

Article

Not peer-reviewed version

Fractional Process with Jumps for Karstic Spring Discharge

[Dániel Boros](#) , [Edit Borbás](#) , Amina Darougi , József Kovács , [László Márkus](#) *

Posted Date: 23 May 2025

doi: 10.20944/preprints202505.1827.v1

Keywords: fractal dimension; karstic aquifer; multifractals; spring discharges; time series; fractional ornstein-uhlenbeck process; jump process; renewal-reward process



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Fractional Process with Jumps for Modeling Karstic Spring Discharge Data

Dániel Boros ^{1,2} , Edit Borbás ¹ , Amina Darougi ^{1,2} , József Kovács ³ 
and László Márkus ^{1,2,*} 

¹ Department of Probability Theory and Statistics, Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary

² Doctoral School of Mathematics, Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary

³ Department of Physical and Applied Geology, Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary

* Correspondence: laszlo.markus@ttk.elte.hu; Tel.: +36-1-372-2500/ext. 8741 (L.M.)

Abstract: Fractal dimensions for the daily discharge data series of several karstic springs in Northeast Hungary have recently been computed and analyzed. We model four of those series, having similar fractal dimensions, using a superposition of a fractional Ornstein-Uhlenbeck process and a jump process of renewal-reward type. Beyond some usual goodness-of-fit measures, simulations of the model show an appealing visual fit. When the fractal dimension is not taken into account in the modeling, the simulated accumulated discharges tend to significantly exceed the realistic one.

Keywords: fractal dimension; karstic aquifer; multifractals; spring discharges; time series; fractional ornstein-uhlenbeck process; jump process; renewal-reward process

MSC: 60G22; 60G55; 60K15; 62M10; 86A05

1. Introduction

Karst spring discharge time series are known to exhibit highly irregular, heterogeneous, and multifractal dynamics [5,20]. These complex patterns reflect the underlying geologic heterogeneity, including the interaction between the porous matrix, fracture networks, and conduit systems [7]. Traditional linear hydrological models, which often rely on simplified assumptions of stationarity and Gaussian-like fluctuations, fail to capture the intricacies inherent in karst hydrology, such as long-range dependence, sudden jumps in flow rates, and the influence of extreme events [17].

The recession curves of karst spring hydrographs, frequently decomposed into distinct flow components (e.g., conduit-driven quick-flow and matrix-driven baseflow) serve as important indicators of aquifer properties, infiltration processes, and hydraulic connectivity [5]. Although conceptual models such as those proposed by Mangin [16] and Fiorillo [5] provide useful insights by distinguishing various storage and drainage compartments, they often lack the ability to represent the non-linear and fractal or multifractal nature of observed discharge records [16,20]. Empirical studies have shown that karstic systems display scale-invariant behavior: the complexity of flow paths and conduit geometries induces fractal patterns in spring discharge time series [23]. The fractal dimension D characterizes the degree of roughness and complexity on multiple scales [1,21], and thus serves as a crucial quantitative measure to understand karst processes [4,10].

Motivated by these findings, it is imperative to develop stochastic modeling frameworks capable of reproducing both the fractal scaling and the jump-like features observed in karst spring discharges. In this study, we employ a fractional Ornstein–Uhlenbeck process (fOUp) enriched with jump components to model and simulate karst spring discharge dynamics. The fOUp is a continuous-time Gaussian process with long-range dependence, characterized by a Hurst exponent H (with $0 < H < 1$), allowing it to capture persistent ($H > 0.5$) and slowly decaying autocorrelations or anti-persistent ($H < 0.5$) rough fluctuations inherent in karst systems [3].

The usual, mean-reverting Ornstein–Uhlenbeck process, (OU_p) $X(t)$ is defined by the stochastic differential equation (SDE):

$$dX(t) = \kappa(\theta - X(t)) dt + \sigma dB(t), \quad (1)$$

where $\kappa > 0$ is the mean-reversion parameter, θ is the long-term mean, $\sigma > 0$ is the volatility, and $B(t)$ is a classic Brownian motion. Extending the OU_p to a fractional context involves replacing $B(t)$ with a fractional Brownian motion $B_H(t)$ of Hurst index H . This introduces long-memory characteristics into the dynamics when $H > 0.5$ and rough fluctuation when $H < 0.5$.

