating stochastic dynamics (Weinreb et al., 2018). Weak Fokker–Planck constraints have been combined with adversarial transport objectives (Ma et al., 2021), while Guan et al. (2024) propose APPEX for linear drift and constant covariance, and then nn-APPEX (Guan et al., 2025) jointly estimates a nonlinear gradient drift and scalar diffusivity by iterative trajectory reconstruction . SNAP-Diffusion instead fixes a finite tensor class and finite weak design, exposing the coercivity condition that supports coefficient recovery and frozen-operator prediction.

Anisotropic recovery versus generation. DDPM, score-SDE, VDM, and EDM methods primarily construct generative processes, often with scalar covariance schedules (Ho et al., 2020; Song et al., 2020; Kingma et al., 2021; Karras et al., 2022). Our scalar models are coefficient-class controls under the same physical drift convention, not implementations of those generators. Riemannian diffusion models (De Bortoli et al., 2022; Huang et al., 2022) and Schrödinger bridges (De Bortoli et al., 2021) provide geometric or endpoint-constrained path laws, but identifying an unknown physical tensor still requires observable structure. Appendix D gives compact geometry and bridge diagnostics.

Local correlation. Dupire's construction links one-dimensional marginals to local volatility (Dupire, 1994), whereas two single-asset marginals do not determine joint correlation. Local-correlation calibration and multidimensional martingale constructions address this dependence (Guyon, 2013; Pallavicini, 2025; Kay, 2025). Our FX experiment uses joint snapshots to esti-

mate a structured local-correlation field and evaluates prices at held-out tenors; quote-based interfaces and implementation details are reported in Appendix G.

3 SNAP-DIFFUSION

3.1 FORWARD MODEL AND STRUCTURED COEFFICIENT CLASS

Consider the Itô diffusion

$$dX_t = b(t, X_t)\,dt + \sqrt{2D(t, X_t)}\,dW_t, \qquad D(t,x) \in \mathbb{S}^d_{++}. \tag{1}$$

Its density satisfies

$$\partial_t p_t = -\nabla \cdot (b p_t) + \sum_{i,j=1}^{d} \partial_{ij}(D_{ij} p_t). \tag{2}$$

Equivalently, defining $\mu = b - \operatorname{div} D$ with $(\operatorname{div} D)_i = \sum_j \partial_j D_{ij}$ gives the divergence form

$$\partial_t p_t = -\nabla \cdot (\mu p_t) + \nabla \cdot (D \nabla p_t). \tag{3}$$

The convention is that D is half the instantaneous covariance rate $2D$; a scalar score-SDE therefore has $D(t,x) = \frac{1}{2} g(t)^2 I$ such as Song et al. (2020).

We observe independent, unpaired samples from marginals $\{p_{t_r}\}_{r=0}^R$ at $0 = t_0 < \cdots < t_R = T$. The drift convention determines the inverse problem. In the fixed-Itô formulation, b is known and D varies; Equation 8 gives its weak system. In the fixed-Fick formulation, μ is known, and varying D_θ changes the Itô drift to $b_\theta = \mu + \operatorname{div} D_\theta$. The transport experiments fix $\mu = 0$, whereas the FX experiment fixes the Itô drift.

The convention determines which generator is held constant during coefficient variation. A spatially varying D contributes derivative terms when the same model is written in the other convention, so the two weak designs coincide only in special cases such as constant coefficients at zero drift. Every experiment therefore declares its convention before constructing A and uses that same operator for downstream propagation.

Assumption 3.1 (Regularity and boundary terms)**.** *The covariance $D^\star$ is continuously differentiable in space and satisfies $0 < \underline{\lambda} I \preceq D^\star(t,x) \preceq \overline{\lambda} I$. In the affine class below, D_{ref} and $B_1, \ldots, B_q$ are symmetric and uniformly belong to $W^{1,\infty}$ in space. The drift grows at most linearly. The densities and test functions are integrable enough for the stated weak identities. Boundary terms vanish because the test functions are compactly supported, the domain is a flat torus, or the boundary conditions enforce this property.*

The admissible model is the following affine, uniformly elliptic class. For known symmetric matrix-valued bases $B_k(t,x)$, define

$$D_\theta(t,x) = D_{\mathrm{ref}}(t,x) + \sum_{k=1}^{q} \theta_k B_k(t,x), \qquad \theta \in \Theta_{\mathrm{spd}}, \tag{4}$$

where

$$\Theta_{\mathrm{spd}} = \{\theta : \underline{\lambda} I \preceq D_\theta(t,x) \preceq \overline{\lambda} I \text{ for all } (t,x)\}. \tag{5}$$

This affine parameterization makes every weak observation linear in θ. The basis may encode constant, spatial, temporal, diagonal, or off-diagonal structure, while Θ_{spd} enforces one physical ellipticity range over the entire domain. The declared class is therefore both an inductive bias and the set on which identifiability is assessed. A neural form such as $LL^\top + \varepsilon I$ remains positive definite, but then local Jacobian conditions replace the global rank statements below.

3.2 WEAK SNAPSHOT SYSTEM

A weak snapshot system begins by choosing admissible test functions $\{\phi_\ell\}_{\ell=1}^L$ and intervals $I_r = [t_{r-1}, t_r]$. Define the endpoint increment by $\delta y_{\ell r} = \mathbb{E}_{p_{t_r}} \phi_\ell - \mathbb{E}_{p_{t_{r-1}}} \phi_\ell$. When the Itô drift b is held

fixed, define

$$a^{\mathrm{Ito}}_{\ell r,0} = \int_{I_r} \mathbb{E}_{p_t}[D_{\mathrm{ref}} : \nabla^2 \phi_\ell] dt, \tag{6}$$

$$A^{\mathrm{Ito}}_{\ell r,k} = \int_{I_r} \mathbb{E}_{p_t}[B_k : \nabla^2 \phi_\ell] dt, \tag{7}$$

$$\widetilde{y}^{\mathrm{Ito}}_{\ell r} = \delta y_{\ell r} - \int_{I_r} \mathbb{E}_{p_t}[b \cdot \nabla \phi_\ell] dt - a^{\mathrm{Ito}}_{\ell r,0}. \tag{8}$$

The resulting weak Fokker–Planck system is $\widetilde{y}^{\mathrm{Ito}} = A^{\mathrm{Ito}}\theta^\star$. If the Fick drift μ is fixed instead, integration by parts in Equation 3 gives

$$a^{\mathrm{Fick}}_{\ell r,0} = \int_{I_r} \mathbb{E}_{p_t}[\nabla \cdot (D_{\mathrm{ref}} \nabla \phi_\ell)] dt, \tag{9}$$

$$A^{\mathrm{Fick}}_{\ell r,k} = \int_{I_r} \mathbb{E}_{p_t}[\nabla \cdot (B_k \nabla \phi_\ell)] dt, \tag{10}$$

$$\widetilde{y}^{\mathrm{Fick}}_{\ell r} = \delta y_{\ell r} - \int_{I_r} \mathbb{E}_{p_t}[\mu \cdot \nabla \phi_\ell] dt - a^{\mathrm{Fick}}_{\ell r,0}, \tag{11}$$

so $\widetilde{y}^{\mathrm{Fick}} = A^{\mathrm{Fick}}\theta^\star$. For state-dependent D_θ, these conventions differ through first-derivative terms. We use A and y for the design and centered response $\widetilde{y}$ under the chosen convention. Mixing a response from one convention with a design from the other defines a different inverse problem.

Canonical quadratic loss. Given the weak design A and response y, quadratic SNAP-Diffusion estimates

$$\widehat{\theta}_{\mathrm{quad}} \in \arg\min_{\theta \in \Theta_{\mathrm{spd}}} \mathcal{L}_{\mathrm{quad}}(\theta), \qquad \mathcal{L}_{\mathrm{quad}}(\theta) = \|A\theta - y\|^2_W, \tag{12}$$

where $\|z\|^2_W = z^\top W z$ for a fixed $W \succ 0$; $W = I$ recovers ordinary least squares. The objective directly fits the finite weak system, with W allowing fixed rescaling of its moment equations. Since D_θ is affine and Θ_{spd} is convex, the resulting estimator is a constrained convex quadratic problem and is the estimator analyzed in Theorem 3.4. No downstream quantity or forward simulation enters the objective.

3.3 Identifiability and stability

Proposition 3.2 (Finite-observation non-identifiability)**.** *Suppose $\theta^\star$ is an interior point of Θ_{spd}. If $\ker(A) \neq \{0\}$, then $\theta^\star$ is not identifiable from the chosen weak observations. Specifically, for every nonzero $h \in \ker(A)$ there is an $\epsilon_h > 0$ such that $\theta^\star + \epsilon h \in \Theta_{\mathrm{spd}}$ and $A(\theta^\star + \epsilon h) = A\theta^\star$ whenever $|\epsilon| < \epsilon_h$.*

In particular, if $q > LR$, then $\mathrm{rank}(A) \leq LR < q$ and these weak moments cannot uniquely recover all parameters without additional structure. The proposition makes the limitation operational: adding particles reduces empirical noise but cannot reveal a direction in $\ker(A)$, whereas a new release or test function can help if it changes that kernel.

Let $\|h\|^2_G = h^\top G h$ and $\|z\|^2_W = z^\top W z$ for $G, W \succ 0$. Define the feasible difference set $\mathcal{C}(\theta^\star) = \{h : \theta^\star + h \in \Theta_{\mathrm{spd}}\}$, which need not be a cone.

Assumption 3.3 (Restricted coercivity)**.** *There is a $\kappa > 0$ such that $\|Ah\|_W \geq \kappa \|h\|_G$ for every $h \in \mathcal{C}(\theta^\star)$.*

Theorem 3.4 (Identifiability and deterministic stability)**.** *Under Assumption 3.3, exact weak observations uniquely identify $\theta^\star$ in Θ_{spd}. If $\widehat{y} = A\theta^\star + \eta$ and*

$$\widehat{\theta} \in \arg\min_{\theta \in \Theta_{\mathrm{spd}}} \|A\theta - \widehat{y}\|^2_W, \tag{13}$$

then

$$\|\widehat{\theta} - \theta^\star\|_G \leq \frac{2}{\kappa} \|\eta\|_W. \tag{14}$$

Let $\widehat{A} = A + E$, let $\widehat{y} = A\theta^\star + \eta$, and define $\epsilon_A = \sup_{h \in \mathcal{C}(\theta^\star)\setminus\{0\}} \|Eh\|_W / \|h\|_G$.

Corollary 3.5 (Perturbed design). *If $\epsilon_A < \kappa$ and*

$$\widehat{\theta} \in \arg\min_{\theta \in \Theta_{\mathrm{spd}}} \|\widehat{A}\theta - \widehat{y}\|_W^2, \tag{15}$$

then every such constrained least-squares solution satisfies

$$\|\widehat{\theta} - \theta^\star\|_G \le \frac{2\|\eta - E\theta^\star\|_W}{\kappa - \epsilon_A}. \tag{16}$$

Restricted coercivity requires separation only over feasible Symmetric Positive Definite(SPD) coefficient differences, with κ quantifying the conditioning of the weak design on this set.

Theorem 3.4 isolates this design conditioning through κ, while Corollary 3.5 places response error and quadrature-induced design error in the numerator and denominator, respectively. Appendix A contains all proofs, and Appendix E reports the robust objective as an empirical variant.

3.4 Downstream prediction guarantee

For the affine class, define

$$\beta_{0,G} = \sup_{h \neq 0} \frac{\|\sum_{k=1}^q h_k B_k\|_{L^\infty([0,T]\times\Omega;\mathrm{op})}}{\|h\|_G}. \tag{17}$$

Then, for $0 \le t \le T$,

$$\varepsilon_D(t) := \operatorname*{ess\,sup}_{0\le s\le t,\, x\in\Omega} \|D_{\widehat{\theta}}(s,x) - D_{\theta^\star}(s,x)\|_{\mathrm{op}} \le \beta_{0,G}\|\widehat{\theta} - \theta^\star\|_G. \tag{18}$$

The next result bounds density error for an unseen release using only the coefficient error; later marginals from that release are not used for fitting.

Proposition 3.6 (Held-out density stability under fixed Fick drift). *Let Ω be a flat torus, and let $p_t^\star$ and $\widehat{p}_t$ be probability densities solving Equation 3 from the same initial density under the same Fick drift μ, with diffusion tensors $D^\star$ and $\widehat{D}$, respectively. Assume $p^\star \in L^2(0,T;H^1(\Omega))$, $\widehat{D} \succeq \lambda I$, and $M_\mu := \|\nabla\cdot\mu\|_{L^\infty([0,T]\times\Omega)} < \infty$. Then*

$$\|\widehat{p}_t - p_t^\star\|_2 \le \frac{\varepsilon_D(t)}{\sqrt{\lambda}} \left[\int_0^t e^{M_\mu(t-s)} \|\nabla p_s^\star\|_2^2 \, ds\right]^{1/2}. \tag{19}$$

Consequently, with $\mathrm{TV}(p,q) = \frac{1}{2}\|p-q\|_1$,

$$\mathrm{TV}(\widehat{p}_t, p_t^\star) \le \tfrac{1}{2}|\Omega|^{1/2}\|\widehat{p}_t - p_t^\star\|_2, \tag{20}$$

$$|\widehat{P}_t(R) - P_t^\star(R)| \le |R|^{1/2}\|\widehat{p}_t - p_t^\star\|_2, \tag{21}$$

$$|\mathbb{E}_{\widehat{p}_t} f - \mathbb{E}_{p_t^\star} f| \le 2\|f\|_\infty \,\mathrm{TV}(\widehat{p}_t, p_t^\star). \tag{22}$$

Moreover, for the torus metric, $W_2(\widehat{p}_t, p_t^\star) \le \mathrm{diam}(\Omega)\sqrt{\mathrm{TV}(\widehat{p}_t, p_t^\star)}$.

Therefore, Section 3.3 completes the downstream stability chain

$$\boxed{\text{snapshots} \longrightarrow (A,y) \longrightarrow \widehat{\theta} \longrightarrow D_{\widehat{\theta}} \longrightarrow \widehat{p}_t \longrightarrow \text{downstream observables}}$$

Restricted coercivity controls the first step through Theorem 3.4 and Corollary 3.5; Equation 19 converts parameter error into diffusion-field error; and Proposition 3.6 propagates this error to the density of an unseen release and hence to total variation, regional probabilities, Wasserstein distance, and bounded observables. For anisotropic truth, restricting the coefficient class to isotropic fields additionally incurs the projection gap in Equations 30–31. Proofs and the corresponding excess-risk result are given in Appendix A.

Table 1: **Experiment 1: prediction from unseen releases.** Errors are averaged over releases and times within each seed, then over ten seeds within each of five media, and finally across media. The last two columns evaluate the recovered coefficient field. Lower is better; bold marks the best non-oracle mean.

