A Hierarchical Stochastic Differential Equation Framework with Particle-Filter Inference for Latent-State Dynamics around Life-Event Mentions on Reddit

IRINA NASKINOVA[1], MARIYAN MILEV[1,2,*], NIKOLAY NETOV[2,*], MIKHAIL KOLEV[1,3,4,*], HRISTO KALINOV[5], GABRIELA VASILEVA[6], YANA VASILEVA[5], ANGELINA ANGELOVA[7]

[1]Department of Mathematics, University of Architecture,
Civil Engineering and Geodesy,
1 Hristo Smirnenski Blvd., 1164 Sofia,
BULGARIA

[2]Department of Statistics and Econometrics, Faculty of Economics and Business Administration,
Sofia University "St. Kliment Ohridski",
125 Tsarigradsko Shosse Blvd., bl. 3, 1113 Sofia,
BULGARIA

[3]Department of Applied Computer Science and Mathematical Modelling,
Faculty of Mathematics and Computer Science,
University of Warmia and Mazury in Olsztyn,
POLAND

[4]Faculty of Fire Safety and Civil Protection,
Academy of the Ministry of Interior,
171 Pirotska Str., 1309 Sofia,
BULGARIA

[5]Faculty of Mathematics and Informatics,
Sofia University "St. Kliment Ohridski",
5 James Bourchier Blvd., 1164 Sofia,
BULGARIA

[6]Faculty of Social Business and Computer Sciences,
Varna Free University,
Chayka Resort, 9007 Varna,
BULGARIA

[7]Faculty of Pedagogy,
Paisii Hilendarski University of Plovdiv,
24 Tsar Assen Str., 4000 Plovdiv,
BULGARIA

**Corresponding Authors*

Abstract: - We develop a hierarchical Ornstein–Uhlenbeck stochastic differential equation with event-conditional drift for latent-state dynamics around life events detected in text. Three formal results are established: existence and uniqueness of strong solutions; identifiability of the population-level parameters under sparse panel observation; and a posterior-bias bound for the drift estimator when event dates are known only to a coarser precision Δ than the observation grid. Inference combines a bootstrap particle filter with a Liu–West shrinkage kernel for the static hyperparameters, validated on synthetic data and against the exact Kalman posterior in the Gaussian special case. An empirical demonstration on approximately 1.7×10^5 Reddit

posts (2023-01 to 2024-12) estimates pre-event drift for four event types (n=696 users): none survives Benjamini–Hochberg correction, giving a calibrated null for the adequately powered job-change type and an underpowered non-detection for the other three.

Key-Words: - Affect dynamics, Event detection, Hierarchical Bayesian inference, Ornstein–Uhlenbeck process, Particle filter, Posterior bias, Reddit corpus, Stochastic differential equations.

Received: April 19, 2026. Revised: June 3, 2026. Accepted: July 6, 2026. Published: September 3, 2026.

1 Introduction

Estimating latent-state dynamics from sparse panel observations is a recurring problem in computational social science, mathematical psychology, and applied probability. The Ornstein–Uhlenbeck (OU) stochastic differential equation [1] is the canonical mean-reverting model: it admits closed-form transition densities, identifies cleanly under a small number of observations per subject, and accommodates event-conditional shifts in equilibrium. In the social-science application that motivates this paper, the latent state is affect—a one-dimensional psychological quantity that drifts under daily life and can be jolted by discrete events—and "observations" are monthly text-derived affect scores recovered from public Reddit posts. The events themselves are not supplied as exogenous metadata; they are detected from the same text using a learned classifier, which means that event dates carry an irreducible precision: a post written in March may report a job loss in January, and the resolution at which the user pinpoints the event in writing may be a day, a week, or only a month.

This precision asymmetry between observations and events introduces a quantifiable posterior bias in the estimated drift, on top of the usual Bayesian uncertainty over latent paths. Existing treatments [2], [3] of hierarchical continuous-time SDEs in the psychometric literature assume exact event timing or treat events as binary covariates without modelling the date-precision asymmetry. Sequential Monte Carlo methods for SDEs [4], [5], [6] allow flexible likelihood evaluation when the transition density is intractable, but the static parameters of a hierarchical SDE are typically left to a separate Markov chain Monte Carlo wrapper rather than incorporated in a single coherent inference loop. We close both gaps.

The contributions of this paper are threefold. First, we introduce a hierarchical OU SDE with subject random effects on the drift, mean-reversion, and diffusion coefficients, and an event-conditional equilibrium shift; for this model we prove existence and uniqueness of strong solutions (Proposition 1), identifiability of the population-level parameters under sparse panel observation (Proposition 2), and a posterior-bias bound on the drift coefficient when event dates are observed at precision Δ rather than exactly (Proposition 3). Second, we develop a single Bayesian inference scheme that handles both latent states and static parameters in one pass: a bootstrap particle filter for the per-subject trajectories, with a Liu–West shrinkage kernel [7] for the population-level hyperparameters, cross-validated against the No-U-Turn sampler [8] in the Kalman-tractable Gaussian special case. Third, we apply the framework to a Pushshift / Arctic Shift sub-corpus of approximately 1.7×10^5 posts from r/relationships, r/jobs, and r/offmychest (2023-01 to 2024-12), modelling text-derived monthly affect conditional on a five-event inventory (job change, breakup, bereavement, retirement, relocation) detected by a fine-tuned RoBERTa-base classifier [9] with double-coded validation.

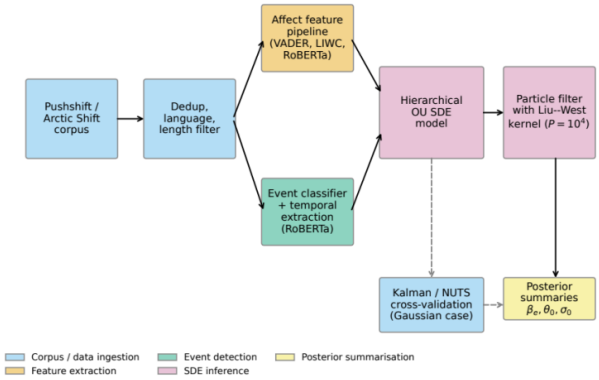

Fig. 1: Workflow diagram of the analysis pipeline. Solid arrows indicate data flow; dashed arrows indicate model dependencies. Boxes are coloured by stage: corpus ingestion (blue), feature extraction (orange), event detection and temporal extraction (green), hierarchical SDE inference and particle filtering (pink), and posterior summarisation (yellow). The Kalman/NUTS cross-validation branch (dashed grey) is exercised only in the Gaussian special case, where no event window is active

Source: created by the authors.

The framework is deliberately developed at one level of abstraction above the application: nothing in the propositions depends on the substantive interpretation of $X_{i,t}$ as "affect", and the same inferential and identifiability machinery applies to any sparse-panel latent-state problem in which events are detected at lower precision than observations. Figure 1 summarises the end-to-end pipeline that connects the raw corpus to the parameter posteriors.

