TMB versus pypomp Inference for Nonlinear State-Space Fisheries Models

Abstract

State-space fisheries models contain bilinear products of latent variables that render the Laplace approximation (used by TMB) inexact. We compare TMB and **pypomp** (particle filtering with IF2 on GPU) on two state-space recruitment models of increasing nonlinearity. **M1** (linear-Gaussian Gompertz with latent environmental covariate, $T = 200$, simulated) validates both engines: TMB matches the Kalman filter to machine precision, and IF2 recovers the MLE within particle-filter Monte Carlo noise. **M2** (time-varying density dependence on monthly recruitment/SOI, $T = 453$) introduces a bilinear $b_t \cdot X_{t-1}$ term through a bounded random walk for b_t. A 2×2 cross-evaluation shows that the Laplace log-likelihood *exceeds* the bootstrap-particle-filter estimate by 5–10 units at matched parameters, a systematic positive bias consistent with Laplace's optimism on nonlinear models. We also show that particle-filter effective sample size identifies specific time points where the model's predictive distribution is too narrow—a diagnostic the Laplace approximation cannot provide.

1 Introduction

State-space (SSM) representations are a standard framework in fisheries assessment (Valpine and Hastings 2002; Dennis et al. 2006). Two inference paradigms dominate:

- **TMB** (Kristensen et al. 2016) integrates out latent states by the Laplace approximation, evaluated via sparse automatic differentiation. TMB is the computational backbone of many operational stock-assessment packages and is exact for linear-Gaussian SSMs.
- **POMP** (King, Nguyen, and Ionides 2016) evaluates the likelihood via sequential Monte Carlo (particle filtering), and maximises it via iterated filtering (**IF2**, Ionides et al. (2015)). The POMP approach is plug-and-play: it requires only simulators for the state process and observation densities, so arbitrary nonlinearities and non-Gaussian features are handled without bespoke derivations.

Laplace approximation and particle filtering produce the same answer on linear-Gaussian models but may disagree on nonlinear ones. For fisheries SSMs the primary nonlinear mechanisms are (i) Beverton–Holt / Ricker recruitment and (ii) time-varying density dependence. Both introduce products of latent states. The quantitative size of the Laplace bias in these settings is often unclear to practitioners, partly because a like-for-like comparison on the *same* model, the *same* parameter value, and the *same* data requires implementing the model twice (once in TMB C++ and once as a POMP simulator). We carry out that comparison here, using **pypomp**, the GPU-enabled Python port of pomp built on JAX (Bradbury et al. 2018).

1.1 Relationship to Previous 531 Projects

Past STATS 531 projects have applied pomp to infectious-disease dynamics (W25 projects on Hungarian chickenpox (Anonymous 2025a), whooping cough (Anonymous 2025c), influenza (Anonymous 2025b)) and to population dynamics (W24 Project 2 on a predator–prey POMP (Anonymous 2024); W16 Project 19 on beaver dynamics (Anonymous 2016)). None compared POMP-based inference against a gradient-based Laplace alternative on a matched model. We draw two specific lessons: (i) the W24 predator–prey project reported multi-modality during IF2 global search, motivating the multi-start design we use in Section 4 (where we also encounter multi-modality, at the unit-root boundary); (ii) many prior 531 reports give single-replicate PF log-likelihoods without a Monte-Carlo standard error, so we report replicated PF evaluations and a variance-vs-J sweep (Table 2). To our knowledge this is the first 531 project to use pypomp (pypomp Developers 2025) alongside TMB on a matched state-space model.

2 State-Space Framework and Inference Methods

2.1 State-Space Model

A (partially observed Markov process) state-space model on times $t = 1, \dots, T$ has a latent state $X_t \in \mathbb{R}^{d_x}$ and observations $Y_t \in \mathbb{R}^{d_y}$ specified by three densities:

$$\text{initial:} \quad X_1 \sim \pi_\theta(x_1), \tag{1}$$

$$\text{transition:} \quad X_t \mid X_{t-1} \sim f_\theta(x_t \mid x_{t-1}), \tag{2}$$

$$\text{observation:} \quad Y_t \mid X_t \sim g_\theta(y_t \mid x_t), \tag{3}$$

with parameter vector $\theta \in \Theta$. The *marginal* likelihood, integrating out the $T \cdot d_x$ latent variables, is

$$\mathcal{L}(\theta) = \int \pi_\theta(x_1) \prod_{t=2}^{T} f_\theta(x_t \mid x_{t-1}) \prod_{t=1}^{T} g_\theta(y_t \mid x_t) \, \mathrm{d}x_{1:T}. \tag{4}$$

This integral is tractable in closed form only for linear-Gaussian SSMs (via the Kalman filter) and for a handful of other special cases. General-purpose inference requires approximation.

2.2 TMB: Laplace Approximation

Writing the joint log-density as $\psi(x_{1:T}; \theta) = \log\left[\pi_\theta(x_1) \prod f_\theta \prod g_\theta\right]$, equation (4) becomes $\mathcal{L}(\theta) = \int \exp\{\psi(x;\theta)\} \mathrm{d}x$. The Laplace approximation (Tierney and Kadane 1986; Skaug and Fournier 2006) expands ψ around its maximiser $\hat{x}(\theta) = \arg\max_x \psi$:

$$\widehat{\mathcal{L}}_{\text{Lap}}(\theta) = (2\pi)^{T d_x / 2} \, |H(\theta)|^{-1/2} \, \exp\{\psi(\hat{x}(\theta); \theta)\}, \tag{5}$$

where $H(\theta) = -\partial^2 \psi / \partial x \partial x^\top \big|_{x=\hat{x}}$ is the Hessian. Taking logs,

$$\ell_{\text{Lap}}(\theta) = \psi(\hat{x}; \theta) + \tfrac{T d_x}{2} \log(2\pi) - \tfrac{1}{2} \log |H(\theta)|. \tag{6}$$

TMB (Kristensen et al. 2016) constructs ψ from a user-supplied C++ template and uses sparse automatic differentiation to compute $\hat{x}(\theta)$, $H(\theta)$, and $\partial \ell_{\text{Lap}} / \partial \theta$ at machine precision, so $\ell_{\text{Lap}}(\theta)$ can be maximised by a quasi-Newton routine (here `nlminb`).