Method	Hellinger	Reg. MAE	Exposure error	Rel.-F (D)	L^∞-op (D)
Oracle coefficient field	0.000305	0.000023	0.000022	0	0
Constant scalar	0.04855	0.003094	0.003046	0.19996	0.01200
Spatial scalar	0.11060	0.007417	0.007358	0.39980	0.03112
Constant Symmetric Positive Definite(SPD)	0.04824	0.003050	0.003015	0.20005	0.01221
Fixed-basis diagonal	0.02769	0.001689	0.001646	0.11329	0.007747
Forward calibration (same structure)	0.008794	0.000862	0.000820	0.03436	0.001790
SNAP-Diffusion (ours), quadratic loss	**0.004196**	**0.000334**	**0.000319**	**0.01628**	**0.001102**

4 Experimental Setup

The central protocol fits one coefficient field per medium from four training releases and freezes it for all downstream tasks. Independent unpaired snapshots determine the weak responses and occupation-based design, and quadratic SNAP-Diffusion solves the resulting constrained inverse problem once. Unseen releases and first-passage configurations are excluded from estimation and used only for downstream evaluation. The resulting tests measure whether a field identified from population snapshots transfers across new initial conditions and functionals.

5 Frozen-Operator Transport Experiments

We evaluate whether a diffusion operator identified from short-horizon periodic snapshots remains predictive outside the identification setting. After fitting, the operator is frozen and evaluated on unseen releases, longer horizons, and absorbing-boundary first-passage tasks.

The ground-truth diffusion belongs to the three-parameter periodic family

$$D_\theta(x,y) = \theta_0 I + \theta_1 \cos(2\pi x)\cos(2\pi y)\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} + \theta_2 \sin(2\pi x)\sin(2\pi y)\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}. \tag{23}$$

Four Gaussian training releases are observed at five times on the periodic unit square. We evaluate five uniformly elliptic media over ten observation seeds, with fixed Fick drift $\mu = 0$.

The baselines progressively restrict the diffusion structure. Constant and spatial scalar fields are isotropic; constant Symmetric Positive Definite(SPD) permits one global tensor; and fixed-basis diagonal fields capture spatially varying directional amplitudes without the full off-diagonal structure.

5.1 Prediction from unseen releases

Eight Gaussian initial densities, all centered outside the training release set, test transfer to new initial conditions. We evaluate density shape by Hellinger distance, regional probability by MAE, and cumulative exposure over both observed and longer time horizons. Training occupancy is used only to stratify test locations by coverage; full settings are given in Appendix E.

SNAP-Diffusion recovers the anisotropic field and improves prediction from unseen releases. SNAP-Diffusion achieves the lowest non-oracle error on all five metrics in Table 1. Relative to forward calibration of the same coefficient family, it reduces mean Hellinger error from 0.008794 to 0.004196, regional MAE from 8.62×10^{-4} to 3.34×10^{-4}, and exposure error from 8.20×10^{-4} to 3.19×10^{-4}. Its coefficient error is also lower, with relative Frobenius error 0.01628 versus 0.03436.

The restricted baselines show why the full anisotropic field matters. Constant scalar and constant Symmetric Positive Definite(SPD) models have nearly identical Hellinger errors (0.04855 and 0.04824), so a single global tensor does not capture the spatially changing anisotropy. Allowing spatial variation without directional structure performs worse (0.11060), whereas the fixed-basis diagonal model improves to 0.02769 by recovering part of the directional variation. SNAP-Diffusion, which represents the full ground-truth family, reduces this error to 0.004196. The same ordering appears in regional probability, cumulative exposure, and coefficient recovery.

Table 2: **Experiment 2: first-passage prediction and release decisions.** Errors are averaged over layouts and start points within each seed, followed by the same seed- and medium-level aggregation as in Table 1. Lower is better; bold marks the best non-oracle mean.

Method	Committor MAE	MFPT rel.-L^1	Decision regret
Oracle coefficient field	0	0	0
Constant scalar	0.01004	0.03322	0.001821
Spatial scalar	0.01972	0.07741	0.004604
Constant Symmetric Positive Definite(SPD)	0.009991	0.03318	0.001821
Fixed-basis diagonal	0.006057	0.02955	0.000258
Forward calibration (same structure)	**0.000480**	0.03002	0.000045
SNAP-Diffusion (ours), quadratic loss	0.000653	**0.006957**	**0.000020**

The gain also persists beyond the identification horizon: the operator is estimated from snapshots through $t = 0.20$ and remains accurate when propagated to $t = 0.30$ and 0.40. Together, these results show that recovering the spatially varying anisotropic field improves prediction across unseen initial conditions, downstream summaries, and longer horizons.

5.2 First-passage and release decisions

We next change the boundary-value problem while keeping the recovered operator fixed. Three pairs of absorbing disks define committor and mean first-passage-time (MFPT) tasks,

$$q(x) = P_x(\tau_B < \tau_A), \qquad m(x) = \mathbb{E}_x[\min(\tau_A, \tau_B)].$$

A fixed set of candidate releases additionally converts the predicted committor into a decision and evaluates regret under the ground-truth medium. None of these absorbing configurations or functionals is used during identification.

The recovered anisotropic operator transfers to first-passage functionals. SNAP-Diffusion attains committor MAE 6.53×10^{-4} and the lowest non-oracle MFPT error, 0.006957. Forward calibration is slightly better on the committor (4.80×10^{-4}) but has substantially larger MFPT error (0.03002). SNAP-Diffusion also gives the lowest mean decision regret, 2.0×10^{-5}.

The coefficient-class controls again expose the role of spatially varying anisotropy. Constant scalar and constant Symmetric Positive Definite(SPD) models both have committor error near 10^{-2} and MFPT error near 3.3×10^{-2}; a global anisotropic tensor therefore offers little improvement. The fixed-basis diagonal model lowers the committor error to 0.006057 and MFPT error to 0.02955, while SNAP-Diffusion reduces them further to 0.000653 and 0.006957. Spatial variation alone is insufficient: the spatial scalar model has the largest errors on both first-passage quantities.

Controls confirm that the gain comes from recoverable anisotropy. Under isotropic truth, the advantage of the anisotropic model disappears; when the declared basis or training releases fail to represent or excite the relevant anisotropic directions, recovery deteriorates. Full ablations and numerical diagnostics are reported in Appendix E.

6 Benchmark and FX Experiments

6.1 Common constant-diffusion benchmark

We compare SNAP-Diffusion with APPEX Guan et al. (2024), its nonlinear gradient-drift extension nn-APPEX Guan et al. (2025), Ma/FPP Ma et al. (2021), and forward calibration on a common setting representable by all methods: zero-drift diffusion on $\mathbb{R}^2$ with three constant diagonal ground truths. Each method receives independent point snapshots from two releases at $t \in \{0, .06, .12\}$, with 64 samples per release and time, and is fitted before any downstream evaluation. Two unseen releases test density propagation, followed by common absorbing-boundary tests of committor and MFPT. Results aggregate nine predeclared truth–seed units; full protocols and paired intervals are given in Appendix F.

Figure 2: **Common-domain Experiment 2: first-arrival prediction.**

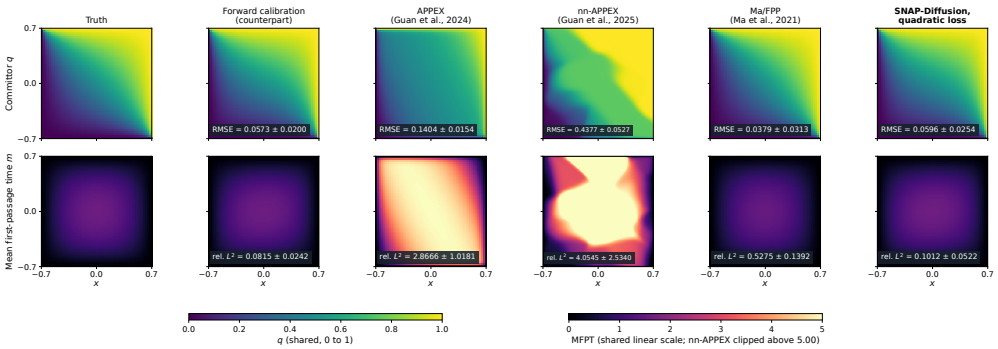


Table 3: **Common benchmark: unseen-release prediction.** Mean ± sample SD over nine truth–seed units after averaging two unseen releases and three evaluation times within each unit. Lower is better.

Method	Gaussian Hellinger	Gaussian W_2
APPEX Guan et al. (2024)	0.9168 ± 0.0175	0.8230 ± 0.0137
nn-APPEX Guan et al. (2025)	0.6772 ± 0.1054	0.5918 ± 0.2298
Ma/FPP Ma et al. (2021)	0.1613 ± 0.0652	0.1130 ± 0.0538
Forward calibration	0.1038 ± 0.0257	0.06322 ± 0.01189
SNAP-Diffusion (ours)	$\mathbf{0.1063 \pm 0.0369}$	$\mathbf{0.06426 \pm 0.01867}$

SNAP-Diffusion matches forward calibration closely on the common constant-diffusion benchmark. Forward calibration gives the lowest unseen-release errors, with Hellinger 0.1038 and W_2 0.06322, while SNAP-Diffusion obtains 0.1063 and 0.06426—remaining quite close to the frontier. APPEX gives 0.9168 and 0.8230; nn-APPEX improves these to 0.6772 and 0.5918 but remains behind Ma/FPP. For first passage, Ma/FPP gives the lowest committor RMSE (0.0379), whereas forward calibration gives the lowest MFPT error (0.0815); SNAP-Diffusion follows at 0.0596 and 0.1012, respectively. SNAP-Diffusion and forward calibration have the same mean decision regret. Thus, on a coefficient class shared by all methods, SNAP-Diffusion achieves propagation and first-passage accuracy close to simulation-based forward calibration while estimating the coefficients directly from weak snapshot equations. nn-APPEX underperforms APPEX in committor RMSE ((0.438) vs. (0.140)) and MFPT error ((4.055) vs. (2.867)), primarily because sparse observations and an unregularized neural potential induce spurious drift wells as analyzed in Appendix F.1.

6.2 Structured local correlation

We next test whether the same snapshot-based recovery principle extends to state-dependent local correlation. Let $X_t = \mathrm{EUR/USD}$, $Y_t = \mathrm{GBP/USD}$, and $Z_t = X_t/Y_t = \mathrm{EUR/GBP}$. With $u = \log(X_t/X_0)$ and $v = \log(Y_t/Y_0)$, we use

$$D_\theta(t,u,v) = \begin{bmatrix} \frac{1}{2}\sigma_X^2 & \frac{1}{2}\rho_\theta\sigma_X\sigma_Y \\ \frac{1}{2}\rho_\theta\sigma_X\sigma_Y & \frac{1}{2}\sigma_Y^2 \end{bmatrix}, \qquad \rho_\theta = \theta_0 + \theta_1\Big(\frac{2t}{T_{\max}} - 1\Big) + \theta_2 \tanh\tfrac{u+v}{c} + \theta_3 \tanh\tfrac{u-v}{c}. \tag{24}$$

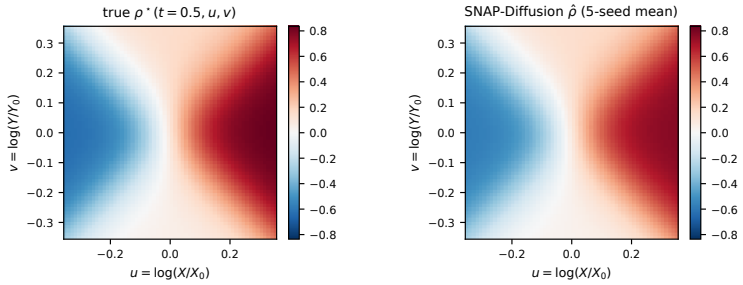
Here $T_{\max} = 1$ year, c is fixed, and the marginal volatilities and log drifts are known. The four coefficients encode baseline correlation, time variation, common moves, and relative moves. The constraint $\sum_k |\theta_k| \le 0.9$ ensures $|\rho_\theta| \le 0.9$ and hence $D_\theta \succ 0$.

Each experiment provides $5{,}000$ unpaired joint samples of (u, v) at 0, 1, 3, 6, and 12 months. SNAP-Diffusion estimates the four correlation coefficients from weak cross moments. Pricing is then evaluated at held-out 2-, 4-, and 9-month tenors using EUR/GBP calls and two-asset worst-of options, which probe cross-asset dependence not used in fitting. We compare SNAP-Diffusion with constant and time-dependent correlation models, same-class forward calibration, and interpolation of empirical snapshot correlations (Table 4).

Table 4: **FX triangle: held-out pricing error.** Absolute error ($\times 10^3$ USD per unit notional) averaged over three held-out tenors and seven contracts within each seed; entries are mean $\pm$ sample SD over five seeds. Lower is better; bold marks the lowest mean.

Truth	SNAP-Diffusion	const ρ	time-dep ρ	fwd-calib	marg. interp
const ρ	$0.438\pm.205$	$0.300\pm.275$	$0.369\pm.150$	$0.455\pm.127$	$\mathbf{0.192}\pm.163$
time-dep ρ	$0.127\pm.031$	$3.276\pm.129$	$\mathbf{0.087}\pm.055$	$0.144\pm.071$	$0.166\pm.092$
state-dep ρ	$\mathbf{0.241}\pm.077$	$0.764\pm.019$	$0.838\pm.152$	$0.298\pm.101$	$0.821\pm.126$
saddle (mis.)	$0.512\pm.156$	$0.696\pm.062$	$0.459\pm.131$	$0.483\pm.144$	$\mathbf{0.226}\pm.177$
cos-wave (mis.)	$0.371\pm.111$	$1.305\pm.328$	$0.348\pm.122$	$0.392\pm.037$	$\mathbf{0.241}\pm.122$

Figure 3: **State-dependent local-correlation recovery at $t = 0.5$.** Ground truth (left) and the field induced by the mean SNAP-Diffusion parameter estimate over five seeds (right). The relative grid-L^2 error of the displayed mean field is 0.086.



SNAP-Diffusion is most effective when pricing requires state-dependent correlation. For the state-dependent truth, SNAP-Diffusion gives the lowest mean held-out pricing error, 0.241×10^{-3}, reducing error by 68% relative to constant correlation, 71% relative to time-dependent correlation, and 19% relative to same-class forward calibration. The mean per-seed relative grid-L^2 field error is 0.129 ± 0.053, and the field induced by the mean parameter estimate has error 0.086 (Figure 3). Thus, state dependence recovered solely from joint snapshots improves pricing at unseen tenors and for contracts absent from fitting.

Model complexity helps only when it matches the correlation structure. For purely time-dependent truth, the correctly specified time-dependent model gives the lowest pricing error (0.087 versus 0.127 for SNAP-Diffusion). Under constant truth, the simpler constant and interpolation baselines also outperform SNAP-Diffusion. These results complement the state-dependent experiment: the four-parameter field is useful when state dependence is present, but offers no advantage when a lower-dimensional correlation model already represents the truth.