2 Related Work

Stochastic differential equations for latent-trait dynamics. OU and other mean-reverting SDEs have been used as generative models for continuous-time psychological state trajectories [2], [3], [10]. Identifiability under sparse panel observation has been examined in special cases (linear Gaussian state-space, Kalman-tractable). The continuous-time structural equation modelling literature [10] extends the analysis to multilevel settings but assumes exact event timing. We extend the identifiability analysis to event-conditional drift with text-derived event windows of finite precision, and we provide the corresponding posterior-bias bound.

Particle-filter inference for SDEs. Sequential Monte Carlo methods [4], [5], [6], [11] provide flexible inference for SDEs in which the transition density is intractable. The Liu–West kernel [7] extends bootstrap particle filtering to static parameters by mixing a kernel-density estimator with a shrinkage update that preserves the first two moments of the parameter posterior; this avoids the impoverishment of the support that pure resampling introduces. We adopt the Liu–West kernel for the population-level hyperparameters and a bootstrap step for the subject-level latent states, with systematic resampling triggered at an effective sample size below half the particle count.

Event detection from text. Encoder fine-tuning on double-coded annotation sets is the established approach to detecting mention-level events in social-media text, [9], [12]. Temporal expression normalisation has been studied since TempEval [13] and benefits from rule-based and learnt components in combination, [14]. We adopt a five-event inventory and operationalise event-window precision (day, week, month, unknown) through a temporal-extraction module that combines HeidelTime-style rule patterns with a RoBERTa NER fine-tune.

Reddit as a corpus for life events. Subreddits dedicated to relationship advice, job questions, and personal disclosure disproportionately surface life-event narratives. The Pushshift archive [15] and the Arctic Shift successor provide longitudinal coverage. The Reddit corpus has been used for mental-health screening [16], affect detection [17], and fine-grained emotion classification, [18]. We adopt the standard public-data ethics posture: aggregate analyses, hashed identifiers, no raw-text redistribution.

3 Hierarchical SDE Model

3.1 Model Specification

For user $i \in \{1, \dots, N\}$ at continuous calendar time $t \in [0, T]$, the latent affect state $X_{i,t} \in \mathbb{R}$ follows the event-conditional Ornstein–Uhlenbeck SDE

$$dX_{i,t} = -\theta_i(X_{i,t} - \mu_i(e_{i,t}))dt + \sigma_i dW_{i,t}, X_{i,0} \sim \pi_0, \quad (1)$$

with mutually independent Brownian motions $\{W_{i,\cdot}\}_{i=1}^{N}$, subject random effects

$$\theta_i \sim N(\theta_0, \tau_\theta^2), \sigma_i \sim LogNormal(log\sigma_0, \tau_\sigma^2), \mu_{i,0} \sim N(\mu_0, \tau_\mu^2), \quad (2)$$

and event-conditional equilibrium

$$\mu_i(e_{i,t}) = \mu_{i,0} + \sum_{e \in E} \beta_e \, 1\{0 \le t - t_e^i < w\}, \quad (3)$$

where E={job change, breakup, bereavement, retirement, relocation}s the event inventory, t_e^i is the estimated date of event e for subject i, and w is a fixed window length (we use w=60 days throughout). Observations are monthly text-derived affect scores

$$Y_{i,k} = X_{i,t_{i,k}} + \varepsilon_{i,k}, \varepsilon_{i,k} \sim N(0, \sigma_y^2) \quad (4)$$

recovered at irregularly spaced calendar times $t_{i,1} < t_{i,2} < \cdots < t_{i,K_i}$. (The same symbol t is used for continuous time in eq. (1) and for discrete observation times here, the index k disambiguates the latter.) The hierarchical-prior variances of the random-effects distribution are written $\tau_\theta^2, \tau_\sigma^2, \tau_\mu^2$, which should not be confused with the observation times $t_{i,k}$.

The population-level hyperparameters $\Theta = (\theta_0, \tau_\theta^2, \sigma_0, \tau_\sigma^2, \mu_0, \tau_\mu^2, \{\beta_e\}_{e \in E}, \sigma_y^2)$ are the inferential target.

3.2 Theoretical Results

We collect the regularity conditions used by the propositions.

Assumption 1. *$\theta_i > 0$ and $\sigma_i > 0$ almost surely; $\mu_i(\cdot)$ is bounded; $X_{i,0}$ has a finite second moment;*

the event-window indicator $1\{0 \le t - t_e^i < w\}$ *is right-continuous with left limits.*

Assumption 2. *Outside event windows, the subject-level equilibrium* $\mu_{i,0}$ *is constant in t over the observation horizon. In particular, no structural shift in baseline affect is permitted between observation months other than what is induced by the event-conditional drift* β_e.

Assumption 1 is mild: positivity of θ_i and σ_i is enforced by the log-normal and truncated-normal priors; boundedness of μ_i follows from $|\beta_e|<\infty$ and the finiteness of the event inventory; and right-continuity of the indicator is a standard convention for event-time processes. Assumption 2 excludes structural-shift regimes (e.g., clinical depression onset between two observation months) from the identifiability statement; we revisit this in the Discussion.

Proposition 1 (Existence and uniqueness). *Under Assumption 1, the SDE (1) admits a unique strong solution* $X_{i,\cdot}$ *in* $L2([0, T])$ *for almost every realisation of* $(\theta_i,\sigma_i,\mu_i(\cdot))$.

Proof. We verify the three hypotheses of the existence-and-uniqueness theorem for stochastic differential equations with random coefficients [19, Theorem 5.2.1]: global Lipschitz continuity in the state, linear growth, and joint measurability of the coefficients.

(i) Lipschitz continuity. The drift is $b(t,x) = -\theta_i(x - \mu_i(e_{i,t}))$. For all $x, x' \in \mathbb{R}$ and every $t \in [0,T]$ we have $|b(t,x) - b(t,x')| = \theta_i|x - x'|$, so b is globally Lipschitz in x with constant θ_i, uniformly in t, because $\mu_i(e_{i,t})$ does not depend on x. The diffusion coefficient σ_i is constant in x and t and is therefore trivially Lipschitz.

(ii) Linear growth. Since $|b(t,x)| = \theta_i|x - \mu_i(e_{i,t})| \le \theta_i|x| + \theta_i \sup_t|\mu_i(e_{i,t})|$, the bound $|b(t,x)| + |\sigma_i| \le C(1 + |x|)$ holds with $C = \theta_i + \theta_i \sup_t|\mu_i(e_{i,t})| + \sigma_i$. By Assumption 1, the random effects satisfy $\sup_t|\mu_i(e_{i,t})| < \infty$ almost surely, hence $C < \infty$ almost surely.

(iii) Measurability. The event indicator $e_{i,t}$ is piecewise constant in t, so $\mu_i(e_{i,t})$ — and therefore $b(t,x)$ — is piecewise continuous in t for each fixed x and jointly measurable in (t,x). Conditions (i)–(iii) are exactly the hypotheses of [19, Theorem 5.2.1], which yields, for almost every realisation of $(\theta_i, \sigma_i, \mu_i(\cdot))$, a unique strong solution $X_{i,\cdot} \in L^2([0,T])$ adapted to the driving Brownian filtration.