The ultimate source for renewing the karstic hydrogeological system is infiltration from precipitation or from melting snow. To address sudden discharge spikes originating from infiltration, we incorporate a jump component into the driving force of the fractional Ornstein–Uhlenbeck (fOU) equation, resulting in a hybrid model that can be schematically written as:

$$dX(t) = \kappa(\theta - X(t)) dt + \sigma dB_H(t) + dJ(t), \quad (2)$$

where $J(t)$ represents the jump component. In general modeling practice, often the jump component, $J(t)$, follows a compound Poisson process. However, it has the shortcoming that the jump times and jump sizes are independent, and the process has the Markov property, enforcing exponential interarrival times (i.e., times between jumps, also called as sojourn times in renewal theory). This property is counterintuitive in our application, since a long delay between jumps – induced mainly by long delay in precipitation events – leaves the aquifer drained. Hence, the water from the next precipitation event – or a quick snowmelt for its turn – is kept back in the aquifer, filling up the porous matrix and the fracture networks, allowing for only lesser peaks in spring water discharge. Inspired by renewal theory and supported by statistical evidence (see Section 6), we suggest using a semi-Markov jump process for $J(t)$ with Weibull-distributed interarrival times and also Weibull-distributed peaks [24]. Contrary to Markov models, semi-Markov models allow interdependence of the process' state and the interarrival time, and essentially place no restriction on the interarrival time distribution, see, e.g. [24]. The proper correlation of the interarrival times and the jump sizes is secured by the probability integral transform and its inverse, along with the Gaussian copula. To the best of our knowledge, this model specification is novel in the literature. The jump component describes the abrupt short term, whereas the fractional component fine tunes the long-term behavior.

We apply this fractional jump Ornstein–Uhlenbeck framework to daily discharge data from karst springs in Northeast Hungary. The associated analysis illustrates the methodology's capacity to replicate both the statistical and scaling features of the observed time series. By comparing simulated and observed discharge data, we demonstrate that this modeling strategy can reproduce essential empirical characteristics, such as fractal scaling, non-Gaussian tails, and sudden discharge extremes. Thus, the jump-fOU_p approach provides new insights into the structure and complexity of karst aquifers and represents a promising avenue for understanding and predicting the non-linear, non-stationary, and fractal-like behavior of karstic spring discharges.

1.1. Outline of the Paper

The remainder of this article is organized as follows. Section 2 introduces the modelling framework—a fractional Ornstein–Uhlenbeck process augmented by a renewal–reward jump term—together with its mathematical properties and a short proof of well-posedness, which can be found in Appendix A. Section 3 reviews existing estimation techniques for the pure fOU process and adapts them into a two-tier procedure that later underpins our jump-fOU_p inference. Section 4 describes the geological setting, the 19-year daily discharge dataset from four Hungarian karst springs, and the exploratory analyses that motivate our model choices. Section 5 presents the complete estimation–simulation workflow: detection of surges, copula-based generation of correlated jump pairs, fGn simulation for the fractional kernel, and the iterative scheme that couples both components. Section 6 compares observed and simulated hydrographs, fractal dimensions, and cumulative discharges, demonstrating

that the jump-fOUp reproduces key statistical and visual features markedly better than short-memory alternatives. Section 7 discusses hydrological implications, the necessity of long-range dependence, and possible extensions such as time-varying parameters or precipitation-driven jump intensities. Finally, Section 8 concludes with the main findings and outlines directions for future research.