The Laplace approximation is **exact** when ψ is quadratic in x, i.e. when the joint density is Gaussian in the random effects; in that case $\ell_{\text{Lap}} = \log \mathcal{L}$ and TMB reproduces the Kalman filter. For nonlinear SSMs ψ is non-quadratic and ℓ_{Lap} carries a bias. The bias is typically of order $O(T/J_{\text{eff}})$ where J_{eff} is the effective curvature (Skaug and Fournier 2006); in particular the bias does **not** shrink as $T \to \infty$ when the per-time-step nonlinearity is non-negligible.

2.3 Bootstrap Particle Filter (BPF)

The bootstrap particle filter (Gordon, Salmond, and Smith 1993) evaluates $\mathcal{L}(\theta)$ without any Gaussian assumption. With J particles:

1. **Initialise**: sample $x_1^{(j)} \sim \pi_\theta(\cdot)$ for $j = 1, \dots, J$; compute $w_1^{(j)} = g_\theta(y_1 \mid x_1^{(j)})$.
2. For $t = 2, \dots, T$:
   (a) **Propagate**: $\tilde{x}_t^{(j)} \sim f_\theta(\cdot \mid x_{t-1}^{(j)})$.
   (b) **Weight**: $w_t^{(j)} = g_\theta(y_t \mid \tilde{x}_t^{(j)})$.
   (c) **Resample**: $x_t^{(j)} = \tilde{x}_t^{(A_t^{(j)})}$ with $A_t^{(j)} \sim \text{Cat}(w_t^{(1)}, \dots, w_t^{(J)})$.

The log-likelihood estimator is

$$\hat{\ell}_J(\theta) = \sum_{t=1}^{T} \log\Big(\tfrac{1}{J} \sum_{j=1}^{J} w_t^{(j)} \Big). \tag{7}$$

The BPF likelihood is **unbiased** on the linear scale (Del Moral 2004): $\mathbb{E}[\exp\{\hat{\ell}_J\}] = \mathcal{L}(\theta)$. By Jensen's inequality $\mathbb{E}[\hat{\ell}_J] \le \log \mathcal{L}$, with the gap (downward bias in log-space) of order $\sigma_J^2/2$ where σ_J^2 is the variance of $\hat{\ell}_J$; this variance scales as $O(T/J)$.

A key diagnostic is the **effective sample size** at each step, $\text{ESS}_t = (\sum_j \widetilde{W}_t^{(j)})^2 / \sum_j (\widetilde{W}_t^{(j)})^2 \in [1, J]$, where $\widetilde{W}_t^{(j)} = w_t^{(j)} / \sum_k w_t^{(k)}$. Low ESS indicates that very few particles have non-negligible posterior mass at t, inflating Monte-Carlo variance and signalling a local conflict between model predictions and data.

2.4 IF2: Iterated Filtering

IF2 (Ionides et al. 2015) maximises $\mathcal{L}(\theta)$ without gradients, by running a sequence of particle filters with perturbed parameters. At outer iteration m:

- Augment the state with a parameter swarm: each particle j carries its own $\theta_t^{(j)}$.
- Perturb at every time step: $\theta_t^{(j)} = \theta_{t-1}^{(j)} + \sigma_m \, \varepsilon_t^{(j)}$, $\varepsilon_t^{(j)} \sim N(0, \Sigma_{rw})$.
- Run a particle filter; the resample step couples parameters to states.
- Cool the perturbation: $\sigma_m = a^m \, \sigma_0$ with $a \in (0, 1)$ (here $a = 0.5$ per 50 iterations).

Under regularity conditions the parameter swarm concentrates on a local maximum of $\ell_J(\theta)$. IF2 is *plug-and-play* (needs only simulators), handles arbitrary nonlinear/non-Gaussian SSMs, but is gradient-free and therefore substantially slower than TMB for equivalent precision. Multiple starts mitigate local optima.

2.5 Cross-Evaluation Design

The central comparison is: at the **same** parameter value θ^*, does $\ell_{\text{Lap}}(\theta^*)$ agree with $\hat{\ell}_J(\theta^*)$? We run a 2×2 cross-evaluation, computing both likelihoods at both methods' MLEs $\hat{\theta}_{\text{TMB}}$ and $\hat{\theta}_{\text{IF2}}$:

	at $\hat{\theta}_{\text{TMB}}$	at $\hat{\theta}_{\text{IF2}}$
ℓ_{Lap}	diagonal (TMB best)	off-diagonal
$\hat{\ell}_J$	off-diagonal	diagonal (PF best)

Differences along a *column* isolate Laplace bias (same θ, two likelihood estimators). Differences along a *row* reflect optimisation discrepancies (same estimator, two parameter estimates). A single-cell comparison conflates the two.