When the true field lies outside the fitted basis, coefficient recovery deteriorates sharply: SNAP-Diffusion's relative field error rises to 0.85–0.89, and empirical interpolation gives the lowest pricing error for both misspecified truths. This identifies basis specification as the main structural limitation of the estimator. Additional FX protocols, ablations, and numerical diagnostics are given in Appendix G.

7 Conclusion

SNAP-Diffusion converts unpaired snapshots into a finite weak system for recovering structured anisotropic diffusion fields and freezes the resulting operator for downstream prediction. Restricted coercivity connects weak-design information to stable coefficient recovery, and the downstream theory propagates this error to held-out predictions. Controlled transport and FX experiments show that the recovered operator transfers across initial conditions and functionals when directional or state-dependent structure matters. Together, the theory and experiments identify two requirements for snapshot-based diffusion recovery: the coefficient class must represent the relevant structure, and the observed releases must excite its identifiable directions.

LLM USE DISCLOSURE

LLM Agent tool (Qoder, https://qoder.com/) was employed in several stages of preparing this manuscript. It was used for manuscript-level editing, including restructuring passages, removing redundant or overly qualified statements, improving consistency of terminology, suggesting alternative title and section-level organization.

In the parts of Abstract and Introduction, it assisted with revision, including improving clarity, concision, terminology, and the presentation of the paper. In the methodological sections, including Sections 3.1 and 3.2, Qoder was used to improve the organization and presentation of the weak Fokker–Planck formulation and estimator. In the theoretical sections, including 3.3 and 3.4, the Agent is used to check the presentation of assumptions, mathematical claims, and the logical connections among weak-design conditioning, coefficient recovery, diffusion-field error, and downstream prediction. The mathematical results and the proofs themselves were derived and verified by the authors; generative AI was not used to produce the proofs.

For the experimental validation, the authors designed the experimental protocol and independently implemented the proof-of-concept (PoC) code. The AI agent was not used to generate the synthetic datasets, which were produced through numerical simulation. Real FX market data used for the empirical analysis were obtained separately from (COBE https://www.cboe.com/). The AI agent was used to accelerate the subsequent coarse-to-fine implementation and code refinement.

All Agent-assisted codes, statements, interpretations, and methodological descriptions were reviewed by the authors against the underlying derivations, implementations, and numerical results. The authors take full responsibility for the final content of the paper, including all claims, results, and AI-assisted material.

ETHICS STATEMENT

This work develops methodology for recovering structured diffusion operators from unpaired snapshots. The transport, common-domain, and FX experiments considered in this paper are controlled numerical studies constructed to evaluate coefficient recovery and downstream prediction. The FX experiment is methodological: it studies state-dependent local-correlation recovery and held-out option pricing under prescribed model settings, rather than making trading or investment recommendations.

The main limitation relevant to downstream use is structural: recovery depends on whether the declared coefficient class represents the relevant diffusion structure and whether the observed releases sufficiently excite its identifiable directions. The paper evaluates these effects through isotropic, coverage, and basis-misspecification controls. Applications to consequential real-world settings would require separate validation of the coefficient class, observation process, and downstream decision model.

REPRODUCIBILITY STATEMENT

We provide the mathematical definitions, assumptions, and estimators needed to reproduce SNAP-Diffusion in the main text. The weak Itô and Fick systems specify the construction of the design and response, and the canonical quadratic objective specifies the constrained estimator. Proofs of the identifiability, stability, downstream-prediction, and representation results are provided in Appendix A.

The experimental protocols are specified together with their aggregation rules and evaluation metrics. Appendix E provides the transport media, training and held-out releases, observation times, test functions and quadrature, generator construction, first-passage layouts, candidate sets, random seeds, ablations, and numerical checks. Appendix F specifies the common constant-diffusion benchmark, including the adaptations and evaluation protocol used for APPEX Guan et al. (2024), nn-APPEX Guan et al. (2025), Ma/FPP Ma et al. (2021), forward calibration, and SNAP-Diffusion. Appendix G provides the FX ground truths, snapshot-generation and pricing protocols, contract and tenor grids, parameter constraints, Monte-Carlo settings, random seeds, solver details, and numerical-sensitivity checks. These specifications are intended to make the reported theoretical and empirical results reproducible without relying on unreported tuning or downstream refitting.

REFERENCES

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17695–17709. Curran Associates, Inc., 2021.

Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. In *Advances in Neural Information Processing Systems*, volume 35, pp. 2406–2422. Curran Associates, Inc., 2022. doi: 10.52202/068431-0175.

Bruno Dupire. Pricing with a smile. *Risk*, 7(1):18–20, January 1994.

Vincent Guan, Joseph Janssen, Hossein Rahmani, Andrew Warren, Stephen Zhang, Elina Robeva, and Geoffrey Schiebinger. Identifying drift, diffusion, and causal structure from temporal snapshots. *arXiv preprint arXiv:2410.22729*, 2024.

Vincent Guan, Joseph Janssen, Nicolas Lanzetti, Antonio Terpin, Geoffrey Schiebinger, and Elina Robeva. Gradient-flow sdes have unique transient population dynamics. *arXiv preprint arXiv:2505.21770*, 2025.

Julien Guyon. A new class of local correlation models. *Available at SSRN 2283419*, 2013.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020.

Chin-Wei Huang, Milad Aghajohari, Joey Bose, Prakash Panangaden, and Aaron Courville. Riemannian diffusion models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 2750–2761. Curran Associates, Inc., 2022. doi: 10.52202/068431-0199.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26565–26577. Curran Associates, Inc., 2022. doi: 10.52202/068431-1926.

Michael M Kay. Multi-dimensional martingales from mutual information. *arXiv preprint arXiv:2512.16544*, 2025.

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. In *Advances in Neural Information Processing Systems*, volume 34, pp. 21696–21707. Curran Associates, Inc., 2021.

Shaojun Ma, Shu Liu, Hongyuan Zha, and Haomin Zhou. Learning stochastic behaviour from aggregate data. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7258–7267. PMLR, 2021.

Andrea Pallavicini. Pricing quanto and composite contracts with local-correlation models. *arXiv preprint arXiv:2501.07200*, 2025.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Caleb Weinreb, Samuel Wolock, Betsabeh K Tusi, Merav Socolovsky, and Allon M Klein. Fundamental limits on dynamic inference from single-cell snapshots. *Proceedings of the National Academy of Sciences*, 115(10):E2467–E2476, 2018.

A PROOFS

A.1 FINITE-OBSERVATION NON-IDENTIFIABILITY

Proof of Proposition 3.2.

Proof. The equality follows from $Ah = 0$. Because $\theta^\star$ is an interior point and D_θ is affine and continuous in θ, sufficiently small perturbations in either direction retain uniform positive definiteness. □

A.2 RESTRICTED-COERCIVITY STABILITY

Proof of Theorem 3.4.

Proof. For exact observations, $A(\theta - \theta^\star) = 0$ and restricted coercivity imply $\theta = \theta^\star$. For perturbed observations, optimality gives $\|A\widehat{\theta} - \widehat{y}\|_W \le \|\eta\|_W$. The triangle inequality then yields $\|A(\widehat{\theta} - \theta^\star)\|_W \le 2\|\eta\|_W$; apply Assumption 3.3. □

Proof of Corollary 3.5.

Proof. For feasible differences, $\|\widehat{A}h\|_W \ge (\kappa - \epsilon_A)\|h\|_G$. Apply the proof of Theorem 3.4 with the effective residual $\eta - E\theta^\star$. □

A.3 HELD-OUT DENSITY STABILITY

Proof of Proposition 3.6.

Proof. Set $w = \widehat{p} - p^\star$ and $\Delta D = \widehat{D} - D^\star$. Subtracting the two Fick equations gives

$$\partial_t w = -\nabla \cdot (\mu w) + \nabla \cdot (\widehat{D}\nabla w) + \nabla \cdot (\Delta D \nabla p^\star). \tag{25}$$

Testing against w on the torus and integrating by parts yields

$$\frac{1}{2}\frac{d}{dt}\|w\|_2^2 = -\frac{1}{2}\int_\Omega (\nabla \cdot \mu) w^2 - \int_\Omega \nabla w^\top \widehat{D} \nabla w - \int_\Omega \nabla w^\top \Delta D \nabla p^\star \tag{26}$$

$$\le \frac{M_\mu}{2}\|w\|_2^2 - \frac{\lambda}{2}\|\nabla w\|_2^2 + \frac{\varepsilon_D(t)^2}{2\lambda}\|\nabla p_t^\star\|_2^2. \tag{27}$$

Hence

$$\frac{d}{dt}\|w\|_2^2 \le M_\mu \|w\|_2^2 + \frac{\varepsilon_D(t)^2}{\lambda}\|\nabla p_t^\star\|_2^2. \tag{28}$$

Because $w(0) = 0$, Grönwall's inequality gives Equation 19. The remaining statements follow from Cauchy–Schwarz, the definition of total variation, and the diameter coupling bound for two equal-mass probability measures. □

A.4 ISOTROPIC PROJECTION AND EXCESS RISK

Fix the truth-induced occupation measure $\rho^\star(dt, dx) = p_t^\star(x)\, dt\, dx$ and use the corresponding Hilbert norm

$$\|H\|_{\mathcal{H}(\rho^\star)}^2 = \int_0^T \mathbb{E}_{p_t^\star}\|H(t, X_t)\|_F^2 dt. \tag{29}$$

Let $\mathcal{V}_{\text{iso}}$ be the isotropic subspace in the model class and

$$\Delta_{\text{iso}}(D^\star) = \inf_{D \in \mathcal{V}_{\text{iso}}} \|D - D^\star\|_{\mathcal{H}(\rho^\star)}. \tag{30}$$

If the scalar field is unrestricted within the ambient function space, the pointwise projection is $(\operatorname{tr} D^\star/d)I$ and

$$\Delta_{\text{iso}}^2 = \int_0^T \mathbb{E}_{p_t^\star}\left\| D^\star - \frac{\operatorname{tr} D^\star}{d} I \right\|_F^2 dt. \tag{31}$$

This quantity measures the traceless anisotropic energy under the truth occupation measure. If the scalar family is parameter restricted, the relevant gap is Equation 30, not Equation 31.

Corollary A.1 (Conditional isotropic excess-risk gap)**.** *Let $\mathcal{J}$ be a population objective with minimizer $D^\star$ and assume quadratic growth*

$$\mathcal{J}(D) - \mathcal{J}(D^\star) \geq \frac{\mu_J}{2}\|D - D^\star\|^2_{\mathcal{H}(\rho^\star)} \tag{32}$$

for all considered D. Then

$$\inf_{D\in\mathcal{V}_{\mathrm{iso}}} \mathcal{J}(D) - \mathcal{J}(D^\star) \geq \frac{\mu_J}{2}\Delta_{\mathrm{iso}}(D^\star)^2. \tag{33}$$

Proof. Apply quadratic growth to every $D \in \mathcal{V}_{\mathrm{iso}}$ and take the infimum. □

At the observation level, the best isotropic fit $\theta_1 = \theta_2 = 0$ leaves residual 0.2997 on weak data from the anisotropic truth. The full three-parameter model has residual 0. Separately, the density-averaged L^2 norm of the traceless field is 0.0096. These values use different norms. The nonzero weak residual is consistent with an isotropic approximation gap, but equals Δ_{iso} only under an additional coercivity or isometry relation between the weak observation operator and $\mathcal{H}(\rho^\star)$.

B Weak-Design and Numerical Verification

B.1 Weak observations from sparse marginals

Empirical snapshots perturb the response $\widehat{y}$ and design $\widehat{A}$ through distinct sampling and quadrature channels. For $\varphi_\ell \in C_c^2(\Omega)$, Itô's formula integrated over $I_r = [t_{r-1}, t_r]$ gives

$$\begin{aligned}&\mathbb{E}_{p_{t_r}}\varphi_\ell - \mathbb{E}_{p_{t_{r-1}}}\varphi_\ell - \int_{I_r}\mathbb{E}_{p_t}[b\cdot\nabla\varphi_\ell]dt\\ &= \int_{I_r}\mathbb{E}_{p_t}[D^\star : \nabla^2\varphi_\ell]dt.\end{aligned} \tag{34}$$

Subtract the contribution of D_{ref} from the left-hand side and denote the result by $y_{\ell r}$. Define

$$A_{(\ell,r),k} = \int_{I_r}\mathbb{E}_{p_t}[B_k(t, X_t) : \nabla^2\varphi_\ell(X_t)]dt. \tag{35}$$

If $D^\star = D_{\theta^\star}$ and the occupation integrals are exact, then $y = A\theta^\star$.

Endpoint snapshots do not determine the interval occupation measure $p_t(x)dt\,dx$. Quadrature and empirical density estimates produce

$$\widehat{y} = A\theta^\star + \eta, \qquad \eta = \eta_{\mathrm{sample}} + \eta_{\mathrm{quad}} + \eta_{\mathrm{model}} + \eta_b. \tag{36}$$

Figure 14 (left) isolates η_{quad}: with observations fixed, refining the occupation quadrature monotonically reduces standardized trapezoid bias.

B.2 Empirical stability bound

We evaluate Theorem 3.4 with 48 weak observations: 3 intervals and 16 Fourier test functions. The design singular values are $(55.27,\ 21.14,\ 10.41)$, so $\kappa = 10.41$. We add Gaussian observation noise η with $\sigma \in \{10^{-4}, 3\times10^{-4}, 10^{-3}, 3\times10^{-3}, 10^{-2}\}$ and run 20 repetitions per level. Every constrained least-squares solution satisfies

$$\|\widehat{\theta} - \theta^\star\|_2 \ \leq\ \frac{2\|\eta\|_2}{\kappa} \tag{37}$$

The inequality holds in every run. At $\sigma = 0$, the recovery error is 1.2×10^{-17}, at machine precision. The worst-case bound is 13–17 times the observed error but captures the $\mathcal{O}(\|\eta\|)$ scaling.

B.3 CONSTRUCTIVE NON-IDENTIFIABILITY

To illustrate Proposition 3.2, we restrict the design to one interval and one Fourier frequency. The resulting matrix has 2 rows, rank 1, and a 2-dimensional kernel. Let h be the right singular vector for the smallest singular value. Then $\|Ah\|_2 = 9.8\times10^{-17}$, while $\theta^\star + 0.005\,h = (0.0299,\ 0.0070,\ 0.0080)$ remains strictly Symmetric Positive Definite(SPD) and has 8.1% relative tensor Frobenius error. The design therefore cannot distinguish these two tensors.

B.4 SNAPSHOT SCALING

Fixing $\sigma = 3 \times 10^{-3}$ and increasing the number of observed intervals improves the observation rank and monotonically reduces recovery error:

Intervals	1	2	3	4	5
κ	4.48	5.43	6.49	10.11	13.54
Rel. tensor error	0.0134	0.0108	0.0076	0.0055	0.0037

This is the finite-dimensional counterpart of Proposition 3.2: independent snapshots increase κ and reduce the stability constant $2/\kappa$.