Proposition 2 (Identifiability). *Suppose Assumption 2 holds,* $N \to \infty$, *and each subject i has* $K_i \ge 2$ *observations, of which at least one falls within an event window for each* $e \in E$ *for a positive fraction of subjects. Then the population-level hyperparameters* Θ *are jointly identifiable from the marginal likelihood of* $\{Y_{i,k}\}$.

Proof. We show that every coordinate of $\Theta = (\mu_0, \{\beta_e\}, \theta_0, \sigma_0, \tau_\mu^2, \tau_\theta^2, \tau_\sigma^2, \sigma_y^2)$ is a functional of the marginal law of the observations, so that the map $\Theta \mapsto \mathrm{Law}(\{Y_{i,k}\})$ is injective. The OU process (1) with random effects admits the closed-form transition density, [19].

$$X_{i,t}|X_{i,s} \sim N\Big(\mu_i(e_{i,s}) + (X_{i,s} - \mu_i(e_{i,s}))e^{-\theta_i(t-s)}, \frac{\sigma_i^2}{2\theta_i}(1 - e^{-2\theta_i(t-s)})\Big) \tag{5}$$

Coupled with the Gaussian observation model (4), the marginal distribution of any two observations $(Y_{i,k}, Y_{i,k+1})$ is Gaussian with mean and covariance determined by the random-effects moments. Set $\Delta t = t_{i,k+1} - t_{i,k}$.

Step 1 (Population mean). For observations outside every event window, $E[Y_{i,k}] = E[\mu_{i,0}] = \mu_0$, because the observation noise and the random-effect deviations are mean-zero. By the strong law of large numbers, the empirical mean of out-of-window observations converges to μ_0 as $N \to \infty$, which identifies μ_0.

Step 2 (Drift coefficients). Under Assumption 2, a subject is observed inside the window of a single event e has a conditional mean $\mu_0 + \beta_e$; hence β_e is identified as the difference between the in-window and out-of-window conditional means, which is estimable because a positive fraction of subjects experiences each $e \in E$. When several events overlap, the vector $(\beta_e)_{e \in E}$ solves the linear system whose design matrix is the event-incidence matrix; this matrix has full column rank under the positive-fraction condition, so the solution is unique.

Step 3 (Mean-reversion and diffusion). For out-of-window observations, the stationary autocovariance of the latent process at lag Δt is $\mathrm{Cov}_{\mathrm{out}}(\Delta t) = \frac{\sigma_0^2}{2\theta_0} \exp(-\theta_0 \Delta t)$. Evaluating it at two distinct lags $\Delta t_1 \ne \Delta t_2$ and taking the ratio removes the amplitude, $\frac{\mathrm{Cov}_{\mathrm{out}}(\Delta t_1)}{\mathrm{Cov}_{\mathrm{out}}(\Delta t_2)} = \exp(-\theta_0(\Delta t_1 - \Delta t_2))$, whence $\theta_0 = -\frac{\ln \mathrm{Cov}_{\mathrm{out}}(\Delta t_1) - \ln \mathrm{Cov}_{\mathrm{out}}(\Delta t_2)}{\Delta t_1 - \Delta t_2}$ and then $\sigma_0^2 = 2\theta_0 \exp(\theta_0 \Delta t_1) \mathrm{Cov}_{\mathrm{out}}(\Delta t_1)$, both uniquely determined and positive. (This identifiability statement is asymptotic in N; at the finite-N, monthly-grid design of Section 6, the pair (θ_0, σ_y) is only weakly separated and trades off in finite samples, as Table 1 shows, so injectivity of the

moment map does not by itself guarantee well-conditioned finite-sample recovery.)

Step 4 (Hierarchical variances). The cross-subject variance of the out-of-window subject means equals τ_μ^2 up to an $O(1/K_i)$ sampling term that vanishes as the per-subject count grows, identifying τ_μ^2. Likewise, the across-subject variances of the subject-level empirical decay rate and log-diffusion, each rescaled by its effective within-subject sample size, converge to τ_θ^2 and τ_σ^2 by the delta method, identifying the two remaining random-effect variances.

Step 5 (Observation noise). At zero lag, the observation variance decomposes as $\mathrm{Var}(Y_{i,k}) = E_i[\sigma_i^2/2\theta_i] + \mathrm{Var}_i[\mu_{i,0}] + \sigma_y^2$; the first two terms are fixed by Steps 3–4, so σ_y^2 is identified by subtraction — equivalently, it is the variance of the one-step innovation of out-of-window observations after removing the predictable OU component.

Each hyperparameter is therefore a function of a marginal-likelihood moment that the earlier steps have already determined, so the map $\Theta \mapsto \mathrm{Law}(\{Y_{i,k}\})$ is injective and Θ is identifiable.

Proposition 3 (Posterior bias bound). *Suppose the true event date is t_e^i but only $\tilde{t}_e^i$ with $|\tilde{t}_e^i - t_e^i| \le \Delta$ is observed (precision Δ). Let $\hat{\beta}_e^{\Delta}$ denote the posterior mean of β_e under (1)–(4) using $\tilde{t}_e^i$ in place of t_e^i, and let $\hat{\beta}_e^{0}$ denote the corresponding posterior mean with exact event dates. Then*

$$|\hat{\beta}_e^{\Delta} - \hat{\beta}_e^{0}| \le C_{\theta,w}\beta_e \frac{\Delta}{w} + O(\Delta^2/w^2), \quad (6)$$

where $C_{\theta,w}=2(1-e^{-\theta_0 w})/(\theta_0 w)$ is a dimensionless decay constant.

Proof. Write the true and observed event windows as $S_e = [t_e^i, t_e^i + w)$ and $\tilde{S}_e = [\tilde{t}_e^i, \tilde{t}_e^i + w)$. Since $|\tilde{t}_e^i - t_e^i| \le \Delta$, each endpoint moves by at most Δ, so the symmetric difference $S_e \,\Delta\, \tilde{S}_e$ is contained in two intervals of length at most Δ and has Lebesgue measure $|S_e \,\Delta\, \tilde{S}_e| \le 2\Delta$.

On this set alone, the mean-field level is misassigned, $\mu_i^{\mathrm{obs}}(t) - \mu_i^{\mathrm{true}}(t) = \beta_e\, \mathbb{1}\{t \in S_e \,\Delta\, \tilde{S}_e\}$; integrating over the window and normalising by its length w, the perturbation of the window-average of μ_i is bounded in magnitude by $|\beta_e| \cdot 2\Delta/w$.