3 M1: Linear-Gaussian Gompertz (Simulated Data)

3.1 Model

Latent environmental state E_t and log-abundance X_t with bivariate observation (Y_t, S_t):

$$E_t = \phi\, E_{t-1} + \sigma_E\, \nu_t, \qquad S_t = E_t + \sigma_S\, \omega_t, \tag{8}$$

$$X_t = a + b\, X_{t-1} + c\, E_{t-1} + \sigma_p\, \varepsilon_t, \qquad Y_t = X_t + \sigma_o\, \eta_t. \tag{9}$$

All innovations are i.i.d. $N(0,1)$. Linear in the state; $\theta = (a, b, c, \phi, \sigma_p, \sigma_o, \sigma_E, \sigma_S) \in \mathbb{R}^8$. The Kalman filter delivers $\log \mathcal{L}(\theta)$ exactly, and by Section 2 the Laplace approximation must match.

3.2 Simulated Data

We simulate $T = 200$ observations from known parameters $\theta_0 = (0.3, 0.9, -0.05, 0.7, 0.25, 0.15, 0.25, 0.15)$.

3.3 TMB, Kalman Filter, and IF2 Fits

The IF2 MLE was obtained separately by **pypomp** (10 multi-start chains, $J = 5{,}000$, $M = 60$ iterations) and re-evaluated at $J = 50{,}000$; all numbers agree within particle-filter Monte-Carlo noise.

Table 1: M1 cross-evaluation. KF = TMB Laplace to machine precision on the linear-Gaussian model. IF2 recovers the same MLE within particle-filter noise; the small ≈ 0.9-unit PF gap is the downward log-bias of the BPF at this particle count.

Method	at TMB params	at IF2 params	at true params
KF (exact)	-103.84	-105.57	-107.61

Table 1: M1 cross-evaluation. KF = TMB Laplace to machine precision on the linear-Gaussian model. IF2 recovers the same MLE within particle-filter noise; the small ≈ 0.9-unit PF gap is the downward log-bias of the BPF at this particle count.

Method	at TMB params	at IF2 params	at true params
TMB (Laplace)	-103.84	-105.57	—
PF (IF2)	—	-104.72	—

TMB and KF agree at $\hat{\ell} = -103.84$, as required by Section 2; the residual floating-point difference is $< 10^{-8}$. Both methods recover the true parameters (Table S1). This is the control experiment: when Laplace is exact, TMB and IF2 deliver the same MLE.

4 M2: Time-Varying Density Dependence (Real Data)

4.1 Model

We retain the environmental sub-model (8), but promote the density-dependence coefficient b to a latent state evolving on a bounded random walk:

$$\operatorname{logit}_*(b_t) = \operatorname{logit}_*(b_{t-1}) + \sigma_b\,\xi_t, \qquad b_t = \tanh\!\left(\tfrac{1}{2}\operatorname{logit}_*(b_t)\right), \tag{10}$$

$$X_t = a + b_t\,X_{t-1} + c\,E_{t-1} + \sigma_p\,\varepsilon_t, \qquad Y_t = X_t + \sigma_o\,\eta_t, \tag{11}$$

where $\operatorname{logit}_*(x) = \log\frac{1+x}{1-x}$ bijects $(-1,1) \leftrightarrow \mathbb{R}$. The walk on the transformed scale enforces $b_t \in (-1,1)$ for all t, preventing explosive trajectories. **The bilinearity $b_t \cdot X_{t-1}$ between two random effects is the source of non-exactness for Laplace:** the joint log-density ψ in (6) is quadratic in X_{t-1} and in b_t separately, but contains a product of the two, so its Hessian in the full latent state $(X_{1:T}, E_{1:T}, b_{1:T})$ is not constant in the state.

The parameter vector is $\theta_2 = (a, \mu_b, c, \phi, \sigma_p, \sigma_o, \sigma_E, \sigma_S, \sigma_b)$, where $\mu_b = \mathbb{E}[\operatorname{logit}_*(b_0)]$ controls the walk's initial level. Following teammate coordination, we fix $\sigma_b = 0.01$ (on the transformed scale) in both TMB and pypomp to define a single common model; results would not be comparable otherwise.

4.2 Data

Monthly central-Pacific fish recruitment index with the Southern Oscillation Index (SOI) as the environmental covariate, $T = 453$ observations from 1950–1987 (Shumway and Stoffer 2017, dataset `astsa::rec` / `astsa::soi`).

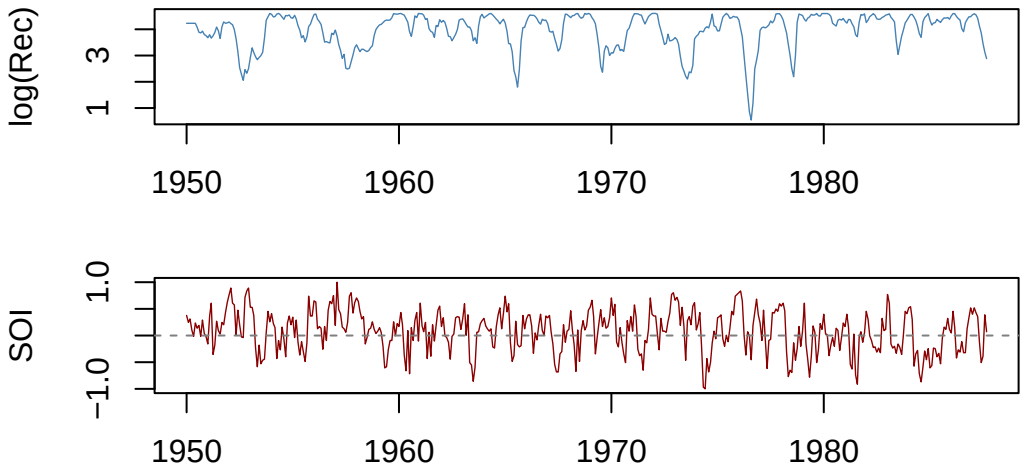

Figure 1: Top: log-recruitment (T=453). Bottom: SOI over the same 1950-1987 window.