B.5 ITÔ/FICK CONVENTION AUDIT

Additional recovery validation. The independent validation case uses truth 0, seed 11, a 32^2 truth grid, a 16^2 fitting grid, and $4{,}000$ particles per snapshot. The design retains rank three in all four whole-release leave-one-release-out folds. The maximum column-normalized condition number is 1.357 (rounded to 1.36). The maximum absolute and relative parameter errors are 0.00438 and 0.132. Replacing the correct Fick action $\nabla \cdot (B_k \nabla \phi)$ with the zero-drift Itô action $B_k : \nabla^2 \phi$ changes the design by relative Frobenius norm 0.124 and raises parameter error from 8.90×10^{-4} to 2.25×10^{-3}, a factor of 2.53. Two independent runs with the same serialized configuration and seed produce identical snapshots, design matrix, response vector, parameter vector, artifact payload, and SHA-256 hashes. These checks are not full-run replicates.

B.6 ANALYTIC SCALAR EXPRESSIVITY CHECK

Analytic scalar expressivity gap. Let $D^\star = \mathrm{diag}(0.06, 0.015)$ and initialize an isotropic Gaussian with standard deviation 0.04. Its exact covariance is $\Sigma_t = \Sigma_0 + 2tD^\star$. A scalar model imposes equal directional variance increments, but the truth has x–y gaps of 0.0045 at $t = .05$ and 0.0090 at $t = .10$. The best trace-matched scalar Gaussian still has Hellinger errors 0.160 and 0.195. This representation gap is independent of the estimator and solver.

C NEURAL AND GENERATIVE EXTENSIONS

These independent extensions parameterize structured covariances for neural models. The weak-moment estimator and main experiments do not use them. The following propositions verify their formal properties.

C.1 MODE-WISE DIFFUSION SCHEDULING

For an orthogonal basis $Q = [q_1, \ldots, q_d]$, choose nonnegative mixture weights π_{ik} summing to one, rates $\lambda_{ik} > 0$, and scales $\tau_i > 0$. Define

$$\alpha_i(t) = \sum_{k=1}^{K_i} \pi_{ik} e^{-\lambda_{ik} t/2}, \tag{38}$$

$$\beta_i(t) = -2\frac{\dot\alpha_i(t)}{\alpha_i(t)}, \tag{39}$$

$$A_t = Q\,\mathrm{diag}(\alpha_i(t))Q^\top, \quad \Lambda_t = Q\,\mathrm{diag}(\beta_i(t))Q^\top. \tag{40}$$

For $\Gamma = Q\,\mathrm{diag}(\tau_i^2)Q^\top$, consider

$$dX_t = -\frac{1}{2}\Lambda_t X_t dt + Q\,\mathrm{diag}(\tau_i\sqrt{\beta_i(t)})Q^\top dW_t. \tag{41}$$

This forward process has an explicit transition kernel.

C.2 Covariance-Valued Aggregation

Let continuously differentiable covariance experts satisfy $C_m(t,x) \succeq \epsilon_m I$ with $\epsilon_m > 0$, and let continuously differentiable weights $\rho_m(t,x) \geq 0$, $\sum_m \rho_m = 1$. CVA outputs

$$D_\vartheta(t,x) = \sum_{m=1}^{M} \rho_m(t,x) C_m(t,x). \tag{42}$$

Each expert may be parameterized as $C_m = L_m L_m^\top + \epsilon_m I$.

C.3 Exact scalar collapse of MDS

Proposition C.1 (Mode-wise transition law)**.** *Equation 41 satisfies*

$$X_t \mid X_0 = x_0 \sim \mathcal{N}(A_t x_0, \Gamma - A_t \Gamma A_t^\top). \tag{43}$$

Proof. In the Q coordinates, each component is a scalar linear SDE with fundamental solution $\alpha_i(t)$. Itô isometry and $d(\alpha_i^{-2})/dt = \beta_i \alpha_i^{-2}$ give conditional variance $\tau_i^2(1-\alpha_i^2)$. □

Proposition C.2 (Exact scalar collapse)**.** *For nonnegative mixture weights and a scalar scale $\tau > 0$, MDS equals the scalar exponential process $A_t = e^{-\beta t/2} I$ and $\Gamma - A_t \Gamma A_t^\top = \tau^2(1 - e^{-\beta t})I$ for every $t \geq 0$ if and only if $\tau_i = \tau$ for all i and every positive-weight component has $\lambda_{ik} = \beta$.*

Proof. Sufficiency follows by substitution. Necessity follows from uniqueness of the Laplace transform of a finite positive measure: $\sum_k \pi_{ik} e^{-\lambda_{ik} t/2} = e^{-\beta t/2}$ for all t implies $\sum_k \pi_{ik} \delta_{\lambda_{ik}/2} = \delta_{\beta/2}$. The covariance identity then gives $\tau_i = \tau$. □

MDS strictly contains scalar exponential schedules. It represents a general monotone VDM schedule only if the component family is enlarged or used as an approximation. An attention decay also need not equal the signal-to-noise ratio of a forward process.

For a numerical check of Proposition C.2, set the mixture weights to $(0.3, 0.7)$. With equal rates $\lambda = (2, 2)$, the attenuation $\alpha(t) = \sum_k \pi_k e^{-\lambda_k t/2}$ matches the scalar exponential $e^{-\beta t/2}$ within 1.1×10^{-16} uniformly on $t \in [0, 4]$. With rates $(1, 3)$, the maximum deviation is 0.042, as predicted by the if-and-only-if characterization.

C.4 CVA positive definiteness and square-root nonlinearity

Proposition C.3 (Uniform positive definiteness)**.** *The CVA output obeys*

$$D_\vartheta(t,x) \succeq \left(\sum_m \rho_m(t,x)\epsilon_m\right) I \succ 0. \tag{44}$$

Proof. For every $v \neq 0$, $v^\top D_\vartheta v = \sum_m \rho_m v^\top C_m v \geq (\sum_m \rho_m \epsilon_m)\|v\|^2$. □

CVA preserves positive definiteness without projection or eigenvalue clipping. Reverse sampling requires both the divergence and a square root R of the *aggregated* matrix:

$$RR^\top = 2D_\vartheta, \qquad \left(\sum_m \rho_m C_m\right)^{1/2} \neq \sum_m \rho_m C_m^{1/2} \quad \text{in general.} \tag{45}$$

For state-dependent weights,

$$\operatorname{div} D_\vartheta = \sum_m \left(C_m \nabla \rho_m + \rho_m \operatorname{div} C_m\right). \tag{46}$$

A neural implementation must map the Symmetric Positive Definite(SPD) output to the hidden dimension, for example through $h_D = W_D \operatorname{vech}(\log D_\vartheta)$. A $d \times d$ matrix cannot be added directly to a d_h-vector.

We test Proposition C.3 and the square-root distinction in 200 random trials. Each trial uses $M = 5$ experts $C_m = L_m L_m^\top + 0.05I$ in dimension 4 with Dirichlet weights. Every aggregated covariance has minimum eigenvalue at least $0.35 > 0$. Relative to $(\sum_m \rho_m C_m)^{1/2}$, the Frobenius difference $\|(\sum_m \rho_m C_m)^{1/2} - \sum_m \rho_m C_m^{1/2}\|_F$ averages 13.2% and reaches 20.9%. A sampler must therefore factor the aggregate rather than average the expert factors.

C.5 Reverse-Time Generation

A recovered tensor can also define a reverse-time sampler, although we do not train or evaluate one. Assume the forward diffusion is non-explosive and its density is positive and sufficiently smooth on every interval $[\epsilon, T]$ with $\epsilon > 0$. For Equation 1, define $s^\star(t, x) = \nabla_x \log p_t(x)$. Using the original time variable from T to 0, the reverse SDE is

$$dX_t = \left[b(t, X_t) - 2 \operatorname{div} D(t, X_t) - 2D(t, X_t) s^\star(t, X_t)\right] dt + \sqrt{2D(t, X_t)}\, d\overline{W}_t. \tag{47}$$

Equivalently, for the increasing reverse clock $Y_s = X_{T-s}$,

$$dY_s = \left[-b(T-s, Y_s) + 2 \operatorname{div} D(T-s, Y_s) + 2D(T-s, Y_s) s^\star(T-s, Y_s)\right] ds + \sqrt{2D(T-s, Y_s)}\, d\widetilde{W}_s. \tag{48}$$

For $b = \mu + \operatorname{div} D$, the reverse drift becomes $-\mu + \operatorname{div} D + 2Ds^\star$. The correction vanishes when $\operatorname{div} D = 0$. State independence is sufficient but not necessary.

On a grid $T = t_N > \cdots > t_0 = 0$ with $\Delta_n = t_n - t_{n-1} > 0$, reverse Euler–Maruyama evaluates

$$D_n = D(t_n, x_n), \quad c_n = \operatorname{div} D(t_n, x_n), \quad s_n = s(t_n, x_n), \tag{49}$$

$$R_n R_n^\top = 2D_n, \tag{50}$$

$$x_{n-1} = x_n - \Delta_n b(t_n, x_n) + 2\Delta_n c_n + 2\Delta_n D_n s_n + \sqrt{\Delta_n} R_n \xi_n, \quad \xi_n \sim \mathcal{N}(0, I). \tag{51}$$

Using CVA in this sampler requires $\operatorname{div} D$ and a matrix square root of $2D$. The score must match the marginals induced by that tensor, so a score trained for another scalar schedule is generally invalid.

The associated probability-flow velocity is

$$v(t, x) = b(t, x) - \operatorname{div} D(t, x) - D(t, x) s^\star(t, x). \tag{52}$$

For $b = \mu + \operatorname{div} D$, this simplifies to $v = \mu - Ds^\star$.

D Geometry and Bridge Experiments

This appendix presents two controlled diagnostics. The S^1 study compares intrinsic operator recovery with a Euclidean-naive misspecification. The bridge study compares endpoint-only interpolation with propagation under a separately recovered operator. These experiments do not evaluate complete Riemannian or Schrödinger-bridge generative models.

Geometric consistency and identification. Let D be a smooth, uniformly positive-definite field on the periodic domain, and set $g = D^{-1}$. The Riemannian volume is $dV_g = (\det D)^{-1/2}\, dx$. Expanding the operators gives

$$\nabla \cdot (D \nabla f) = \Delta_g f + \tfrac{1}{2}(D \nabla \log \det D) \cdot \nabla f. \tag{53}$$

Thus $g^{-1} = D$ alone does not equate unweighted Riemannian Brownian motion with the fixed-Fick model under Lebesgue measure. The Fick generator corresponds to a weighted Laplace–Beltrami operator with weight $w = \sqrt{\det D}$. A Lebesgue density p maps to the volume density $p\sqrt{\det D}$. Any consistency check must therefore use the same reference measure, estimate $\widehat{D}$, initial condition, and boundary conditions.

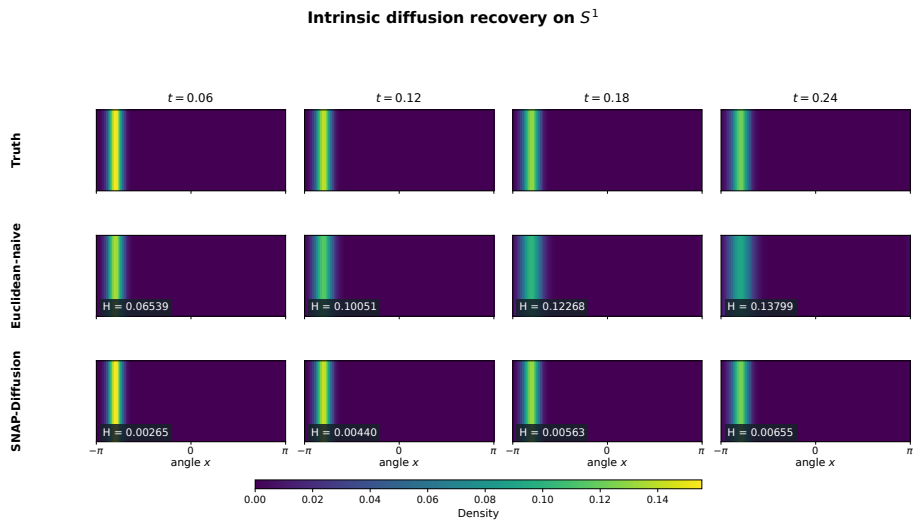


Figure 4: Intrinsic S^1 diagnostic for the predeclared case with seed 11 and unseen release -2.50. Rows show the intrinsic truth, Euclidean-naive fit, and SNAP-Diffusion at $t = .06, .12, .18, .24$ on one linear scale; calibration releases are not shown. Boxes give the selected-case geodesic-distribution Hellinger distance. Table 5 reports the three-seed aggregate.

Table 5: S^1 recovery: three-seed means. Lower is better.

Method	Param. rel.-L^2	Geodesic Hell.	Fourier RMSE
SNAP-Diffusion (ours)	0.05598	0.004179	0.001156
Euclidean-naive	1.1591	0.08967	0.02988
Oracle	0	0	0

Endpoint constraints and path laws. A fixed reference process makes a Schrödinger bridge unique, but the endpoints do not identify the physical generator. We set the reference from training information and exclude evaluation marginals from bridge fitting. Intermediate densities and finite-window hitting probabilities then test behavior not fixed by endpoint agreement. For a time-dependent bridge on $[0, T]$, the relevant quantity is $P_{s,x}(\tau_B < \tau_A, \tau_B \le T)$ at start time s, rather than the time-homogeneous infinite-horizon committor used in the transport experiment. Fitting the bridge to a new source's test endpoint would invalidate the unseen-release evaluation.

Example of an unobserved diffusion direction. Suppose every observed release on the periodic unit square is uniform in x_2, so $p_t(x_1, x_2) = r_t(x_1)$. For a constant field $D = \mathrm{diag}(d_1, d_2)$ with $d_1, d_2 > 0$, the fixed-Fick equation reduces to $\partial_t r = d_1 \partial_{11} r$. The observations are then identical for every d_2, although releases localized in x_2 can evolve differently. Only releases that probe this direction add identifying information. A cross-medium prior may select a plausible d_2, but cannot make the observation map injective.

D.1 Intrinsic S^1 recovery

Represent the circle by $x \in [-\pi, \pi)$ and use the conservative intrinsic generator $\mathrm{div}_{S^1}(D(x)\nabla_{S^1})$, where $D(x) = d_0 + d_c \cos(2x) + d_s \sin(2x)$ and $(d_0, d_c, d_s) = (.055, .018, -.012)$. We observe four calibration releases at $t \in \{0, .06, .12, .18, .24\}$ and reserve four disjoint centers for evaluation. The intrinsic weak fit and a Euclidean-naive unwrapped-variance scalar fit use the same three predeclared seeds. SNAP-Diffusion has lower error at every displayed time and in the three-seed aggregate (Figure 4, Table 5). Because $\mathrm{div}_{S^1}(D\nabla_{S^1})$ has fixed-Fick form with periodic, boundary-free test functions, Proposition 3.6 extends in form to this setting. The observed gap is empirical.