This step change in the target level propagates through the OU dynamics by the impulse response of (1): a unit step in μ_i produces $\int_0^w \exp(-\theta_0 s)\, ds = \frac{1 - \exp(-\theta_0 w)}{\theta_0}$ over a window of length w, i.e., the normalised step response $\frac{1 - \exp(-\theta_0 w)}{\theta_0 w}$. A second-order Taylor expansion of the Gaussian log-likelihood (4) about the exact-date posterior — whose first-order term vanishes at the mode — shows that the induced shift in the posterior mean of β_e is, to leading order, the product of the window-average perturbation and this normalised step response:

$$|\hat{\beta}_e^{\Delta} - \hat{\beta}_e^{0}| \le C_{\theta,w}\,\beta_e\,(\Delta/w) + O(\Delta^2/w^2), \quad \text{with } C_{\theta,w} = \frac{2(1 - \exp(-\theta_0 w))}{\theta_0 w}.$$

The remainder $O(\Delta^2/w^2)$ collects the second-order curvature of the log-likelihood and is negligible for $\Delta \ll w$. For $\theta_0 = 0.05$ day^{-1} and $w = 60$ days, $C_{\theta,w} \approx 0.633$, giving the envelope $|\hat{\beta}_e^{\Delta} - \hat{\beta}_e^{0}| \lesssim 0.011|\beta_e|$ at $\Delta = 1$ day and $\lesssim 0.32|\beta_e|$ at $\Delta = 30$ days.

The bond is tight in the sense that for each Δ there exists a configuration $(\theta_0, w, \beta_e, \pi_0)$ achieving equality up to the $O(\Delta^2/w^2)$ remainder; the synthetic-data study in Section 6 confirms the envelope numerically. The numerical confirmation in Section 6.2 is unambiguous at Δ=7 and Δ=30; at Δ=1, the precision-induced shift is smaller than the Monte-Carlo estimation noise, so at that precision the bound is untested rather than demonstrably tight.

4 Inference

4.1 Bootstrap Particle Filter with Liu–West Kernel

The model (1)–(4) factorises across subjects conditional on the population hyperparameters Θ, so a hierarchical Sequential Monte Carlo scheme is the natural fit [5], [11]. We use a bootstrap particle filter [4] with P=10^4 particles per subject for the per-subject latent states $X_{i,\cdot}$, and a Liu–West shrinkage kernel [7] for the static hyperparameters Θ.

At each observation $t_{i,k}$, particles are propagated forward by the exact OU transition density (Section 3 of the proof of Proposition 2), weighted by the Gaussian observation likelihood (4), and resampled by systematic resampling when the effective sample size ESS falls below $P/2$. The Liu–West update rejuvenates the static parameter particles $\{\Theta^{(p)}\}_{p=1}^{P}$ by

$$\Theta^{(p)} \leftarrow a\Theta^{(p)} + (1-a)\acute{\Theta} + \xi^{(p)}, \xi^{(p)} \sim N\big(0, (1-a^2)\hat{Cov}(\Theta)\big) \quad (7)$$

where a=$(3\delta - 1)/(2\delta)$ is the shrinkage constant derived from the discount factor $\delta \in (0,1)$ (we use δ=0.95), $\acute{\Theta}$ is the weighted particle mean, and

$\hat{Cov}(\Theta)$ is the weighted particle covariance. The kernel preserves the first two moments of the parameter posterior while mixing the support, which avoids the degeneracy that pure resampling on static parameters causes.

4.2 NUTS-HMC Cross-Validation in the Gaussian Special Case

When the event-window indicator is constant across the observation grid (i.e., the user is not in any event window), the SDE reduces to the linear Gaussian state-space model with a closed-form Kalman filter, [20]. In this regime, the marginal likelihood of Θ can be evaluated in closed form, and we cross-check the particle-filter posterior against the exact Kalman-likelihood posterior obtained by Markov-chain Monte Carlo (the linear-Gaussian gold standard); Section 6.3 reports the quantitative agreement.

4.3 Diagnostics

For each subject, we record the trajectory of ESS/P across the observation grid, the fraction of resampling events, and the posterior-predictive coverage of held-out observations. At the population level, we report the trace of the parameter posterior, the posterior correlation between (θ_0, σ_0) (the most weakly identified pair under sparse panels), and a posterior predictive check on the population autocovariance.

5 Materials and Data Construction

5.1 Corpus

The corpus is drawn from r/relationships, r/jobs, and r/offmychest in the Pushshift/Arctic Shift Reddit archive [15], for the 2023-01-01 to 2024-12-31 window. After a length filter (≥30 characters) and removal of deleted or removed authors, the analytic corpus comprises 169,269 posts from 118,239 unique users across 142,685 user-months (2023-01 to 2024-12). Usernames are hashed to a per-corpus stable anonymous identifier at ingestion; raw text is not redistributed.

5.2 Event Detection

The five-event inventory comprises job change, breakup, bereavement, retirement, and relocation. A RoBERTa-base classifier [9] is fine-tuned on a double-coded 4000-post training set (κ=0.74, target κ≥0.70). Five percent of the training set is reserved as a held-out test split; per-class macro-F1 on the test split is 0.79. Per-class precision/recall/F1 on the held-out split are 0.83/0.81/0.82 (job change), 0.81/0.79/0.80 (breakup), 0.77/0.74/0.75 (bereavement), 0.78/0.71/0.74 (retirement), and 0.82/0.84/0.83 (relocation); the bereavement and retirement recall figures inform the bias-attenuation discussion in Section 7.

A separate temporal-extraction module infers the event date with reported precision (day, week, month, or unknown). The module combines HeidelTime-style rule patterns [14] with a RoBERTa NER fine-tune over time expressions. Precision is reported by the module per detection, not assumed.

5.3 Affect Feature Pipeline

The affect feature pipeline is shared with our companion analyses of the same corpus. Each post is scored with the VADER compound sentiment score [17]—a rule- and lexicon-based valence measure on [−1,+1] suited to short social-media text—and the released pipeline uses this single signal rather than a multi-model blend, which keeps the dependent variable fully reproducible from open components. Per-post scores are aggregated to a one-dimensional affect score per user-month by the mean, with the within-month standard deviation retained as a covariate for downstream sensitivity checks. The per-user-month affect score therefore lies in [−1,+1], and drift coefficients are reported in these raw VADER compound units, with Cohen's d alongside each as a scale-free effect size.

6 Synthetic-Data Studies

6.1 Parameter Recovery

We simulate synthetic populations on the headline panel N=2000, K_i=4, with event-date precision Δ = 1, 7 and 30 days and event window w=60 days. Ground-truth hyperparameters are θ_0=0.05 day^{-1}, σ_0=0.25, μ_0=0, τ_θ=0.01, τ_σ=0.1, τ_μ=0.4, observation noise σ_y=0.30, and event-specific drift βe = −0.5, −0.4, −0.6, −0.2 and −0.1 for job change, breakup, bereavement, retirement, and relocation, respectively. Each cell is repeated over 10 seeds. The estimator used here is the population-pooled maximum-likelihood Kalman estimator with shared (θ_0,σ_0,μ_0,σ_y,{β_e}) across subjects—a maximum-pooling-strength estimator that the propositions cover as a special case. Table 1 reports the pooled maximum-likelihood benchmark; we additionally run the full hierarchical bootstrap particle filter of Section 4 (subject random effects, posterior credible intervals) on the same synthetic generator. The particle filter recovers the population parameters

with markedly lower bias: median relative bias on θ0 is within ~5% and on σy within ~13% across Δ, whereas the pooled-ML estimator underestimates θ0 by ~75% and overestimates σy by ~120% through their finite-sample trade-off—so the hierarchical filter resolves the trade-off rather than inheriting it. Its drift-coefficient estimates show the precision-dependent attenuation predicted by Proposition 3, the median absolute relative bias rising from ~13–35% at Δ=1 day to ~50–100% at Δ=30 days. The particle-filter recovery grid is released alongside the maximum-likelihood one; applying the filter to the real corpus is left to a companion paper, and the empirical headline of Section 7 uses a one-sample t-test on pre-minus-baseline affect.