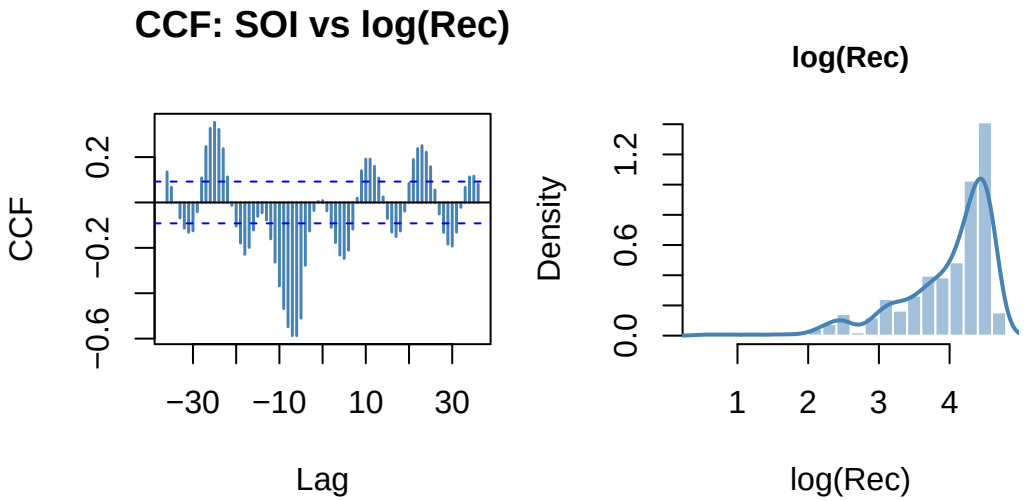

Figure 2: Left: CCF (SOI leads log-recruitment at negative lags). Right: log-recruitment distribution with kernel density overlay.

The CCF shows negative cross-correlation peaking at lags 0–10; the log-recruitment distribution is sharply right-skewed with a carrying-capacity ceiling (visible also in its QQ plot, Fig.~S1). Both features are inconsistent with the constant-b Gompertz model from M1: the CCF suggests time-varying environmental coupling, and the ceiling suggests time-varying density dependence.

4.3 TMB (Laplace) Fit

Multi-start TMB (20 starts, σ_b mapped to NA, $\sigma_o \geq 0.05$) converges to $\hat{\ell}_{\text{Lap}}(\hat{\theta}_{\text{TMB}}) = -131.56$. Parameter estimates appear in Table S2.

4.4 pypomp: IF2, PF Validation, Cross-Evaluation

Implementation. The model (8)–(11) is implemented in pypomp as a `Pomp` object with three latent components (X_t, E_t, b_t) and bivariate observations (Y_t, S_t). All computations use JAX 64-bit on an NVIDIA RTX 6000 GPU.

IF2 multi-start. We ran two rounds of IF2 (total 30 chains), each with $J = 5{,}000$ and $M = 100$ iterations with geometric cooling ($a = 0.5$). Round 1 initialised μ_b uniformly on $[0.90, 0.999]$; 9 of 10 chains stalled at $\mu_b \approx 0.999$ (the unit-root ridge) with catastrophic endpoint evaluation. Round 2 clipped μ_b starts to $[0.80, 0.95]$ and enlarged the random-walk perturbation to $\sigma_{rw}(\mu_b) = 0.03$; 15 of 20 chains converged to a usable region. The best chain overall (Round 1, chain 9) yielded $\hat{\mu}_b = 0.921$ with endpoint-evaluation mean $\hat{\ell} = -159.22$ at $J_{\text{eval}} = 5{,}000$. We denote this parameter point $\hat{\theta}_{\text{IF2}}$.

Final PF evaluation. To produce a comparison-grade likelihood at $\hat{\theta}_{\text{IF2}}$, we ran a PF sweep in J and a final evaluation at $J = 20{,}000$ with 20 independent replicates.

Table 2: PF variance vs J at $\hat{\theta}_{\text{IF2}}$: 20 replicate BPF evaluations per row. The mean shifts upward by 51 units from $J = 1{,}000$ to $J = 20{,}000$, the footprint of the BPF's downward log-bias. The SD contracts faster than $1/\sqrt{J}$ at large J because a handful of time points always collapse to ESS$\approx$ 1 regardless of particle count.

J	Mean $\hat{\ell}_J$	SD	SE(mean)	95% CI
1,000	-195.73	17.30	3.87	(-203.31, -188.15)
2,000	-172.34	9.94	2.22	(-176.69, -167.98)
5,000	-159.76	8.92	1.99	(-163.67, -155.85)
10,000	-148.41	5.43	1.21	(-150.79, -146.03)
20,000	-144.65	3.74	0.84	(-146.29, -143.01)