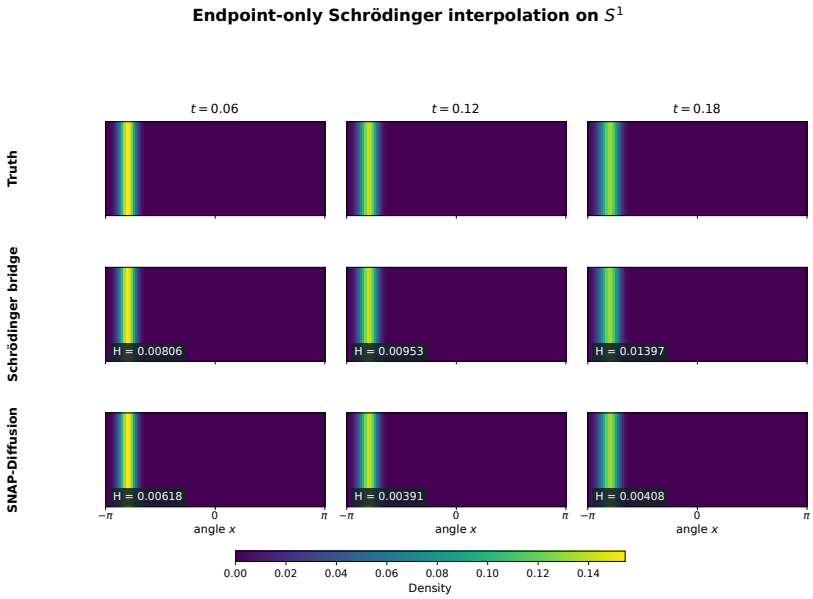


Figure 5: Endpoint-only Schrödinger bridge diagnostic on S^1 for the predeclared case with seed 11 and release 0. Rows show the intrinsic truth, the log-domain Sinkhorn bridge fitted to the two endpoints, and propagation under the SNAP-Diffusion operator at the unobserved times $t = .06, .12, .18$, all on one linear scale. Boxes give selected-case Hellinger distances; Table 6 reports the three-seed aggregate.

Table 6: Endpoint-only S^1 interpolation: three-seed means. Lower is better.

Method	Intermediate Hell.	Fourier RMSE
SNAP-Diffusion (ours), frozen operator	0.008381	0.002861
Schrödinger bridge	0.01253	0.003087
Oracle	0	0

D.2 Endpoint-only Schrödinger interpolation

For each unseen S^1 release, the log-domain Sinkhorn solver receives the $t = 0$ and $T = .24$ histograms and a predeclared constant-diffusion reference kernel with $D_{\mathrm{ref}} = .04$. We withhold truth marginals at $t \in \{.06, .12, .18\}$ until evaluation. All twelve seed–release problems converge in 30–40 iterations, with maximum endpoint residual 8.32×10^{-11}. The resulting bridge satisfies the endpoint constraints. Propagation under the separately recovered SNAP-Diffusion operator has lower mean intermediate error in both the displayed case and the three-seed aggregate (Figure 5, Table 6).

E Transport Experimental Details

E.1 Robust objective

The following robust variant changes the loss while retaining the same weak design. Theorem 3.4 covers the quadratic estimator applied to either declared design in Equations 8 and 11, labeled "SNAP-Diffusion (ours), quadratic loss" in the experiments. The transport study also reports the penalized M-estimator

$$\sum_i 2\big(\sqrt{1 + r_i(\theta)^2} - 1\big) + 10^{-4} \left\| \frac{\theta - \theta_{\mathrm{prior}}}{0.03} \right\|_2^2, \qquad r(\theta) = \frac{A\theta - y}{\mathrm{median}(|y|) + 10^{-12}}, \tag{54}$$

Table 7: Robust-loss transport sensitivity. Aggregation matches Tables 1 and 2. These rows are reported separately because quadratic SNAP-Diffusion is the canonical main-text estimator. Lower is better.

Method	Hellinger	Reg. MAE	Exposure	Rel.-F (D)	L^∞-op (D)	Committor MAE	MFPT rel.-L^1	Regret
SNAP-Diffusion (ours), robust loss	0.004127	0.000329	0.000315	0.01601	0.001080	0.000639	0.006927	0.000020

Table 8: Common-domain robust-loss sensitivity: mean $\pm$ sample SD over nine predeclared truth–seed units. Lower is better.

Method	Hellinger	W_2	Committor RMSE	MFPT rel.-L^2	Regret
SNAP-Diffusion (ours), robust loss	0.1176 ± 0.0575	0.06977 ± 0.03045	0.06903 ± 0.04880	0.1600 ± 0.1926	0.01278 ± 0.03833

with $\theta_{\text{prior}} = (0.03, 0, 0)$ and the same bounds and ellipticity constraints. This is "SNAP-Diffusion (ours), robust loss." It differs from the quadratic instance only in the objective; Theorem 3.4 applies to the quadratic estimator.

Experimental use of the two losses. The periodic and common-domain studies evaluate both objectives, including the robust objective in Equation 54. The FX snapshot study, synthetic-quote interface, and per-medium cross-medium fit use the quadratic objective. This design isolates the effect of the fitting loss while keeping the weak-moment construction fixed; the robust variant is assessed by the reported paired experiments.

E.2 Forward-search comparison and solver validation

Forward calibration estimates θ by propagating each candidate and matching the observed marginals. It uses the same snapshots, parameter family, drift convention, bounds, and priors as SNAP-Diffusion. Thus it tests linear weak-moment inversion against forward search. Robust SNAP-Diffusion has lower density error in Experiment 1 (0.004127 versus 0.008794), while the paired committor interval against forward calibration includes zero in Experiment 2. The transport study does not record a separate common timing comparison.

The scalar baselines restrict the physical coefficient under the same fixed-Fick drift. They isolate directional expressivity in this transport problem; they are not implementations of DDPM, VP/VE-SDE, VDM, or EDM.

Solver and validation. All downstream PDEs use the same symmetric conservative graph generator, built from the x, y, $(1, 1)$, and $(1, -1)$ tensor directions. Matrix-exponential propagation neither clips negative values nor renormalizes mass. The generator also defines the absorbing-boundary systems for the committor and mean first-passage time (MFPT). Before evaluation, we verify symmetry, zero row and column sums, nonnegative off-diagonal entries, negative semidefiniteness, strict dissipation of three checkerboard modes, and positivity on every grid. At resolution 128, relative Fourier-rate errors are 1.07×10^{-3} for positive cross diffusion and 4.13×10^{-4} for negative cross diffusion. The maximum Gaussian covariance-propagation error is 9.56×10^{-7}. Committor and MFPT linear-system residuals are 9.24×10^{-14} and 3.15×10^{-13}.

E.3 Detailed unseen-release results

Experiment 1 results. Figure 6 shows the predeclared example, and Table 1 gives the aggregate. Bracketed t_4 intervals use the five predeclared medium-level paired means and summarize across-medium variation. Relative to robust SNAP-Diffusion, the constant-scalar Hellinger excess is 0.04442 $[0.03510, 0.05374]$; the fixed-basis diagonal excess is 0.02357 $[0.00647, 0.04067]$. Quadratic and robust SNAP-Diffusion are not separated at this resolution. The quadratic instance has Hellinger excess 6.83×10^{-5} $[-1.00 \times 10^{-4}, 2.37 \times 10^{-4}]$ and relative coefficient-field error excess 2.76×10^{-4} $[-4.06 \times 10^{-4}, 9.59 \times 10^{-4}]$. The result supports recovery in the structured tensor class, not superiority of the robust loss.

For robust SNAP-Diffusion, mean Hellinger errors across the low-, middle-, and high-coverage tertiles are 0.00437, 0.00418, and 0.00385. The fixed-basis diagonal model shows the opposite

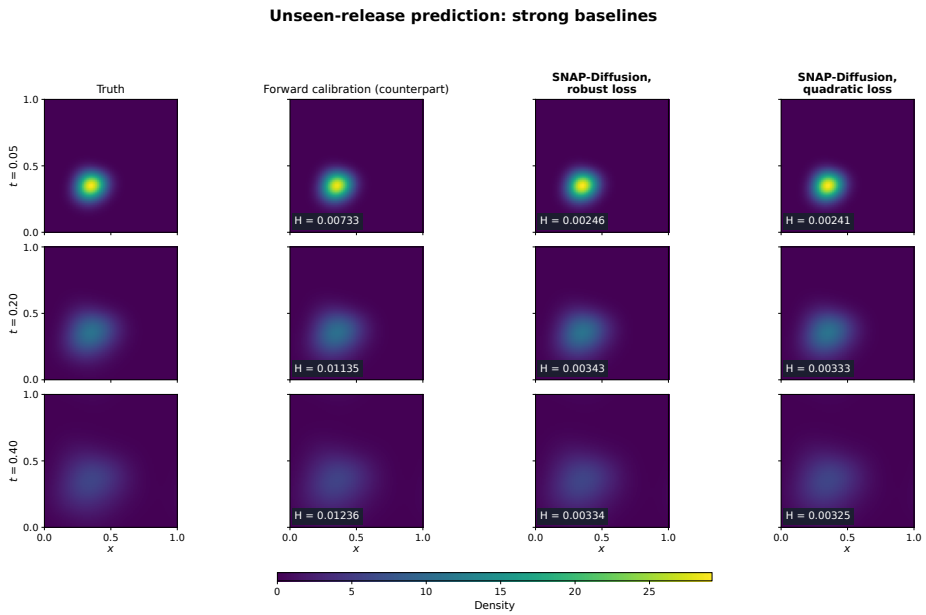


Figure 6: **Experiment 1: robust-loss sensitivity for unseen-release prediction.** Rows show $t = .05, .20, .40$; columns show the truth, forward calibration, robust SNAP-Diffusion, and quadratic SNAP-Diffusion on one linear density scale. This is the first predeclared medium, seed, and test center $(0.35, 0.35)$, without outcome-based selection. The stored 256^2 truth is block-averaged to 128^2. Every method uses its frozen coefficient without refitting. Annotations give selected-case Hellinger errors. Tables 1 and 7 report the aggregate; Figure 8 reports scalar, regional, exposure, and coverage diagnostics.

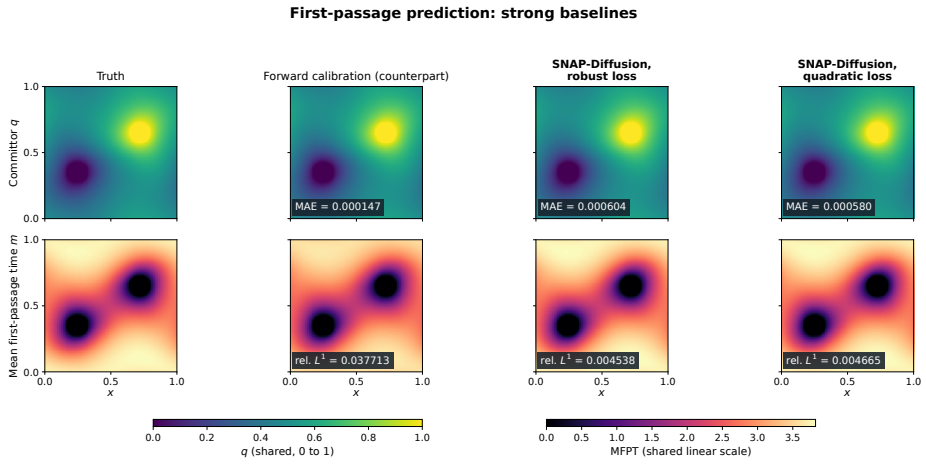


Figure 7: **Experiment 2: robust-loss sensitivity for first-passage prediction.** Columns show the stored discrete truth, forward calibration, robust SNAP-Diffusion, and quadratic SNAP-Diffusion. The top row shows committors on a shared $[0, 1]$ scale; the bottom row shows MFPTs on one linear scale. This is the first predeclared medium, seed, and absorbing layout. Selected-case errors use 243 off-absorber starts with periodic bilinear sampling; grid cells are not independent replicates. Tables 2 and 7 report the aggregate, and Figure 9 gives candidate locations and the censoring-limited SDE check.

pattern, while the oracle is flat. Coverage is therefore not a monotone predictor of error and should be interpreted with the design ablation. All eight test centers were fixed before evaluation.

E.4 Detailed first-passage results and censoring

Detailed first-passage results. Table 2 gives the aggregate; Figures 7 and 9 show the fields, fixed candidates, and SDE check. In the displayed case, robust SNAP-Diffusion, quadratic SNAP-Diffusion, and forward calibration all have zero decision regret. This coarse tie does not imply equal field accuracy.

Relative to robust SNAP-Diffusion, forward calibration has MFPT excess 0.02309 $[0.01698, 0.02921]$. Their mean committor MAEs are not significantly different: 4.80×10^{-4} versus 6.39×10^{-4}, with paired difference -1.58×10^{-4} and interval $[-3.70 \times 10^{-4}, 5.3 \times 10^{-5}]$. Across Tables 1 and 2, robust SNAP-Diffusion has the lowest non-oracle mean in all five Experiment 1 columns and leads or ties on MFPT and regret, while forward calibration has the lower committor mean.

The independent SDE check couples increments at $\Delta t = .002$ and $.001$ and retains all timeouts. The maximum timeout probability ranges from 26.7% to 39.8% across the five media, above the predeclared 10% threshold. This check supports only the censoring-aware finite-window bounds $P(\tau_B \le H) \le q \le P(\tau_B \le H) + P(\tau > H)$ and step-halving consistency. It does not validate infinite-horizon committors or MFPTs.

Numerical checks. Across all full runs, propagation mass error is at most 2.02×10^{-13}, and no negative mass is introduced. The smallest fitted coefficient-field eigenvalue is 0.00505. The largest committor and MFPT residuals are 6.14×10^{-12} and 2.21×10^{-11}. All 400 fitted method records pass the optimizer and ellipticity gates.

E.5 Negative controls and data ablations

Robustness and negative controls. Each predeclared ablation uses one medium and five seeds. The independent unit is the seed-level aggregate, not a grid point. Increasing particles per snapshot from 500 to 20,000 lowers mean held-out Hellinger error from 0.0330 to 0.00716 and mean relative coefficient-field error from 0.132 to 0.0281. At 100,000 particles per release, increasing snapshots from 3 to 7 lowers these means from 0.01145 to 0.00659 and from 0.0423 to 0.0267, although the intermediate metrics are not all monotone. Concentrating training releases in one corner lowers the normalized minimum singular value by 0.2352 $[0.2289, 0.2415]$. It increases Hellinger error by 0.01026 $[0.00293, 0.01760]$ and relative coefficient-field error by 0.05086 $[0.01511, 0.08661]$.