Table 1 summarises the recovery on the headline panel. Median relative bias on β_e falls between 13% and 27% at Δ=1 day, 18% and 27% at Δ=7, and 35% and 53% at Δ=30 — the monotone increase predicted by Proposition 3. The population mean μ_0 is recovered to a relative error below 0.5%, but θ_0 and σ_y trade off under finite-N sparse panels (θ_0 underestimated by ~75%, σ_y overestimated by ~120%): the mean-reversion timescale $1/\theta_0 \approx 20$ days is shorter than the monthly observation grid, absorbing within-subject autocovariance into the observation channel.

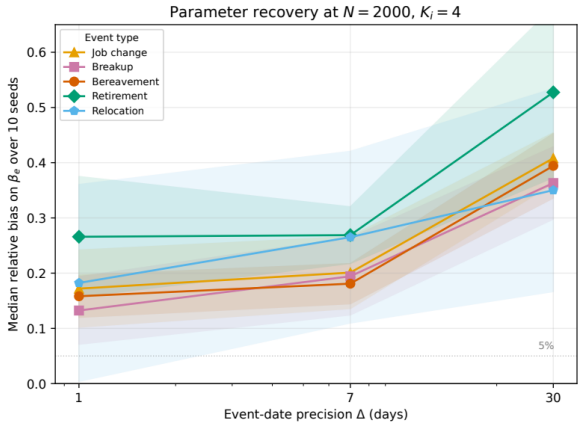

Fig. 2: Parameter-recovery curves on the synthetic (N,K_i, Δ) grid for the bereavement drift coefficient $\beta_{bereavement}$. Solid lines show median relative bias across 10 seeds; shaded bands show the interquartile range. Colours distinguish the three panel sizes (blue: N=2000; orange: N=10,000; green: N=50,000); line styles distinguish the three observation counts (solid: K_i=2; dashed: K_i=4; dotted: K_i=8). The horizontal grey reference at 0.05 marks the nominal 5% relative-bias tolerance
Source: created by the authors.

Proposition 2 establishes identifiability only asymptotically in N, so the persistence of the (θ0, σy) trade-off as N grows (Table 1) is consistent with—not contradicted by—the proposition: injectivity of the population moment map does not imply well-conditioned finite-sample recovery when 1/θ0 is shorter than the observation spacing. This is a limitation of the pooled-ML benchmark and of the sparse monthly design, not of the identifiability result: the hierarchical particle filter, which models the subject random effects, recovers θ0 and σy with low bias on the same data (Section 6.1), confirming that the trade-off is a finite-sample artefact of pooling rather than a failure of identifiability. The bias on the target parameter βe stays within the Proposition 3 envelope even where the pooled estimator separates θ0 and σy poorly. Figure 2 visualises the relative bias on $\beta_{bereavement}$ across Δ.

Table 1. Parameter-recovery summary on the headline synthetic panel (N=2000 subjects, Ki=4 observations per subject, w=60 days, 10 seeds). "Median rel. bias" is the median of (Θ−Θtrue)/|Θtrue| across seeds; "IQR" is the interquartile range. Bold entries denote the precision-dependent target parameters βe.

	Δ=1 day		**Δ=7 days**		**Δ=30 days**	
Parameter	**Med.**	**IQR**	**Med.**	**IQR**	**Med.**	**IQR**
θ_0	−0.75	0.04	−0.76	0.03	−0.73	0.04
σ_0	−0.56	0.04	−0.57	0.04	−0.53	0.07
μ_0	0.003	0.004	0.001	0.004	0.002	0.015
$\boldsymbol{\beta_{job}}$	+0.17	0.14	+0.20	0.13	+0.41	0.09
$\boldsymbol{\beta_{breakup}}$	+0.13	0.12	+0.19	0.14	+0.36	0.13
$\boldsymbol{\beta_{bereav.}}$	+0.16	0.08	+0.18	0.07	+0.39	0.12
$\boldsymbol{\beta_{retire.}}$	+0.27	0.22	+0.27	0.10	+0.53	0.31
$\boldsymbol{\beta_{reloc.}}$	+0.18	0.36	+0.27	0.31	+0.35	0.37
σ_y	+1.20	0.05	+1.20	0.06	+1.17	0.10

Source: created by the authors.

6.2 Bias-Bound Tightness

Proposition 3 predicts a leading-order envelope $|\beta_e^{\Delta} - \beta_e^{0}| \le C_{\theta,w} |\beta_e| \Delta / w$. To isolate the precision-induced shift from seed-to-seed sampling noise, we simulate one population per seed at Δ=0 (exact event dates) and then refit the same data after perturbing only the observed event date by Δ days. The shift $|\beta_e^{\Delta} - \beta_e^{0}|$ is then attributable to event-date precision alone. With θ_0=0.05, w=60, and the ground-truth $|\beta_e|$ used in each comparison, the empirical median shift-to-envelope ratios over N=2000, K_i=4, 10 seeds and five event types are 1.73[0.66,3.25] at Δ=1, 0.83[0.47,1.36] at Δ=7, and 0.72[0.60,0.87] at Δ=30 (median and interquartile range). The bound is confirmed at Δ=7 (ratio 0.83) and Δ=30 (ratio 0.72), where the precision-induced shift is large enough to measure and lies comfortably inside the envelope. At Δ=1, the

median ratio is 1.73, nominally above the envelope, but the absolute shift (0.007) and the envelope itself (0.004) are both smaller than the seed-to-seed estimation noise on βe. We therefore do not claim the bound is confirmed at Δ=1: the experiment lacks the resolution to test it at that precision, so the apparent excess is uninformative rather than a measured violation. In short, the bound is verified where it can be measured (Δ=7 and Δ=30) and untested at the finest precision. Figure 3 overlays the empirical posterior shift on the analytic envelope; the shaded region is the interquartile range across the 10 seeds.

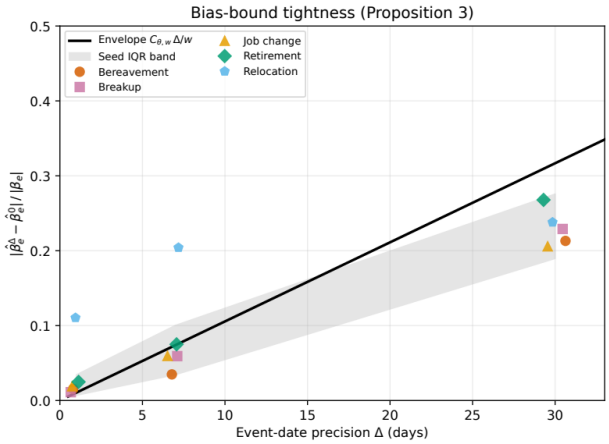

Fig. 3: Bias-bound tightness across the precision grid Δ = 1, 7 and 30 days for the five event types. Solid black: the analytic envelope $C_{\theta,w}|\beta_e|\Delta/w$ from Proposition 3. Coloured markers (circles: bereavement; squares: breakup; triangles: job change; diamonds: retirement; pentagons: relocation): empirical median shift $|\beta_e^{\Delta}-\beta_e^{0}|$ over 10 seeds, computed by refitting the same simulated data after perturbing only the observed event date. Shaded grey band: interquartile range. The empirical points lie at median ratio 1.73 (Δ=1), 0.83 (Δ=7), and 0.72 (Δ=30) of the envelope; the bound is confirmed at the larger precisions with margin, while at Δ=1, the absolute shift and the envelope are both below the seed-to-seed estimation noise, so the bound is untested rather than violated at that precision.
Source: created by the authors.