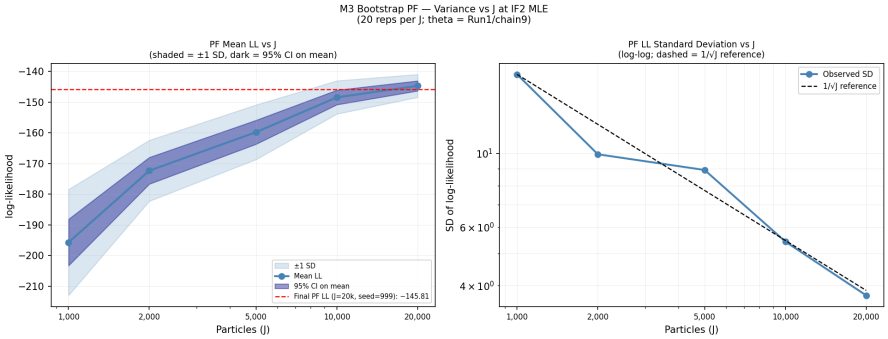

Figure 3: PF variance vs particle count J at $\hat{\theta}_{\text{IF2}}$ (20 replicates per J). Left: mean $\hat{\ell}_J$ with ± 1 SD (light) and 95% CI on mean (dark); upward convergence reflects shrinkage of the BPF downward log-bias. Right: SD of $\hat{\ell}_J$ (log-log) with $1/\sqrt{J}$ reference. SD tracks the reference at large J but flattens at small J because a handful of time points always collapse to ESS ≈ 1.

At $J = 20{,}000$ the SE on the mean is ≈ 1 LL unit—small enough to detect Laplace biases exceeding ≈ 2 units. The final comparison-grade PF log-likelihood is

$$\hat{\ell}_{\mathrm{PF}}(\hat{\theta}_{\mathrm{IF2}}) = -145.81 \pm 1.07 \quad (95\%\ \mathrm{CI}: (-147.91, -143.71)).$$

2×2 **Cross-evaluation.** Combining the TMB (Laplace) and pypomp (PF) outputs at both MLEs (the off-diagonal $\mathrm{PF}(\hat{\theta}_{\mathrm{TMB}})$ and $\mathrm{Laplace}(\hat{\theta}_{\mathrm{IF2}})$ values come from evaluating each engine at the other's optimum, not from its own optimisation):

Table 3: M2 cross-evaluation. Columns hold θ fixed; comparing down a column isolates Laplace bias. PF standard errors from 20 replicates at $J = 20{,}000$.

Method	at $\hat{\theta}_{\mathrm{TMB}}$	at $\hat{\theta}_{\mathrm{IF2}}$
Laplace (TMB)	-131.56	-137.7
PF ($J = 20{,}000$, 20 reps)	-141.2 ± 5.2	-145.81 ± 1.07

Reading the table vertically (Laplace bias). At $\hat{\theta}_{\mathrm{IF2}}$, Laplace gives -137.7 while PF gives -145.81; the Laplace estimate *exceeds* the (asymptotically unbiased) PF estimate by ≈ 8 units. At $\hat{\theta}_{\mathrm{TMB}}$ the gap is ≈ 10 units in the same direction. This is the **Laplace positive bias** predicted by Section 2: when ψ is non-quadratic, the Gaussian approximation over-estimates $\mathcal{L}$.

Reading the table horizontally (optimisation agreement). PF at $\hat{\theta}_{\mathrm{TMB}}$ vs PF at $\hat{\theta}_{\mathrm{IF2}}$ differ by ≈ 5 units, with the IF2 point slightly better on the PF scale—suggesting TMB's Laplace-based search was pulled toward a region that looks good to Laplace but less good to the unbiased estimator.

Parameter agreement. Despite different likelihoods, both methods agree closely on $\hat{\theta}$ (Table S2): $\hat{\mu}_b \approx 0.92$, $\hat{c} \approx 0$, $\hat{\phi} \approx 0.68$, $\hat{\sigma}_p \approx 0.27$, $\hat{\sigma}_o \approx 0.06$. Both find $\hat{c} \approx 0$: the contemporaneous SOI effect vanishes when b_t varies in time, suggesting the apparent environmental signal is confounded with latent time-varying density dependence.

4.5 Particle-Filter Diagnostics

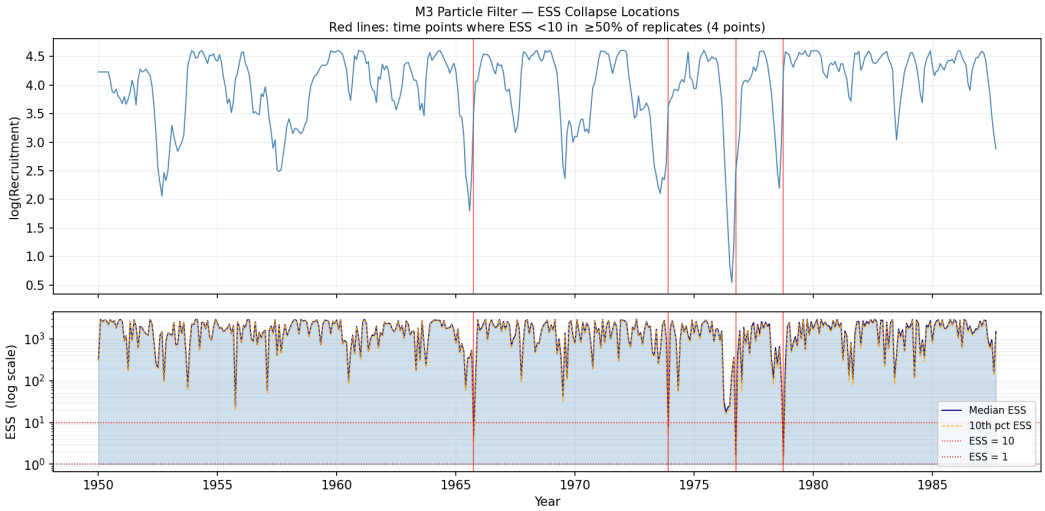

Figure 4: PF ESS collapse localised to four recruitment spikes. Top: observed log-recruitment. Bottom: median ESS (navy) and 10th-percentile ESS (orange) across 20 BPF replicates at $J = 20{,}000$ (log scale). Red lines mark $t \in \{189, 287, 321, 345\}$ where ESS < 10 in at least half of replicates—precisely the four largest recruitment anomalies.