An unrestricted anisotropic fit is harmful when the truth is isotropic. SNAP estimates $\sqrt{\widehat{\theta}_1^2 + \widehat{\theta}_2^2} = 0.00546$ and has Hellinger error 0.02165, compared with 0.00673 for the scalar fit. The paired excess is 0.01492 $[0.00426, 0.02558]$. This amplitude is specific to the three-parameter family in Equation 23; it is not a general function-space norm. Corollary A.1 instead controls excess risk through the isotropic projection gap Δ_{iso} in Equation 30. Thus the larger tensor class can reduce accuracy.

Training and downstream task geometry. The four identification releases are centered at $(.25, .25), (.25, .75), (.75, .25), (.75, .75)$ and are observed independently at $t \in \{0, .05, .10, .15, .20\}$ with 20,000 particles per snapshot. The eight unseen-release centers are $(.35, .35), (.35, .65), (.65, .35), (.65, .65), (.35, .50), (.65, .50), (.50, .35)$, and $(.50, .65)$. Hellinger distance and 4×4 regional-probability MAE are evaluated at $t \in \{.05, .10, .20, .30, .40\}$; cumulative exposure integrates solver outputs spaced by $.005$ over $[0, .4]$ and normalizes by 0.4.

The three absorbing-disk layouts use radii $.08$, $.075$, and $.07$, with centers (A, B) equal to $((.25, .35), (.72, .65))$, $((.18, .72), (.78, .28))$, and $((.30, .18), (.62, .82))$. Committor and MFPT are solved on a 128^2 grid and evaluated on a fixed 16×16 start grid after the predeclared geometric exclusion around each disk pair. The decision task selects among nine fixed candidates on $x = .5$ by predicted committor.

Media, observations, release widths, and grids. The five coefficient vectors in Equation 23 are $(.030, .012, .008)$, $(.032, -.010, .009)$, $(.034, .015, 0)$, $(.030, 0, -.010)$, and $(.036, .010, -.008)$. Each medium uses the ten predeclared seeds $\{11, 23, 37, 53, 71, 89, 107, 131, 151, 173\}$. Identification releases are periodic Gaussians with standard deviation 0.035; the eight Experiment 1 test releases use 0.05. A separate constant-scalar solver check uses an isotropic Gaussian with standard deviation 0.04. This release belongs to neither identification nor evaluation.

We propagate truth snapshots on a 256^2 grid. Recovery uses a 32^2 grid, sixteen weak test functions, trapezoidal time quadrature, and Gaussian histogram smoothing with width 0.15 grid units. Prediction and first-passage evaluation use 128^2 grids. Sections 5.1 and 5.2 give the downstream protocols.

Statistical aggregation. The independent unit is an observation seed nested within a medium. Within each $(\text{medium}, \text{seed}, \text{method})$ unit, we average Hellinger and regional errors over eight releases and five reporting times, exposure error over eight releases, and first-passage metrics over three layouts. Coefficient-field metrics are computed once per fitted field and deduplicated across raw rows. We then average ten seed-level values within each medium and weight the five medium means equally.

Reported intervals are descriptive across-medium t_4 intervals over five predeclared truth-level paired means. They measure variation in paired differences across the full set of media. A nested paired bootstrap with $100{,}000$ draws, resampling media and then seeds within media, gives the same qualitative conclusions. Grid points, releases, times, and layouts are not independent replicates. Unless noted otherwise, $\pm$ denotes sample standard deviation over the named units. The FX sparsity, misspecification, and quote figures instead show seed-level standard errors, as stated in their captions.

Numerical validation and censoring. The graph generator is symmetric, conservative, Metzler, and negative semidefinite. It also strictly dissipates three checkerboard modes. Propagation uses a matrix exponential without positivity clipping or mass renormalization. The oracle's nonzero density and exposure errors in Table 1 arise from comparing block-averaged 256^2 truth with 128^2 oracle propagation, not from coefficient error. The SDE cross-check uses $2{,}000$ particles per start, horizon $H = 3$, and paired Brownian increments at $\Delta t = .002$ and $.001$. Timeouts remain censored, so the pass conditions cover only finite-window hitting bounds and step-halving consistency.

Complementary diagnostics. The following panels use the locked run to add the spatial-scalar comparison and functional diagnostics to the main quadratic-SNAP-Diffusion and forward-calibration figures. They add no fits or replicates.

Rows show $t = .15, .30, .60$ for the predeclared release test_south on one linear scale. Fields use operators averaged over the nine truth–seed units, the nn-APPEX field is the density of its moment-matched Gaussian summary. Boxes give the corresponding nine-unit mean and 95% CI of Gaussian Hellinger distance. Table 3 reports the aggregate. This $\mathbb{R}^2$ benchmark differs from Figure 6.

F Common-Domain and Cross-Medium Experiments

F.1 Common-domain implementation and full paired results

Common comparison domain. The principal APPEX Guan et al. (2024) model has the linear form $dX_t = AX_t\,dt + G\,dW_t$ with constant covariance $GG^\top$; nn-APPEX Guan et al. (2025) instead learns a nonlinear conservative drift and scalar covariance rate $\sigma^2 I$. These baselines and Ma/FPP Ma et al. (2021) do not estimate the periodic spatial tensor in Equation 23. We therefore compare them separately from Section 5 and Tables 1–2. The predeclared shared domain is $\mathbb{R}^2$ with zero drift and three constant diagonal diffusion truths. Each method receives independent point snapshots from two releases at $t \in \{0, .06, .12\}$, with 64 samples per release and time. All fits are completed before the two downstream tests. nn-APPEX is correctly specified for the isotropic truth, but diffusion-class misspecified for both anisotropic truths.

The APPEX implementation retains the authors' function logic; only automatic module-level execution is disabled so it can accept the shared snapshots. This is not a reproduction of the APPEX OU experiment. Ma/FPP uses its native nonperiodic critic and generator at the predeclared final iteration,

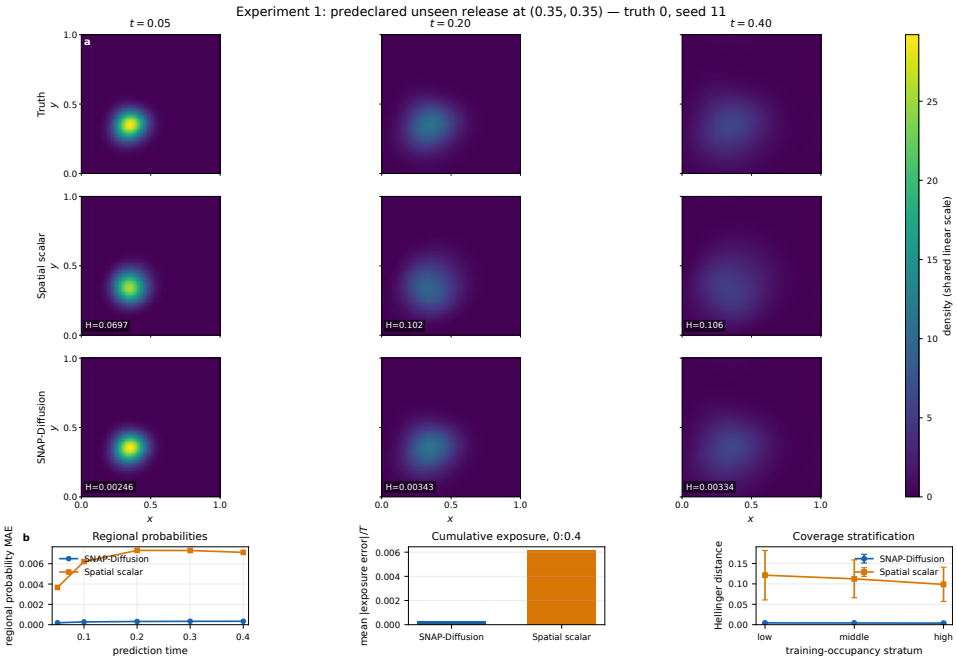


Figure 8: **Experiment 1: complementary scalar and functional diagnostics.** For the first predeclared medium, seed, and test center $(0.35, 0.35)$, the top grid compares truth, SNAP-Diffusion, and spatial-scalar densities at $t = .05, .20, .40$ on one linear scale. The lower panels show regional-probability error, cumulative exposure error over $[0, .4]$, and training-occupancy strata. Aggregate conclusions use five media and ten seeds per medium; releases and times are not independent replicates. Figure 6 gives the primary baseline comparison.

without checkpoint selection from truth or downstream outcomes. All nine truth–seed fits finish. The largest finite-difference backward-PDE residual is 1.30×10^{-11}.

At zero drift, the fixed-Itô and fixed-Fick weak designs coincide. Because the domain is $\mathbb{R}^2$ rather than a torus, Proposition 3.6 does not apply directly. The reported density and first-arrival results are empirical.

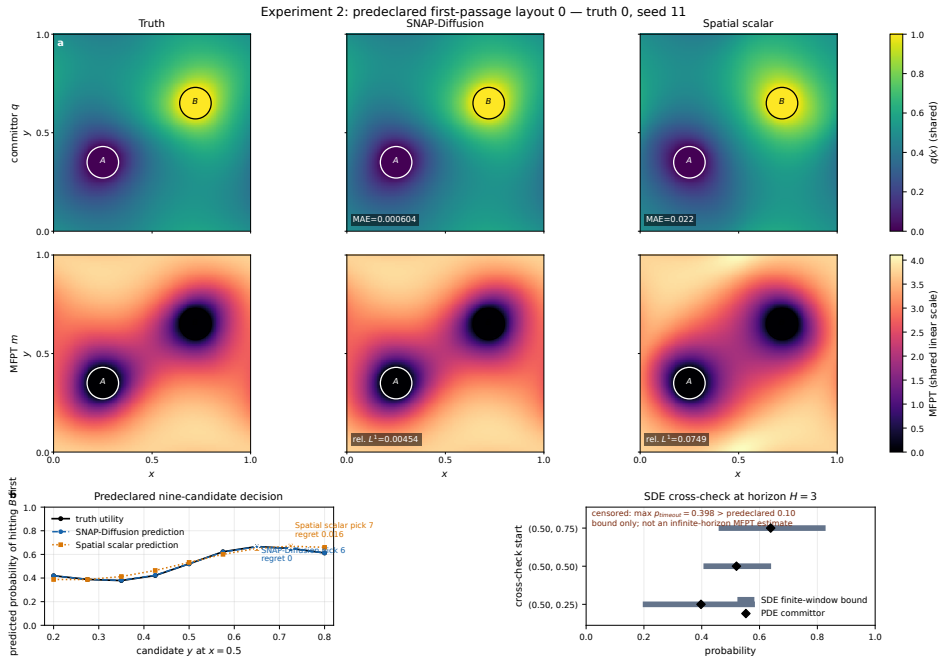


Figure 9: **Experiment 2: complementary decision and SDE diagnostics.** For the first predeclared medium, seed, and absorbing layout, the panels compare truth, SNAP-Diffusion, and spatial-scalar committors on a shared $[0, 1]$ scale and MFPTs on one linear scale. The lower panels show nine fixed candidates and SDE finite-window bounds.

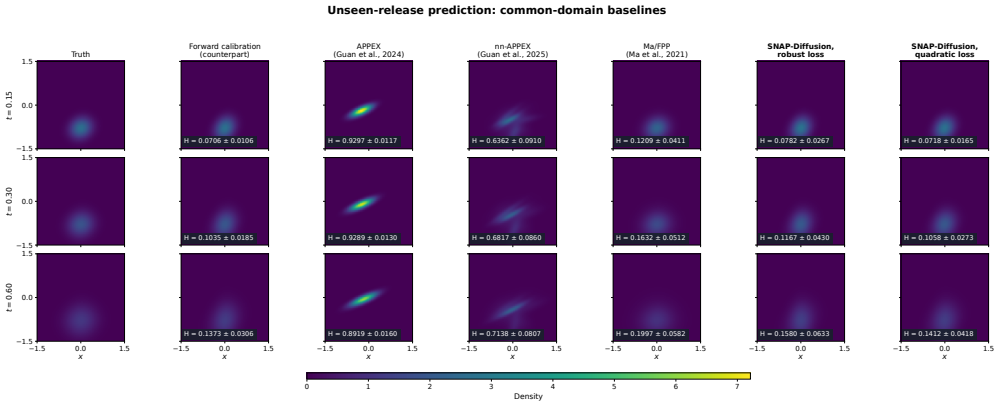


Figure 10: **Common-domain Experiment: unseen-release prediction.** Columns show the exact Gaussian truth and estimates from APPEX Guan et al. (2024), nn-APPEX Guan et al. (2025), Ma/FPP Ma et al. (2021), forward calibration, robust SNAP-Diffusion, and quadratic SNAP-Diffusion.

Table 9: Common-domain paired differences (SNAP-Diffusion minus opponent): means and two-sided 95% Student-t intervals over nine truth–seed units. Negative values favor SNAP; $\star$ marks intervals that exclude zero. Robust SNAP-Diffusion has lower error than APPEX on both Experiment 1 metrics, committor, and MFPT, and lower error than Ma/FPP on MFPT, with the corresponding intervals excluding zero. Its intervals against quadratic SNAP-Diffusion and forward calibration include zero on every metric. Regret is an exact tie.