2. Fractional Ornstein–Uhlenbeck Process with Renewal Jumps

2.1. The Pure fOUp

The (mean reverting) fractional Ornstein–Uhlenbeck process (fOUp) extends the usual Ornstein–Uhlenbeck model by incorporating the fractional Brownian motion (fBm) as its driving noise term. To distinguish it from its jump-enriched counterpart, we refer to it as pure fOUp. The fractional Brownian motion generalizes the classic Brownian motion ($H = 0.5$) to incorporate long-range dependence or rough paths. If $H > 0.5$, the increments are positively correlated, leading to persistent behavior, i.e., long memory. If $H < 0.5$, the increments are negatively correlated, resulting in rough paths and antipersistent behavior. Let $X(t)$ be a continuous-time stochastic process satisfying the stochastic differential equation (SDE)

$$dX(t) = \kappa(\theta - X(t)) dt + \sigma dB_H(t), \quad (3)$$

where $\kappa > 0$ is the mean-reversion rate, θ is the long-term mean toward which $X(t)$ reverts, $\sigma > 0$ is the volatility parameter, $B_H(t)$ is a fractional Brownian motion with Hurst parameter $H \in (0, 1)$. The solution of the fOU equation exists and is unique up to initial condition $X(0)$. It can be given by [3]:

$$X(t) = e^{-\kappa t} X(0) + \theta(1 - e^{-\kappa t}) + \sigma \int_0^t e^{-\kappa(t-s)} dB_H(s). \quad (4)$$

For $H = 0.5$, the fOUp reduces to the usual OUp driven by a classic Brownian motion. The long-memory or rough path behavior is inherited by the fOUp from fBm according to the Hurst parameter. Unlike Bm and fBm the fOUp has a stationary regime. In contrast to the usual OUp, the fOUp is non-Markovian for $H \neq 0.5$. For more details on the fOUp see, e.g., [3].

2.2. Adding Jumps to the fOUp

We extend the fOUp dynamics by including jumps in the driving force to capture the long-memory or rough behavior of hydrological processes alongside sporadic, strictly positive surges that can be extreme in magnitude. The fOUp component encodes persistent correlations or rough fluctuations observed in hydrology. The jump components are responsible for the shocks that arrive from extreme water infiltration. In the simplest analogue, the jump diffusion setting, one might add a compound Poisson process whose interarrival times are exponentially distributed and whose jump sizes can in principle have arbitrary distributions. However, in hydrological systems where only positive, heavy-tailed surges are relevant and where jump sizes and interarrival times can exhibit dependence, the assumption of a compound Poisson structure becomes too restrictive. Our approach uses an analogy to renewal theory and incorporates an additive renewal-reward term into the fOU equation.

We enrich fOUp with jumps by first specifying a renewal-reward process $J(t)$. The process $J(t)$ is specified by a sequence of increasing jump times $T_k = \sum_{i=1}^k \tau_i$, $k = 1, 2, \dots$ with $T_0 = 0$. The interarrival times τ_i , $i = 1, 2, \dots$ are positive random variables, and the jump sizes S_i , $i = 1, 2, \dots$ are positive random variables. We set $S_0 = 0$. With these $J(t, \omega) = \sum_{i=0}^k S_i(\omega)$ if $T_k(\omega) < t \leq T_{k+1}(\omega)$. Both interarrival times τ_i and jump sizes S_i have a Weibull distribution in our application, and variables τ_i and S_i are correlated. In this way, $J(t)$ is a point process, with a nondecreasing, right-continuous, step-function sample path structure, for which stochastic integrals can be defined in terms of standard Lebesgue-Stieltjes integration theory. As a result, its differential $dJ(t)$ can also be viewed as a stochastic measure that puts jump size weights S_i on jump time points T_i . Now we define our jump-fOUp process $\tilde{X}(t)$ as:

$$\tilde{X}(t) = e^{-\kappa t} X(0) + \theta(1 - e^{-\kappa t}) + \sigma \int_0^t e^{-\kappa(t-s)} dB_H(s) + \int_0^t e^{-\kappa(t-s)} dJ(s). \quad (5)$$

Theorem 1. Let $\tilde{X}(t)$ be defined by equation (16). Then $\tilde{X}(t)$ satisfies the stochastic differential equation,

$$d\tilde{X}(t) = \kappa(\theta - \tilde{X}(t)) dt + \sigma dB_H(t) + dJ(t), \quad (6)$$

where, $\kappa > 0$ is the mean reversion rate, θ is the equilibrium level, $\sigma > 0$ is the volatility, $B_H(t)$ is a fractional Brownian motion (same as $W_H(t)$ used in appendix - standardized notation to $B_H(t)$), and $J(t)$ is the renewal-reward process defined above.