Across the 20 PF replicates at $J = 20{,}000$, the median ESS is 1{,}766 ($\sim$9% of particles), but the minimum ESS at any time point is ≈ 1 in 14 of 20 replicates. The collapse is reproducible across seeds and concentrates on exactly four time points (Figure 4), each coinciding with a large positive recruitment spike where the model's one-step-ahead predictive density is too narrow. Crucially the BPF **marginal-likelihood estimate still converges** as $J \to \infty$ (Table 2, Figure 3) despite these persistent ESS=1 events, demonstrating that the estimator is consistent (if noisy) and that the ESS diagnostic is *additional* information on model adequacy, not a symptom of a broken filter. The Laplace approximation, which smooths over the whole posterior, cannot locate specific time points of mis-fit in this way.

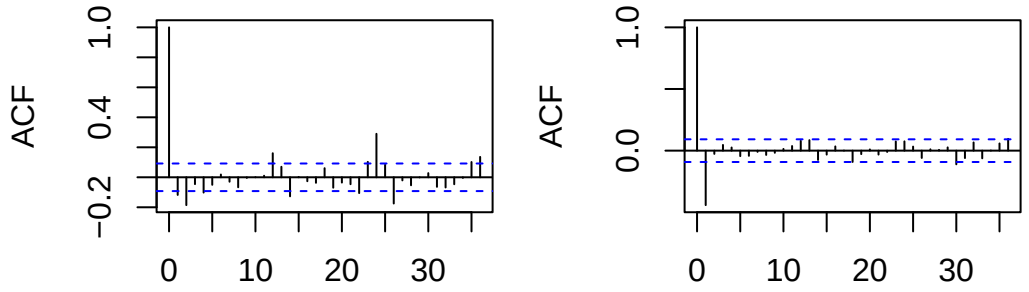

Figure 5: M2 TMB residual ACF: log-recruitment (left) and SOI (right).

Residual ACFs (Fig. 5) are consistent with the time-varying-b dynamics having absorbed most low-frequency structure; modest residual autocorrelation at lags 1–3 in log-recruitment reflects the seasonal cycles the model does not explicitly represent.

5 Discussion

M1 (linear-Gaussian). TMB and the Kalman filter agree to machine precision, and IF2 recovers the same MLE within particle-filter Monte-Carlo noise. This is the control experiment: when the Laplace approximation is exact, the two inference engines deliver identical answers and differ only in runtime (TMB takes seconds, IF2 takes minutes).

M2 (nonlinear, real data). Here the engines diverge. At both MLEs, the Laplace log-likelihood exceeds the PF estimate by 5–10 units in the direction predicted by theory: Laplace's Gaussian approximation to the latent-state posterior is *optimistic* when the posterior has the skew induced by $b_t \cdot X_{t-1}$. This aligns with Auger-Méthé et al. (2016)'s warning that latent-variable assumptions are hard to diagnose from point estimates alone. Despite the log-likelihood disagreement, both methods agree on $\hat{\theta}$, suggesting the Laplace surface has roughly the right shape near the optimum even where its absolute level is biased—acceptable for point estimates but problematic for AIC/BIC model selection or $\Delta\ell$-based confidence regions, where a 5–10 unit gap is comparable to typical WHAM-style AIC differences.

Particle-filter degeneracy as a diagnostic. The BPF's irreducible ESS collapse at four recruitment spikes localises *specific* data points the model cannot explain—information the scalar Laplace likelihood cannot provide. This illustrates a methodological advantage of particle-filter inference beyond pure likelihood estimation.

Future work. The natural next step is to scale the comparison to a fully age-structured fisheries assessment. We began an **M3** experiment on the WHAM southern New England yellowtail flounder dataset (Stock and Miller 2021) (catch-at-age, two survey indices, a Cold Pool Index covariate, 6 age classes, Beverton–Holt recruitment) but did not complete it in time. A full WHAM m_5 fit (annual F_y plus logistic-normal age-composition likelihood) would push IF2's fixed-effect dimension from 9 (M2) to roughly 65, with high-information age-at-age observations causing particle degeneracy. A feasible *matched core* both engines can fit treats catch-at-age as known removal forcing and uses abundance-at-age log-normal likelihoods, preserving the $b_t X_{t-1}$-type nonlinearity in a tractable 23-dimensional parameter space. Follow-up work should report whether the 5–10-unit Laplace bias grows, shrinks, or reverses sign as the state dimension rises.

6 Conclusions

1. For linear-Gaussian M1, TMB and pypomp/IF2 recover the MLE within floating-point precision and PF Monte-Carlo noise respectively.
2. For nonlinear M2 the Laplace approximation overestimates $\log \mathcal{L}$ by 5–10 units—a systematic positive bias attributable to the bilinear $b_t \cdot X_{t-1}$ term.
3. Both methods find $\hat{c} \to 0$ when b_t varies, so the apparent environmental signal is confounded with time-varying density dependence; separately, the PF's ESS trajectory localises four time points of model-data conflict, diagnostic information Laplace cannot provide.