Experiment	Opponent	Metric	SNAP-Diffusion − opponent
Unseen release	APPEX Guan et al. (2024)	Hellinger	$-0.7992\ [-0.8399, -0.7586]\ \star$
Unseen release	APPEX Guan et al. (2024)	W_2	$-0.7532\ [-0.7711, -0.7354]\ \star$
Unseen release	nn-APPEX Guan et al. (2025)	Hellinger	$-0.5596\ [-0.6481, -0.4711]\ \star$
Unseen release	nn-APPEX Guan et al. (2025)	W_2	$-0.5220\ [-0.7063, -0.3377]\ \star$
Unseen release	Ma/FPP Ma et al. (2021)	Hellinger	$-0.0436\ [-0.1059, +0.0186]$
Unseen release	Ma/FPP Ma et al. (2021)	W_2	$-0.0433\ [-0.0868, +0.0003]$
Unseen release	SNAP-Diffusion quadratic loss	Hellinger	$+0.0114\ [-0.0106, +0.0333]$
Unseen release	SNAP-Diffusion quadratic loss	W_2	$+0.0055\ [-0.0061, +0.0171]$
Unseen release	Forward calibration	Hellinger	$+0.0138\ [-0.0180, +0.0456]$
Unseen release	Forward calibration	W_2	$+0.0066\ [-0.0103, +0.0234]$
First arrival	APPEX Guan et al. (2024)	Committor RMSE	$-0.0713\ [-0.0976, -0.0451]\ \star$
First arrival	APPEX Guan et al. (2024)	MFPT rel.-L^2	$-2.7065\ [-3.6026, -1.8105]\ \star$
First arrival	nn-APPEX Guan et al. (2025)	Committor RMSE	$-0.3130\ [-0.3926, -0.2335]\ \star$
First arrival	nn-APPEX Guan et al. (2025)	MFPT rel.-L^2	$-232.5864\ [-530.3465, +65.1738]$
First arrival	nn-APPEX Guan et al. (2025)	Regret	$0.0000\ [-0.0442, +0.0442]$
First arrival	Ma/FPP Ma et al. (2021)	MFPT rel.-L^2	$-0.3675\ [-0.5457, -0.1893]\ \star$
First arrival	Ma/FPP Ma et al. (2021)	Committor RMSE	$+0.0311\ [-0.0213, +0.0835]$
First arrival	SNAP-Diffusion quadratic loss	Committor RMSE	$+0.0095\ [-0.0074, +0.0263]$
First arrival	SNAP-Diffusion quadratic loss	MFPT rel.-L^2	$+0.0589\ [-0.0436, +0.1614]$
First arrival	Forward calibration	Committor RMSE	$+0.0117\ [-0.0123, +0.0358]$
First arrival	Forward calibration	MFPT rel.-L^2	$+0.0786\ [-0.0493, +0.2065]$
First arrival	SNAP-Diffusion quadratic loss / Forward calibration	Regret	0 (exact tie)

Common-domain Experiment. Every paired 95% interval among robust SNAP-Diffusion, quadratic SNAP-Diffusion, and forward calibration includes zero on both unseen-propagation metrics. The corresponding Ma/FPP Ma et al. (2021) intervals also include zero. Robust SNAP-Diffusion has lower error than APPEX Guan et al. (2024) on both metrics, with paired intervals excluding zero. This comparison concerns the shared zero-drift $\mathbb{R}^2$ problem rather than APPEX's native OU setting. Theorem 3.4 covers quadratic SNAP; the robust result is empirical.

Why nn-APPEX underperforms APPEX in first-passage prediction? In Figure 2, After numerical stabilization, five of the nine nn-APPEX runs still have MFPT relative-L^2 error above 100. Even on the favorable retained subset ($n = 4$), nn-APPEX has committor RMSE 0.4377 ± 0.0527, MFPT relative-L^2 error 4.0545 ± 2.5340, and decision regret 0.0671 ± 0.1257, compared with 0.1404 ± 0.0154, 2.8666 ± 1.0181, and 0.0510 ± 0.0588 for APPEX over all nine units;

And in table 9, Scalar-diffusion misspecification contributes for the two anisotropic truths but cannot fully explain the failure because the isotropic runs are also unstable. For completeness, the original 57×57 upwind paired-difference table (SNAP-Diffusion minus nn-APPEX) reports $-0.3130\ [-0.3926, -0.2335]$ for committor RMSE, $-232.5864\ [-530.3465, +65.1738]$ for MFPT relative-L^2, and $0.0000\ [-0.0442, +0.0442]$ for regret.

These are descriptive post-selection means with 95% Student-t half-widths. Grid refinement, M-matrix checks, and scaled backward errors rule out PDE-solver divergence. Under this sparse design, the official nn-APPEX loop fits an approximately 1.7×10^4-parameter neural potential given only 128 observed snapshot samples per time level, reinitializes the network at each outer iteration, and provides no explicit regularization, validation-based selection, or early stopping. These factors are consistent with the observed drift RMS values of roughly 7–15 despite zero true drift and with the resulting spurious potential wells, which moderately distort committors but catastrophically inflate passage times. Scalar-diffusion misspecification contributes for the two anisotropic truths but cannot fully explain the failure because the isotropic runs are also unstable.

Method labels. Quadratic and robust SNAP-Diffusion use the same declared weak design from Equation 8 or 11; robust SNAP-Diffusion uses the objective in Equation 54. In the FX figures, "SNAP (ours)" denotes the quadratic instance specified in Appendix G. Its MMD baseline in Table 4 is the FX counterpart of forward calibration.

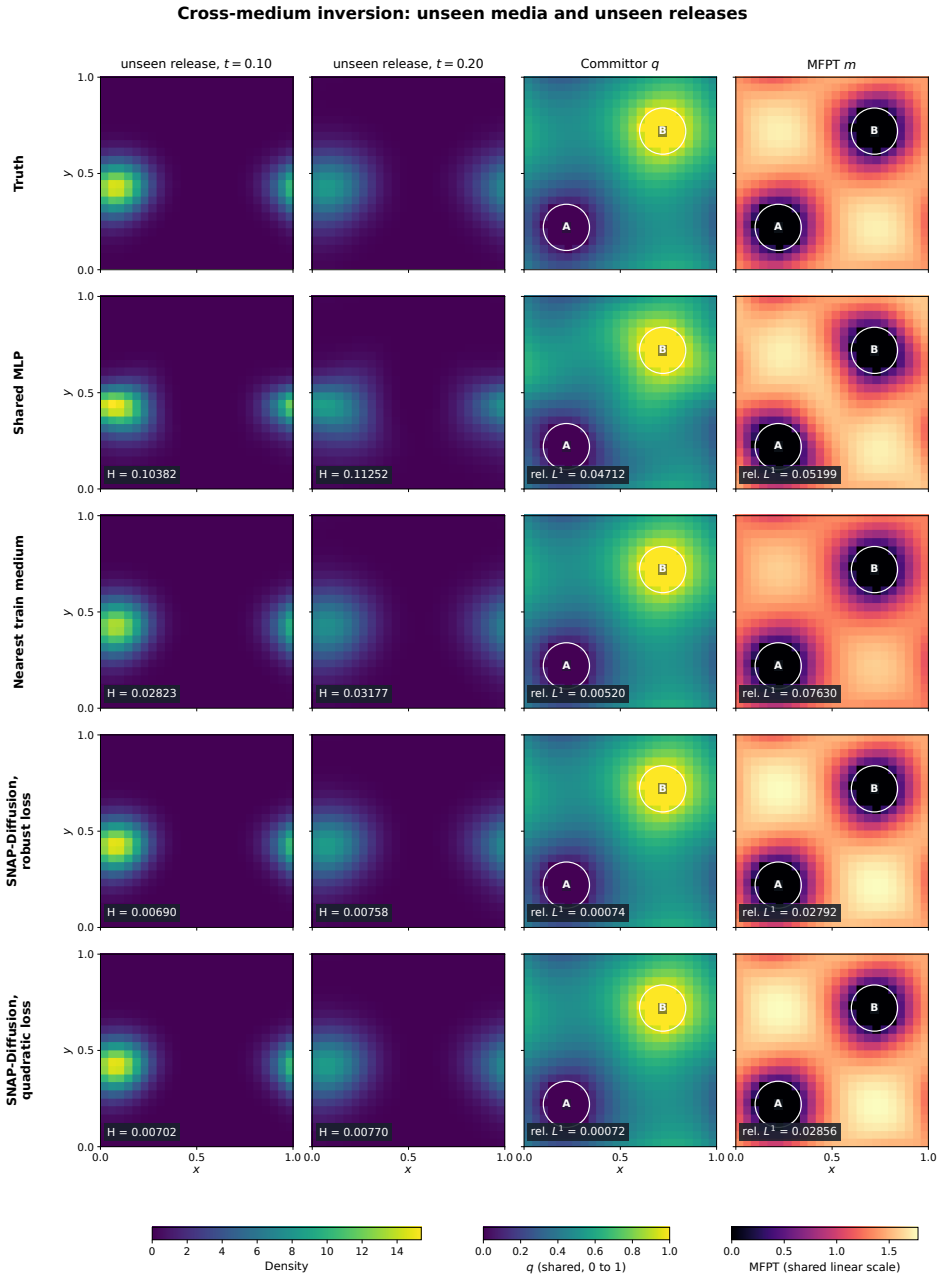


Figure 11: Cross-medium inversion for a held-out medium and release (test medium 0, seed 11, query release $(0.08, 0.42)$). Rows show the truth, shared MLP, nearest-training-medium baseline, post-hoc robust SNAP-Diffusion, and per-medium quadratic SNAP-Diffusion. Columns show unseen-release densities at $t = .10, .20$ on one linear scale, followed by committor and MFPT fields on shared scales; absorbing disks A and B are marked. Boxes give selected-case errors against the stored truth. Table 10 reports the three-seed aggregate for the predeclared rows. It includes quadratic SNAP-Diffusion but excludes the post-hoc robust row. The shared MLP does not generalize successfully in this pilot.

Table 10: Cross-medium pilot: mean over three seed-level aggregates. Lower is better; oracle is zero for every metric. Bold marks the best non-oracle mean.

Method	Param. rel.-L^2	Release Hell.	Committor rel.-L^1	MFPT rel.-L^1
Shared amortized MLP	0.6990	0.1025	0.03758	**0.1396**
Nearest train medium	0.2483	**0.04369**	0.01201	0.1630
SNAP-Diffusion (ours), quadratic loss	**0.2327**	0.04750	**0.009240**	0.2268

F.2 Cross-medium amortization pilot

Cross-medium amortization pilot. For each seed, we train one two-layer MLP across twelve media. We select its ridge parameter on three validation media and evaluate on three other media excluded from training and selection. Inputs are release-aggregated weak sufficient statistics. Test support releases and four query releases do not overlap the training centers. Figure 11 and Table 10 summarize three predeclared seed-level aggregates. The shared MLP has larger parameter, unseen-release Hellinger, and committor errors than both per-medium quadratic SNAP-Diffusion and the nearest-training-medium baseline. It has smaller mean MFPT error.

G FX Details and Additional Experiments

G.1 Observation models and empirical scale

Observation model. The main FX experiment uses simulated joint snapshots from Equation 24. It applies the constrained quadratic weak-moment estimator from Section 3.3, which is covered by Theorem 3.4. It does not use the robust objective in Equation 54.

The synthetic-quote audit uses a different observation model. An option price is $C(K,T) = B_T^{-1}\int (s-K)^+ p_T(s)ds$, and a cross-pair price depends on the joint law. Calibration then acts through the nonlinear map $F_m(\theta) = e^{-\int r}\int h_m(x,y)p^{\theta}_{T_m}(x,y)dxdy$ and requires Jacobian conditions rather than the linear design condition in Theorem 3.4. Table 11 evaluates only this synthetic interface. The available data contain no EUR/GBP option chain, while listed FX-ETF options are American and cover few expiries. We therefore make no market-calibration claim. Because the FX experiment fixes the Itô drift, the fixed-Fick stability result does not apply directly. Its pricing results are empirical simulations.

Empirical scale check. Daily EUR/USD and GBP/USD spot histories from 2003–2026 ($\sim$5,900 days) yield annualized realized volatilities $\hat\sigma_X = 10.9\%$ and $\hat\sigma_Y = 9.3\%$, close to the simulated $10\%/12\%$. Listed FX-ETF ATM implied volatilities are 8.8% for EUR and 12.0% for GBP. The 63-day rolling realized correlation has range $[0.12, 0.92]$ and mean 0.65 (Figure 12). This variation supports a time-dependent correlation model but neither establishes state dependence nor validates the risk-neutral simulator.

Because $Z = X/Y$, $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$. With σ_X, σ_Y fixed, a 1-year at-the-money EUR/GBP call is worth 0.042, 0.033, and 0.020 of notional at $\rho = 0.25, 0.55, 0.85$, respectively. Thus the held-out prices remain sensitive to correlation.

G.2 Synthetic-quote audit

Synthetic-quote audit. A separate three-seed experiment replaces snapshots with quote surfaces from the same simulator, then evaluates held-out contracts. It compares the quadratic SNAP-Diffusion quote interface with Black–Scholes flat correlation (Dupire, 1994), linear local correlation, sticky local volatility, and reduced approximations of Guyon (2013), Pallavicini (2025), and Kay (2025). The "-style" methods are independent reductions, not author code or complete calibration systems. All methods use common random numbers, and every result is retained. The finite-path oracle is omitted because its error is zero by construction.

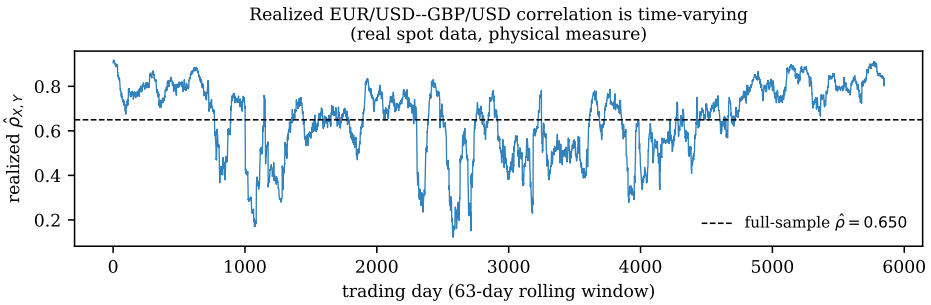


Figure 12: Sixty-three-day rolling EUR/USD–GBP/USD realized correlation from 2003–2026 under the physical measure. The range is $[0.12, 0.92]$, and the full-sample mean is 0.65. The variation motivates a time-dependent model but does not establish state dependence or risk-neutral local correlation.

Table 11: Controlled synthetic-quote MAE, $\times 10^3$. Values are averaged within seed and then across three seeds. Lower is better; bold marks the smallest observed mean in each column. Kay-style has the lowest mean under constant and state-dependent correlation, Pallavicini-style under time-varying correlation, and SNAP-Diffusion (ours) ranks in the middle for all three truths. "-style" denotes a reduced approximation rather than an author-code reproduction. All quotes are simulated.

Method	Constant	Time-varying	State-dependent
Linear-in-time ρ (control)	0.4844	1.5782	0.8774
Moneyness-interpolated ρ (control)	0.2078	1.9839	0.8752
BS flat Dupire (1994)	0.1998	1.9809	0.9586
Guyon-style Guyon (2013)	0.6100	2.0210	1.3477
Pallavicini-style Pallavicini (2025)	0.4568	**1.5354**	0.7779
Kay-style Kay (2025)	**0.01866**	1.8727	**0.7270**
SNAP-Diffusion (ours), quad.	0.4826	1.5825	0.9140

Synthetic-quote results. Table 11 show truth-dependent rankings. Under constant correlation, the Kay-style midpoint approximation has the lowest MAE: 0.0187 versus 0.19–0.61 for the field-based methods. Under time-varying correlation, Pallavicini-style has the lowest mean (1.5354); SNAP-Diffusion, linear local correlation, and sticky local volatility range from 1.58 to 1.98. Under state-dependent correlation, Kay-style again has the lowest mean (0.7270). SNAP-Diffusion ranks in the middle at 0.9140, below Black–Scholes and Guyon-style. In this three-seed audit, the reduced financial approximations therefore match or outperform the SNAP-Diffusion quote interface. This result is separate from the snapshot-observation study.