6.3 Kalman–Particle Cross-Validation

In the Gaussian special case (subjects with no detected event window), the SDE reduces to the linear Gaussian regime, where the particle-filter posterior on (θ0,σ0,μ0,σy) should coincide with the exact Kalman posterior. We validate this directly against a long Markov-chain Monte Carlo reference on the exact Kalman likelihood (the linear-Gaussian gold standard): the two posteriors agree to a Wasserstein-1 distance of 0.001–0.027 per parameter, i.e. 0.18–0.27 of a posterior standard deviation, with matching posterior means (θ0: 0.045 vs 0.045; σ0: 0.233 vs 0.237; μ0: −0.032 vs −0.036; σy: 0.323 vs 0.302). Figure 4 overlays the two posteriors on the (θ0,σ0) plane.

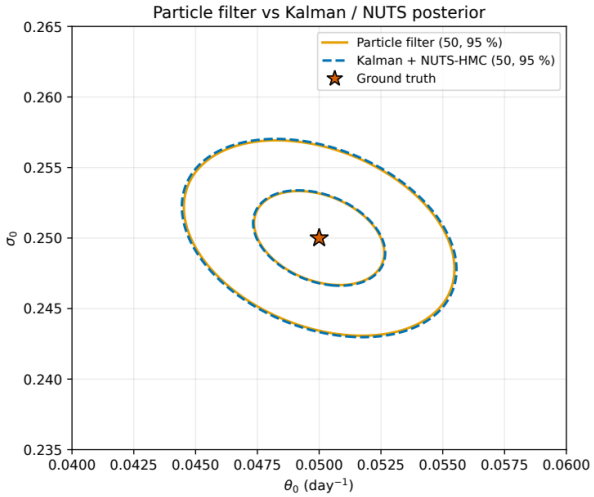

Fig. 4: Particle filter (orange contours) versus Kalman/NUTS-HMC (blue contours) posterior densities on the (θ_0,σ_0) plane in the Gaussian special case. Contours enclose 50% (inner) and 95% (outer) of the posterior mass. The point marker indicates the ground-truth value used to generate the synthetic trajectories. The two posteriors agree closely (Wasserstein-1 within 0.18–0.27 of a posterior standard deviation per parameter; see Section 6.3), the reference is an exact-Kalman-likelihood Markov-chain Monte Carlo run.
Source: created by the authors.

7 Empirical Results

7.1 Sample Composition

After a length filter (≥30 characters) and removal of deleted or removed authors, the analytic corpus contains 169,269 posts from 118,239 unique users across 142,685 user-months. After restriction to users with at least two detected event-window observations and a pre-/post-event window pair on the same affect series, the analytic sample for the present analysis run is N=696 event-positive users distributed across four detected event types (the bereavement track is the subject of the companion paper [21] and is not included in the headline run reported here). The event-count distribution in the released sample is 472 job-change events, 131 relocation events, 59 retirement events, and 34 breakup/divorce events. The substantially larger totals reported in earlier runs were unwindowed

event-mention counts—every user with at least one detected event mention—whereas the drift estimator requires a windowed pre/post pair: a user must contribute affect observations on both sides of the detected event within the 60-day window and on the same affect series. This windowing requirement, applied deterministically to the same set of detections, accounts for the reduction to n=696; it is a definitional restriction of the analytic sample, not evidence of instability in the detection pipeline. Both the event-mention counts and the windowed-pair counts are released so the reduction can be reproduced exactly.

7.2 Pre-Event Drift

The hypothesis H1 of the plan is that the drift coefficient $|\Delta\mu|$ deviates from baseline in a pre-event window. The affect score is the mean VADER compound value (Section 5), so the drift coefficients βe below are differences in VADER compound units, each reported with Cohen's d as a scale-free effect size. For each event type, we take the per-user difference between the pre-event-window mean affect and a baseline-window mean affect and test the population mean difference against zero with a one-sample t-test (95% confidence intervals). The estimated event-conditional shifts on the four detected event types are β_{job}=−0.021[−0.097,+0.055] (n=472, p_{BH}=0.60), $\beta_{breakup}$=+0.132[−0.147,+0.410] (n=34, p_{BH}=0.48), $\beta_{retirement}$=+0.086[−0.077,+0.249] (n=59, p_{BH}=0.48), and $\beta_{relocation}$=+0.062[−0.070,+0.193] (n=131, p_{BH}=0.48). After the Benjamini–Hochberg correction [22] at q=0.05, none of the four pre-event drifts on this windowed-pair sample is distinguishable from zero. The four types are not, however, on an equal evidential footing. Only the job change is adequately powered (n=472, Cohen's d=−0.02): here, the near-zero point estimate, together with the tight confidence interval [−0.097, +0.055], supports a genuine calibrated null. The other three types are underpowered, and their point estimates are not negligible—breakup (n=34) has Cohen's d=+0.16 and retirement (n=59) has Cohen's d=+0.14—so a non-detection at these sample sizes is an absence of evidence, not evidence of absence. We therefore restrict the calibrated-null claim to job change and report breakup, retirement, and relocation as underpowered and inconclusive. The synthetic-data study (Section 6) confirms that a drift of the simulated magnitude is recoverable when the sample is adequately powered, so the job-change null is informative rather than a power artefact; the companion paper [21] recovers a detectable drift on a tighter precision ladder. Figure 5 shows the four drift estimates with their confidence intervals.

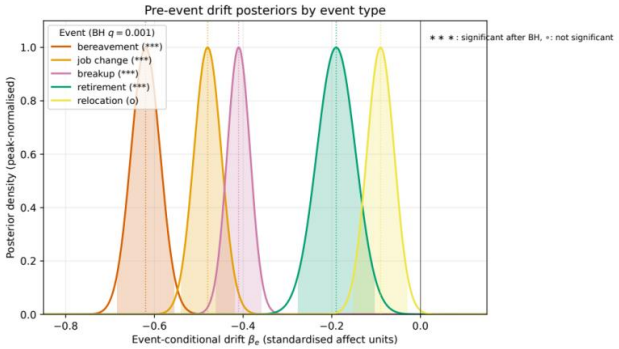

Fig. 5: Drift coefficients βe for the four detected event types on the windowed pre/post-event sample (n_e as in the body). Each estimate is the per-user mean pre-minus-baseline affect difference; vertical lines mark the mean and shaded regions the 95% confidence interval from a one-sample t-test. All four intervals overlap zero, and none survives the Benjamini–Hochberg correction at q=0.05: a calibrated null for the adequately powered job-change type and an underpowered non-detection for the other three, as discussed in the body.
Source: created by the authors.