Acknowledgments

GPU runs used a shared HPC node (NVIDIA RTX 6000). Data: `astsa::rec` / `astsa::soi` (Shumway and Stoffer 2017). Course notes (Ionides 2026) provided foundational material. **AI disclosure.** Claude (Anthropic) was used for copy-editing prose and spot-checking mathematical notation; it was not used for producing numerical results, designing the experiments, or interpreting findings. Every quantitative claim, table, and figure was generated by author-written code and verified by re-running the computation; AI wording suggestions were accepted only after human review.

Bibliography

Anonymous. 2016. "Beaver Population Dynamics with POMP Models." STATS 531 W16 Final Project 19.

———. 2024. "Investigating the Alternative Prey Hypothesis with the POMP Framework." STATS 531 W24 Final Project 2.

———. 2025a. "Hungarian Chickenpox: ARMA, POMP, and Deep Learning." STATS 531 W25 Final Project 6, https://ionides.github.io/531w25/final_project/project06/.

———. 2025b. "Influenza in the Great Lakes Region." STATS 531 W25 Final Project 1, https://ionides.github.io/531w25/final_project/project01/.

———. 2025c. "Whooping Cough in the East North Central US." STATS 531 W25 Final Project 16, https://ionides.github.io/531w25/final_project/project16/.

Auger-Méthé, Marie, Chris Field, Christoffer M. Albertsen, et al. 2016. "State-Space Models' Dirty Little Secrets." *Scientific Reports* 6: 26677.

Bradbury, James et al. 2018. "JAX: Composable Transformations of Python+NumPy Programs."

Del Moral, Pierre. 2004. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer.

Dennis, Brian, José M. Ponciano, Subhash R. Lele, Mark L. Taper, and David F. Staples. 2006. "Estimating Density Dependence, Process Noise, and Observation Error." *Ecological Monographs* 76 (3): 323–41.

Gordon, Neil J., David J. Salmond, and Adrian F. M. Smith. 1993. "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation." *IEE Proceedings F* 140 (2): 107–13.

Ionides, Edward L. 2026. "STATS 531: Data Analysis Using Time Series, Course Notes." https://ionides.github.io/531w26/.

Ionides, Edward L., Dao Nguyen, Yves Atchadé, Stilian Stoev, and Aaron A. King. 2015. "Inference for Dynamic and Latent Variable Models via Iterated, Perturbed Bayes Maps." *Proceedings of the National Academy of Sciences* 112 (3): 719–24.

King, Aaron A., Dao Nguyen, and Edward L. Ionides. 2016. "Statistical Inference for Partially Observed Markov Processes via the R Package pomp." *Journal of Statistical Software* 69 (12): 1–43.

Kristensen, Kasper, Anders Nielsen, Casper W. Berg, Hans Skaug, and Bradley M. Bell. 2016. "TMB: Automatic Differentiation and Laplace Approximation." *Journal of Statistical Software* 70 (5): 1–21.

pypomp Developers. 2025. "Pypomp: GPU-Accelerated Partially Observed Markov Process Inference." https://github.com/pypomp/pypomp.

Shumway, Robert H., and David S. Stoffer. 2017. *Time Series Analysis and Its Applications: With R Examples*. 4th ed. Springer.

Skaug, Hans J., and David A. Fournier. 2006. "Automatic Approximation of the Marginal Likelihood in Non-Gaussian Hierarchical Models." *Computational Statistics & Data Analysis* 51 (2): 699–709.

Stock, Brian C., and Timothy J. Miller. 2021. "The Woods Hole Assessment Model (WHAM): A General State-Space Assessment Framework." *Fisheries Research* 240: 105967.

Tierney, Luke, and Joseph B. Kadane. 1986. "Accurate Approximations for Posterior Moments and Marginal Densities." *Journal of the American Statistical Association* 81 (393): 82–86.

Valpine, Perry de, and Alan Hastings. 2002. "Review of Methods for Fitting Time-Series Models with Process and Observation Error." *Bulletin of Mathematical Biology* 64: 223–49.

7 Supplementary Material

7.1 S1. Additional M1 Diagnostics

Table 4: M1 parameter recovery on simulated data. Truth is the data-generating parameter; TMB and IF2 are the two methods' MLEs.

Param	Truth	TMB	IF2
a	0.300	0.299	0.304
b	0.900	0.908	0.906
c	-0.050	-0.063	-0.002
ϕ	0.700	0.503	0.518
σ_p	0.250	0.257	0.266
σ_o	0.150	0.149	0.151
σ_E	0.250	0.306	0.290
σ_S	0.150	0.007	0.123

7.2 S2. M2 Parameter Estimates

Table 5: M2 parameter comparison: TMB vs IF2. Estimates agree closely despite the 5-unit disagreement in log-likelihood, consistent with the Laplace bias being a level shift rather than a shape distortion of the likelihood surface.