G.3 Misspecification, numerical sensitivity, and source audit

Misspecification study. This separate three-seed study fixes the fitted tanh basis and uses three truths outside it: non-tanh local structure, nonlinear time dependence, and a state-interaction/high-frequency field. The comparisons are quadratic SNAP-Diffusion, feasible two-step GMM, an increment-based effective-correlation estimator, a constant weak fit, and a time-only weak fit. Figure 13 and Table 12 show different field and price rankings. For the non-tanh local truth, constant weak has the smallest field error, while feasible GMM has the smallest quote MAE. Under nonlinear time dependence, effective ρ has the smallest error on both metrics. Feasible GMM has the smallest price MAE for two of the three truths.

The source audit found neither the earlier FX fitting source nor an FX-specific license sufficient for a reproduction claim. We therefore describe these FX baselines as independent or reduced implementations.

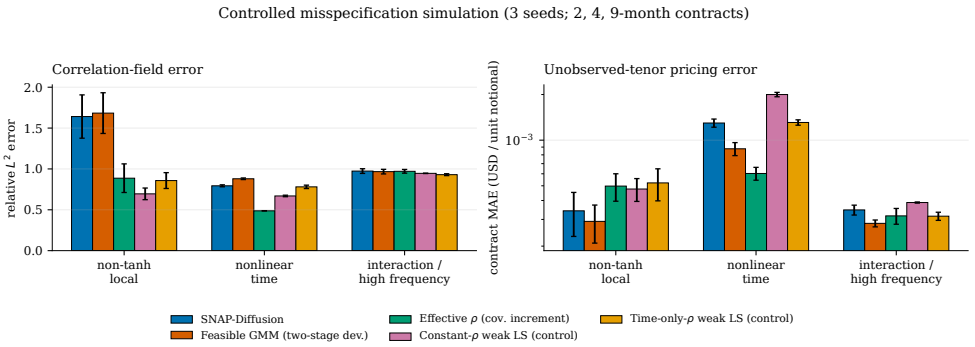


Figure 13: Controlled FX misspecification study over three seeds. Methods are quadratic SNAP-Diffusion (blue), a two-stage feasible-GMM development variant, effective-ρ covariance increments, and constant-ρ and time-only-ρ weak-LS controls. The left panel reports relative correlation-field L^2 error. The right reports MAE over 21 held-out synthetic contracts per seed on a logarithmic axis. Error bars are seed-level standard errors.

Table 12: Additional FX misspecification study. Entries are three-seed means; pricing MAE is in 10^{-3} USD per unit notional. Contracts are aggregated within seed. Lower is better; bold marks the smallest observed mean in each row.

Metric	SNAP-Diffusion	Feasible GMM	Effective ρ	Constant weak	Time-only weak
Field rel.-L^2	1.641	1.683	0.8868	**0.6951**	0.8577
Price MAE $\times 10^3$	0.3419	**0.2913**	0.4979	0.4761	0.5225
Field rel.-L^2	0.7936	0.8797	**0.4877**	0.6694	0.7807
Price MAE $\times 10^3$	1.298	0.8772	**0.6029**	2.000	1.308
Field rel.-L^2	0.9739	0.9670	0.9709	0.9460	**0.9299**
Price MAE $\times 10^3$	0.3467	**0.2827**	0.3166	0.3878	0.3152

Additional snapshot evidence. The following sections report three-seed sparsity scaling and a realized-correlation scale check. The main text reports the five-seed snapshot pricing study in Table 4 and Figure 3. Open problems include calibration from observed market prices, bid–ask-weighted Jacobian identifiability, and task-level error bounds.

G.4 Snapshot sparsity and sample-size scaling

We vary the number of unpaired samples per observed snapshot over $N \in \{500, 2000, 5000\}$ and compare archived panels with two, four, or six observed intervals. Here "observed" means supplied to operator recovery. The main four-interval protocol observes $\{0, 1, 3, 6, 12\}$ months and evaluates contracts at the unobserved $\{2, 4, 9\}$-month tenors. The archive does not retain the exact snapshot times for the two- and six-interval panels. Their pricing aggregates are therefore identified only by the evaluated tenor set.

At every archived sparsity level, recovery and pricing errors decrease monotonically with N. At matched N, the four-interval panel improves on the two-interval panel, providing empirical evidence that additional occupied-time information strengthens recovery in this design. For the state-dependent truth with two intervals, relative ρ errors are 0.43 ± 0.09, 0.29 ± 0.16, and 0.16 ± 0.06 as N increases from 500 to 2000 to 5000; pricing errors are 1.13, 0.45, and 0.24 ($\times 10^{-3}$). With four intervals, the corresponding sequences are 0.28 ± 0.14, 0.21 ± 0.08, and 0.09 ± 0.03 for recovery, and 0.42, 0.31, and 0.20 for pricing. In the four-interval panel, the constant-ρ baseline does not improve with N (0.81, 0.86, 0.77 $\times 10^{-3}$), whereas SNAP-Diffusion does. This contrast is consistent with structural misspecification rather than sample size alone. Figure 15 shows both trends.

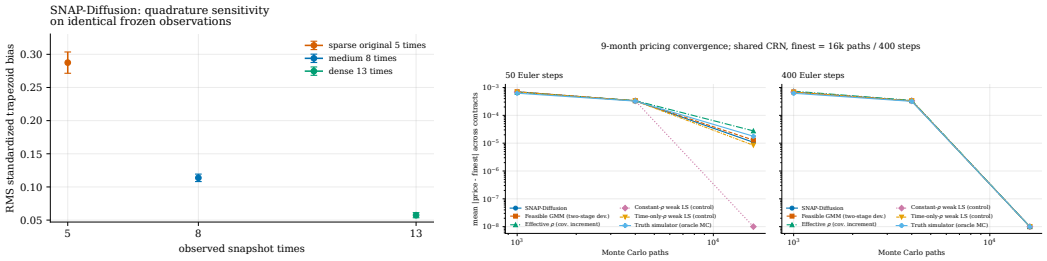


Figure 14: FX numerical sensitivity with fixed observations and common random numbers. Left: denser occupation quadrature reduces standardized trapezoid-bias RMS from 0.2875 at five snapshot times to 0.1137 at eight and 0.0574 at thirteen. Downstream prices do not select the setting. Right: Euler-step and path-count convergence for 9-month contracts, measured against each method's own 16,000-path, 400-step value. The truth simulator provides the oracle pricing reference. Common random numbers reduce paired variance but not Monte-Carlo error or Euler bias. Zero error at the finest point is definitional.

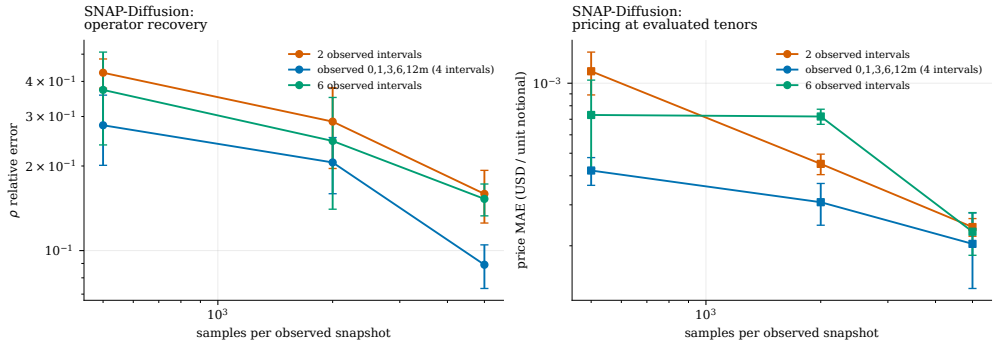


Figure 15: FX snapshot sparsity and sample-size ablation for the state-dependent truth over three seeds. The left panel shows relative ρ-recovery error; the right shows pricing MAE for 2m, 4m, and 9m contracts. The main four-interval panel observes $\{0, 1, 3, 6, 12\}$ months, so all three pricing tenors are held out. Exact snapshot times are unavailable for the archived two- and six-interval panels. Color encodes interval count; circles, squares, and triangles encode 500, 2000, and 5000 samples per snapshot.

G.5 Realized-correlation scale check

We fit the rolling realized correlation in Figure 12 with $\rho(t) \approx 0.65 + 0.04(2t - 1)$ and use this affine profile as the truth in a three-seed controlled experiment. Its residual RMS is 0.16, so the observed profile is far from affine. SNAP-Diffusion recovers $\theta \approx (0.639, 0.012, -0.017, 0.015)$, with relative ρ error 0.085 and held-out pricing error 0.22×10^{-3}. The constant, time-dependent, forward-calibration (MMD), and marginal-interpolation baselines have pricing errors 0.43, 0.21, 0.26, and 0.07 ($\times10^{-3}$), respectively. The fitted affine profile is nearly constant, which explains why interpolation is competitive. This agrees with the snapshot-study result that a more flexible field need not help when the truth has little structure. The rolling statistic is a physical-measure proxy for correlation scale and temporal variation, not a risk-neutral local correlation.

G.6 Measured FX fitting time

On our machine, mean per-fit wall-clock times are 0.090 ± 0.133 s for SNAP and 24.31 ± 1.93 s for simulation-in-the-loop forward calibration. These are sample standard deviations over 25 fits. The measured ratio is $270\times$; we retain the relatively large spread of the faster fit.

G.7 FX PROTOCOLS AND REPRODUCIBILITY

We use log coordinates (u, v), fixed volatilities $\sigma_X = 0.10$ and $\sigma_Y = 0.12$, zero rates, state scale $c = 0.15$, parameter budget $\sum_k |\theta_k| \le 0.9$, and $T_{\max} = 1$y. The main study observes snapshots at $\{0, 1, 3, 6, 12\}$m with $5{,}000$ unpaired samples. The sparsity study uses $N \in \{500, 2000, 5000\}$. Snapshot simulation applies Euler–Maruyama with 8 substeps per interval.

The FX SNAP-Diffusion estimator uses nine Gaussian bumps with $s = 0.12$, centered on a 3×3 grid over $[-0.15, 0.15]^2$, and the global polynomial tests uv, u^2v, uv^2, and u^2v^2. SLSQP minimizes the quadratic weak-moment loss over the ℓ_1 ball. Cross-derivative moments from the localized Gaussian probes supply the main information about state dependence. The polynomial tests stabilize low-order aggregate directions but are insufficient alone. The forward-calibration MMD baseline uses 256 random Fourier features, Gaussian kernel length 0.15, $8{,}000$ common-random-number paths, and SLSQP.

Pricing uses Euler–Maruyama with 200 substeps and 2×10^5 paths per tenor, with common random numbers across methods. EUR/GBP strikes are $0.94, 1.00, 1.06, 1.12\times$ forward; worst-of strikes are $0.90, 0.95, 1.00$. The main evaluation holds out all $\{2, 4, 9\}$m tenors. As noted above, this condition does not apply to the dense sparsity panel. Main-study seeds are $\{11, 23, 37, 53, 71\}$; the sparsity and realism checks use $\{11, 23, 37\}$. The empirical scale check uses daily spot data from 2003-12–2026-09 obtained through a public API, a 63-day rolling window, and one read of FXE/FXB option chains for implied volatility.

G.8 ROLES OF THE FX FIGURES

Under fixed protocols, Figure 13 evaluates SNAP-Diffusion, its feasible two-step GMM development variant, effective-ρ covariance increments, and constant- and time-only-ρ weak fits under three misspecified truths. Three figures isolate numerical or data sensitivity. Figure 15 varies the sample and interval counts. Figure 14 varies occupation quadrature in its left panel and pricing time steps and paths in its right panel. Figure 16 reports product-level held-out pricing error for SNAP-Diffusion, a development variant, and controls in a three-seed pilot.

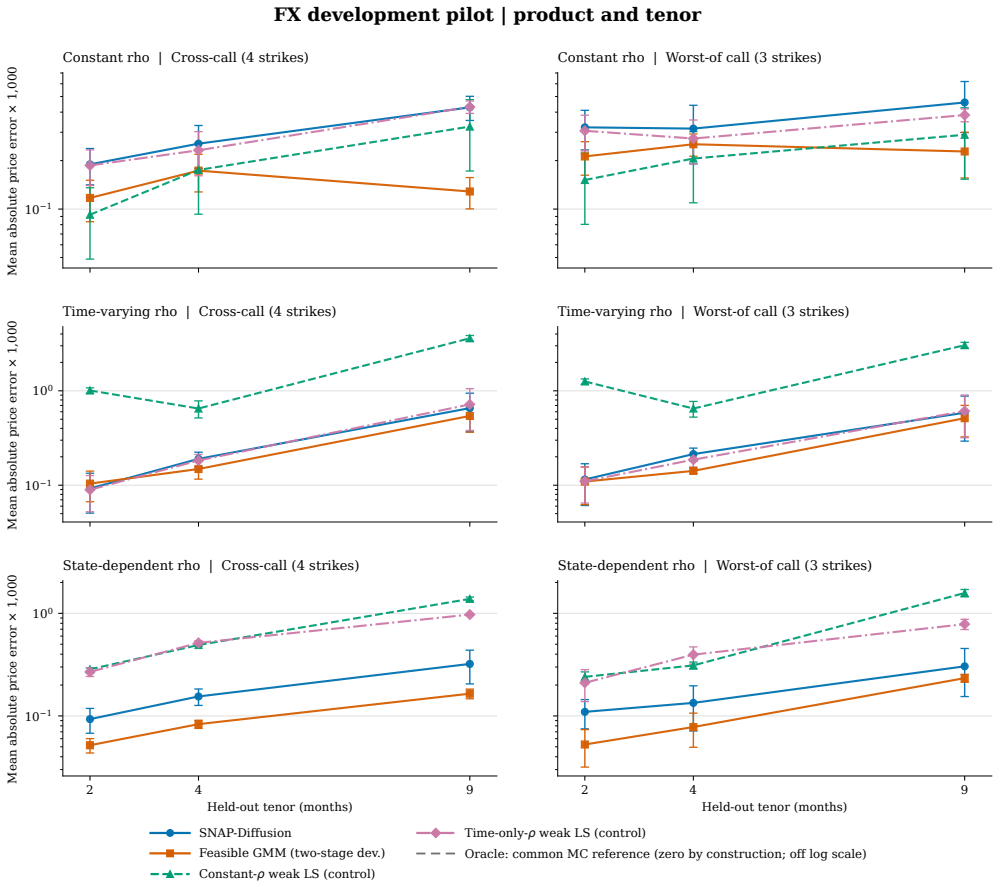


Figure 16: FX development pilot: held-out-tenor pricing error by product for SNAP-Diffusion, the two-stage feasible-GMM variant, and constant-ρ and time-only-ρ weak-LS controls. The dashed line is the shared Monte-Carlo oracle reference and equals zero by construction. Points are seed-level means over equally weighted strikes; whiskers show one standard error over seeds 11/23/37. Feasible GMM has the lowest mean under constant and state-dependent ρ; the overlapping SNAP-Diffusion curves have the lowest mean under time-varying ρ.