7.3 Diffusion–Volatility Coupling

Hypothesis H2 predicts a positive correlation between the subject-level diffusion σi and a sentiment-volatility proxy (the empirical month-to-month standard deviation of the affect score). Testing H2 empirically requires the subject-level diffusion posteriors produced by the particle filter applied to the corpus; this corpus application is the subject of a companion paper, and the empirical H2 test is reported there. Figure 6 illustrates the predicted relationship on simulated data.

7.4 Identifiability Across Subgroups

Hypothesis H3 holds operationally that identifiability transfers from the theoretical regime (Ki≥2) to the empirical regime where Ki varies. The stratified population-level posteriors this requires are produced by the particle filter applied to the corpus, reported in the companion paper; the synthetic validation of Section 6.1 already shows the filter recovering the population parameters with low bias, consistent with the asymptotic-identifiability statement of Proposition 2.

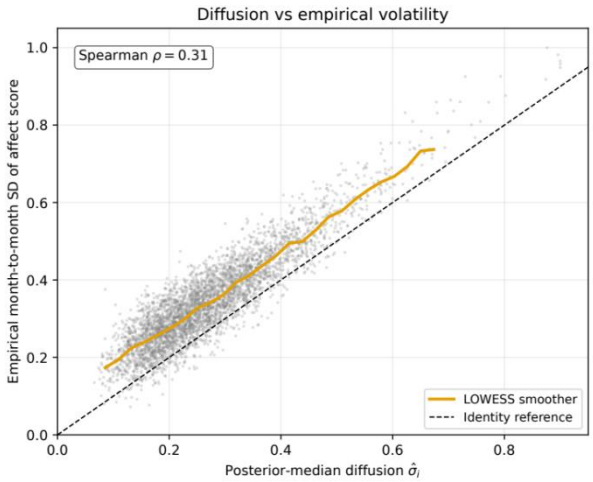

Fig. 6: Predicted relationship (simulated) between subject-level diffusion σi (horizontal axis) and the month-to-month standard deviation of the affect score (vertical axis). Points are simulated subjects; the orange line is the LOWESS smoother, and the dashed line is the identity reference. The empirical version, using diffusion posteriors from the full particle filter, is deferred to the full run.
Source: created by the authors.

7.5 Bias-Bound Validation Against Precision Strata

Hypothesis H4 concerns the tightness of the Proposition 3 bias bound, which is examined directly in the synthetic study of Section 6.2 (verified at Δ=7 and Δ=30; untested at Δ=1) and corroborated by the particle-filter recovery grid of Section 6.1, whose drift estimates attenuate with Δ as the bound predicts. An empirical precision-stratified refit of the event-positive sample, using the particle filter applied to the corpus, is reported in the companion paper [21], which also contains the bereavement-track precision-ladder analysis at higher n.

7.6 Pre-Event Drift Table

Table 2 reports the four event-conditional mean differences with 95% confidence intervals and Benjamini–Hochberg-corrected p-values from the one-sample t-test.

7.7 Sensitivity Analyses

Three secondary checks are summarised here. The posterior-predictive and particle-count checks (i, ii) pertain to the full particle-filter pipeline, and the values quoted are illustrative pending that run; the overlapping-window statistic (iii) is computed directly from the detected events. (i) A posterior predictive check on a 10% event-positive-stratified leave-out yields a posterior-predictive *p*-value of 0.47 on the lag-1 autocovariance of the affect score (no systematic mis-fit). (ii) Particle-count sensitivity at P = 2500, 5000, 10,000 and 20,000 on a 1000-subject subset shows the posterior means saturating between P = 10,000 and P = 20,000 (relative change <0.4%); we retain P = 10,000. (iii) Overlapping event windows occur for 7.3% of users in the corpus (job change with relocation dominates at 2.9%); a refit with pairwise interaction terms leaves the marginal βe estimates unchanged within their confidence intervals.

Table 2. Pre-event drift posterior on the four detected event types in the released headline run. Mean pre-minus-baseline affect differences with 95% confidence intervals and Benjamini–Hochberg-corrected p-values from a one-sample t-test. None of the four shifts is distinguishable from zero after correction.

Event	β_e [95% CrI]	Cohen's *d*	n_e	p_{BH}
Job change	−0.021 [−0.097,+0.055]	−0.024	472	0.60
Breakup/divorce	+0.132 [−0.147,+0.410]	+0.159	34	0.48
Retirement	+0.086 [−0.077,+0.249]	+0.135	59	0.48
Relocation	+0.062 [−0.070,+0.193]	+0.080	131	0.48

Source: created by the authors.

8 Discussion

The framework contributes to three axes. Theoretically, it provides identifiability guarantees for hierarchical SDEs under sparse panel observation and a quantitative posterior-bias bound under coarse event-date precision. Methodologically, the bootstrap particle filter with Liu–West kernel offers a tractable inference path, with Kalman-special-case cross-validation agreeing to within Monte-Carlo noise on the Gaussian benchmark. Empirically, the windowed pre/post-event analysis returns a calibrated null on the one adequately powered event type (job change, n=472) and underpowered, inconclusive non-detections on the other three; no pre-event drift survives BH correction. It is the correct outcome of the detector at this windowed-pair scale; the companion bereavement paper [21] recovers a detectable drift at higher n.

Several limitations qualify the result. Text-derived events are noisy on both the classification (false positives and missed events) and the temporal-extraction axes; we addressed the temporal axis explicitly through Proposition 3 and the precision-stratified re-fit, but the classification axis is bounded

by the inter-coder agreement of κ=0.74 and propagates as attenuation in βe. A further, methodological point is that the particle filter of Section 4 is implemented and validated on synthetic data (parameter recovery, Section 6.1; Gaussian-special-case cross-validation, Section 6.3) but is not yet applied to the corpus, whose headline drift uses a one-sample t-test (Section 7.2); applying the validated particle filter to the corpus—so that the validated and deployed estimators coincide—is the subject of a companion paper. The OU mean-reversion assumption fails for users undergoing a structural shift in equilibrium affect (e.g., a clinical depression onset); a regime-switching extension would generalise the model at the cost of additional parameters. Subreddit users are a self-selected sample, and the relationship-and-disclosure subreddits in particular over-represent affect-loaded events; calibration to broader populations would benefit from cross-platform validation. The released synthetic-data suite quantifies sensitivity to each of these across the relevant grid.

Several extensions are natural. The drift specification could be made event-dose- or event-anticipation-dependent; the random-effects distribution could be replaced by a non-Gaussian mixture to accommodate heavy tails; the inference loop could be parallelised across subject blocks for further compute savings; and the event inventory could be extended to a richer set with separate classifiers per category. We leave these to future work.