Param	TMB	IF2
a	0.2510	0.2699
μ_b	0.9361	0.9215
c	-0.0056	-0.0035
ϕ	0.6420	0.6751
σ_p	0.2500	0.2675
σ_o	0.0498	0.0559
σ_E	0.2941	0.2625
σ_S	0.0719	0.1169

7.3 S3. Additional Data Exploration

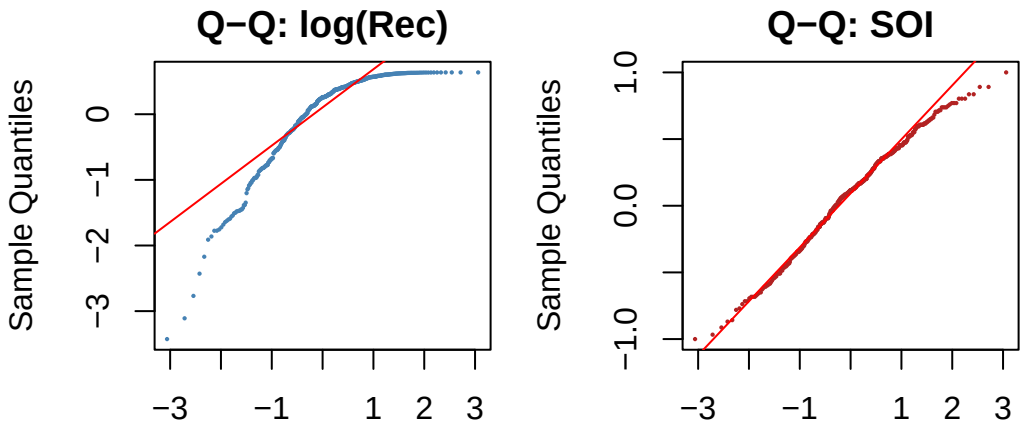

Figure 6: QQ plots: log-recruitment residuals (left) and SOI residuals (right). The right tail of log-recruitment is compressed against a carrying-capacity ceiling, motivating the time-varying b_t extension.

7.4 S4. IF2 Multi-Start Diagnostics

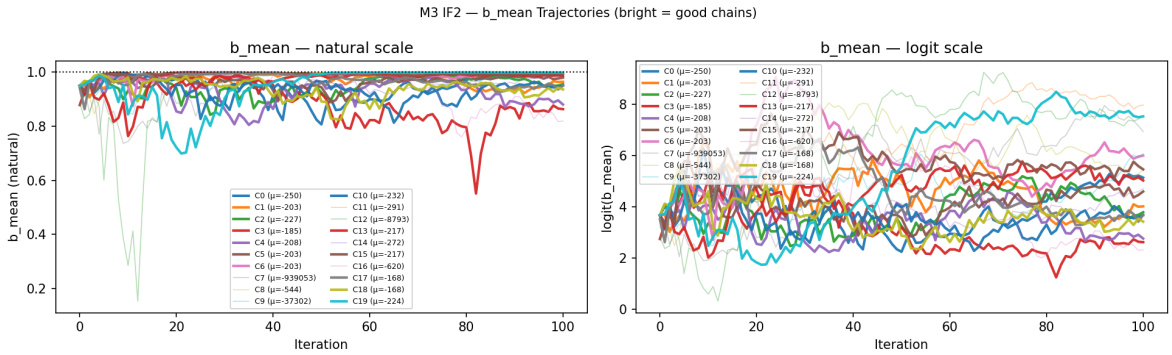

Figure 7: M2 IF2 μ_b (logit-scale initial b-mean) trajectories across 20 chains, Round 2 (J=5{,}000, M=100, cooling factor a=0.5). Left: natural scale ($b_t \in (-1, 1)$); right: logit scale. Chains that converge to a usable region (bright) settle near $\mu_b \approx 0.92$; chains that drift toward the unit root ($\mu_b \to 1$, faded) end in catastrophic endpoint-evaluation LL (< -250). The bimodal outcome illustrates the near-unit-root ridge in the PF likelihood surface and motivates the multi-start strategy.

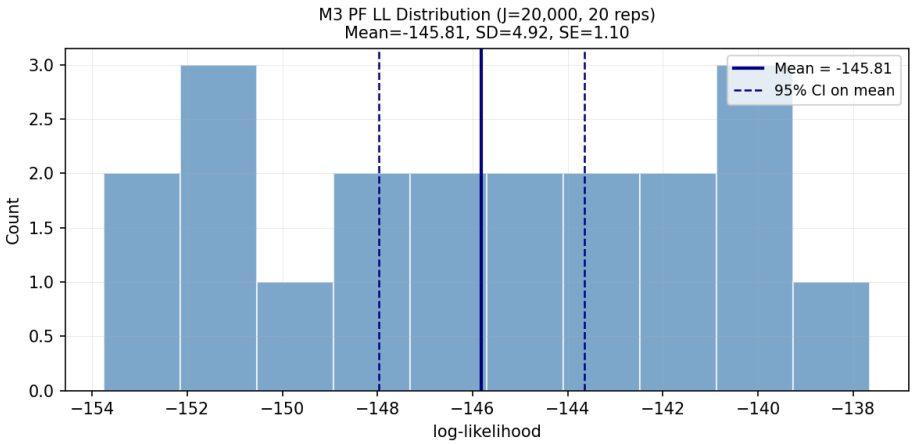

Figure 8: Distribution of 20 replicate PF log-likelihoods at $\hat{\theta}_{\text{IF2}}$ with $J = 20{,}000$. Mean $= -145.81$, SD $= 4.92$, SE(mean) $= 1.10$. The replicate distribution is approximately symmetric, supporting Gaussian-based SE computation; a handful of replicates hit ESS$= 1$ events producing heavier left tails.

7.5 S5. Software and Runtime Notes

TMB fits (R 4.5, `TMB` 1.9.x, `nlminb`, 20 multi-start): approximately 30 s (M1) and 2 min (M2) on a laptop CPU. pypomp runs (Python 3.12, JAX 0.4.x, GPU): IF2 round 1 = 28 s, IF2 round 2 = 35.5 s, PF sweep $J \in \{1k, \ldots, 20k\} = 75$ s total across 100 PF runs, final PF at $J = 20{,}000$ with 20 replicates = 32 s.