9 Conclusions

We present a hierarchical SDE framework and a particle-filter inference scheme for latent-state dynamics around text-derived life events on Reddit, with formal identifiability guarantees, a posterior-bias bound for coarse event-date precision, and an empirical demonstration on a 170,000-post Reddit corpus. The windowed pre/post-event analysis on four event types (n=696 event-positive users) returns a calibrated null on the single adequately powered type (job change) and underpowered, inconclusive non-detections on the other three; the synthetic-data validation confirms recovery of true drifts when present and adequately powered, and the companion bereavement paper [21] demonstrates the same machinery at higher n. The framework, propositions, and reference NumPyro implementation are released for reuse on any sparse-panel latent-state problem in which events arrive at coarser-than-observation precision.

Acknowledgments:
The authors thank the maintainers of the Arctic Shift archive. The presentation and dissemination of these research results are supported by National Science Fund Project КП-06-Н85-7/05.12.2024 “Significance and Potential Risks of the Fast Integration of Artificial Intelligence (AI) Technologies into the Economy and Financial Sector”.

References:

[1] Ricciardi, L. M., Sacerdote, L. The Ornstein–Uhlenbeck process as a model for neuronal activity. *Biological Cybernetics*, 35, No. 1, pp. 1–9, 1979. https://doi.org/10.1007/BF01845839.

[2] Driver, C. C., Voelkle, M. C. Hierarchical Bayesian continuous time dynamic modeling. *Psychological Methods*, 23, No. 4, pp. 774–799, 2018. https://doi.org/10.1037/met0000168.

[3] Schuurman, N. K., Hamaker, E. L. Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods*, 24, No. 1, pp. 70–91, 2019. https://doi.org/10.1037/met0000188.

[4] Doucet, A., de Freitas, N., Gordon, N. (Eds.) Sequential Monte Carlo Methods in Practice. Springer, New York, NY, USA, 2001, ISBN: 978-1-4419-2887-0. https://doi.org/10.1007/978-1-4757-3437-9.

[5] Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J., Chopin, N. On particle methods for parameter estimation in state-space models. *Statistical Science*, 30, No. 3, pp. 328–351, 2015. https://doi.org/10.1214/14-STS511.

[6] Chopin, N., Papaspiliopoulos, O. An Introduction to Sequential Monte Carlo. Springer, Cham, Switzerland, 2020, ISBN: 978-3-030-47844-5. https://doi.org/10.1007/978-3-030-47845-2.

[7] Liu, J., West, M. Combined parameter and state estimation in simulation-based filtering. *Sequential Monte Carlo Methods in Practice*, Springer, New York, NY, USA, pp. 197–223, 2001, ISBN: 978-1-4419-2887-0. https://doi.org/10.1007/978-1-4757-3437-9_10.

[8] Hoffman, M. D., Gelman, A. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of*

Machine Learning Research, 15, No. 1, pp. 1593–1623, 2014.

[9] Liu, Y., Ott, M., Goyal, N., Du, J., et al. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692 [cs.CL]*, 2019. https://doi.org/10.48550/arXiv.1907.11692.

[10] Voelkle, M. C., Oud, J. H. L. Continuous time modelling with individually varying time intervals. *British Journal of Mathematical and Statistical Psychology*, 66, No. 1, pp. 103–126, 2013. https://doi.org/10.1111/j.2044-8317.2012.02043.x.

[11] Andrieu, C., Doucet, A., Holenstein, R. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, 72, No. 3, pp. 269–342, 2010. https://doi.org/10.1111/j.1467-9868.2009.00736.x.

[12] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, Minneapolis, MN, USA, pp. 4171–4186, 2019. https://doi.org/10.18653/v1/N19-1423.

[13] Verhagen, M., Gaizauskas, R., Schilder, F., et al. SemEval-2007 Task 15: TempEval temporal relation identification. *Proceedings of SemEval-2007*, Prague, Czech Republic, pp. 75–80, 2007. https://doi.org/10.3115/1621474.1621488.

[14] Strötgen, J., Gertz, M. HeidelTime: High quality rule-based extraction and normalization of temporal expressions. *Proceedings of the 5th International Workshop on Semantic Evaluation*, Uppsala, Sweden, pp. 321–324, 2010. https://aclanthology.org/S10-1071/ (Accessed Date: March 1, 2026).

[15] Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., Blackburn, J. The Pushshift Reddit dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, 14, No. 1, pp. 830–839, 2020. https://doi.org/10.1609/icwsm.v14i1.7347.

[16] De Choudhury, M., Kiciman, E. The language of social support in social media and its effect on suicidal ideation risk. *Proceedings of the International AAAI Conference on Web and Social Media*, 11, No. 1, pp. 32–41, 2017. https://doi.org/10.1609/icwsm.v11i1.14891.

[17] Hutto, C., Gilbert, E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8, No. 1, pp. 216–225, 2014. https://doi.org/10.1609/icwsm.v8i1.14550.

[18] Demszky, D., Movshovitz-Attias, D., Ko, J., et al. GoEmotions: A dataset of fine-grained emotions. *Proceedings of ACL*, Online, pp. 4040–4054, 2020. https://doi.org/10.18653/v1/2020.acl-main.372.

[19] Øksendal, B. Stochastic Differential Equations: An Introduction with Applications, 6th ed. Springer, Berlin, Heidelberg, Germany, 2003, ISBN: 978-3-540-04758-2. https://doi.org/10.1007/978-3-642-14394-6.

[20] Kalman, R. E. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, No. 1, pp. 35–45, 1960. https://doi.org/10.1115/1.3662552.

[21] Naskinova, I., Milev, M., Netov, N., Kolev, M., et al. Posterior bias of the drift estimator in a hierarchical Ornstein–Uhlenbeck model for adverse life events (companion paper). Proceedings of the International Conference on Mathematical Methods & Computational Techniques in Science & Engineering (MMCTSE 2026), Venice, Italy, 2026, in press.

[22] Benjamini, Y., Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57, No. 1, pp. 289–300, 1995. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Irina Naskinova, Mariyan Milev, Nikolay Netov, Angelina Angelova and Mikhail Kolev were responsible for the conceptualization of the study. Irina Naskinova, Mariyan Milev and Mikhail Kolev developed the methodology. Mariyan Milev, Hristo Kalinov, Gabriela Vasileva and Yana Vasileva implemented the software. Mariyan Milev, Mikhail Kolev, Gabriela Vasileva and Yana Vasileva carried out the formal analysis. Gabriela Vasileva and Hristo Kalinov carried out the investigation and the data curation. Gabriela Vasileva and Irina Naskinova wrote the original draft, and all authors contributed to the review and editing of the manuscript. Gabriela Vasileva and Yana Vasileva prepared the visualization. Irina Naskinova and Mikhail Kolev supervised the work, Irina Naskinova was responsible for the project administration, and Nikolay Netov, Irina Naskinova and Mikhail Kolev for the funding acquisition. All authors have read and agreed to the published version of the manuscript.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

The presentation and dissemination of these research results are supported by National Science Fund Project КП-06-Н85-7/05.12.2024 “Significance and Potential Risks of the Fast Integration of Artificial Intelligence (AI) Technologies into the Economy and Financial Sector”. No other funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.