We leave the proof of the theorem in Appendix A.

The jump term $J(t)$ accumulates at random times governed by Weibull-distributed interarrivals and delivers jump magnitudes that are also Weibull-distributed. This construction ensures that each surge contributes a strictly positive, heavy-tailed (as the Weibull shape is invariably less than one in these applications, see Section 6) increment to $\tilde{X}(t)$, and that the negative correlation of interarrival times and jump sizes is preserved. The semi-Markov property emerges because the distribution of waiting times until the next jump depends on previously realized jump sizes (and vice versa), thereby departing from the memoryless exponential interarrivals of a compound Poisson process.

The flexibility of the model, resulting from the semi-Markov construction, allows the simultaneous reproduction of heavy-tailed surges and their further spreading effect and long-memory fluctuations. This resolves a common challenge in hydrological modeling, where only positive, subexponential jumps with significant extremal impact can occur. By merging Weibull-based jump processes with fractional diffusion, one captures rare yet sizeable excursions on top of a persistently correlated or anti-persistent baseline.

3. Parameter Estimation Methods

Much effort has been devoted to the parameter estimation problem for fOU processes. Assuming that a continuous record of observations is available in the non-mean-reverting case ($\theta = 0$), Kleptsina and Le Breton [14] obtained first the formula for the drift parameter's maximum likelihood (ML) estimation and proved almost sure convergence. Bercu, Courtin and Savy [2] proved a central limit theorem for the MLE in the case of $H > 1/2$. They claimed without proof that the above convergence is also valid for $H < 1/2$. Finally, Tanaka et al. [26] completed the case by studying the ML estimates when $\kappa > 0$ and $H \in (0, 1/2)$.

Turning to mean reversion with $\theta > 0$ and keeping the assumption of continuous observation, Lohvinenko and Ralchenko [19] obtained the ML estimates of κ and θ when $\kappa > 0$ and $H \in (1/2, 1)$.

Estimating the parameter vector $\Theta = (\kappa, \theta, \sigma, H)$ of the fOUp in the more realistic case of discrete, time-equidistant observations $\{X(t)\}_{t=1}^N$ poses significant challenges due to the fractional nature of the driving noise. In the non-mean-reverting case specialized statistical methods were developed in [12,13], using the relationship between the Skorokhod (or divergence) and Stratonovich integrals. These works also provide limit theorems for the estimation as both the time horizon and the mesh size grow to infinity simultaneously, and the horizon increases faster than the mesh size. However, this condition is difficult to verify in a real application with finite sample size.

3.1. A Two-Tier Estimation Framework for the fOU Parameters

This section consolidates, in a single presentation, the continuous-record methodology of [12,13] and its exact, discrete-sample analogue due to [29]. Both procedures target the parameter vector $\Theta = (\kappa, \theta, \sigma, H)$ of a fOUp

Tier I — continuous record ([13]).

Assuming full observation of the path $\{X_t\}_{0 \leq t \leq T}$, the mean-reversion speed κ can be consistently recovered by either *least squares*

$$\hat{\kappa}_{\text{LSE}} = - \frac{\int_0^T X(t) dX(t)}{\int_0^T X(t)^2 dt}, \quad (7)$$

or the *energy-type* statistic;

$$\hat{\kappa}_{\text{ET}} = \left(\frac{1}{\sigma^2 H \Gamma(2H) T} \int_0^T X(t)^2 dt \right)^{-1/(2H)}, \quad (8)$$

where Γ denotes Euler's gamma function. Both estimators are strongly consistent and asymptotically normal as $T \rightarrow \infty$.

Tier II — exact discrete record ([29]).

Let $X(k)$ be an equally spaced sample of size N with the grid step $\delta = T/N$, $k = 1, \dots, N$. Define second-order differences and their quadratic variations.

$$D_\delta(k) = X(k+2) - 2X(k+1) + X(k), \quad S_2 = \sum_{k=1}^{N-2} D_\delta(k)^2, \quad S_1 = \sum_{k=1}^{\lfloor (N-2)/2 \rfloor} D_{2\delta}(2k)^2.$$

Closed-form, strongly consistent estimators are then obtained as

$$\hat{H} = \frac{1}{2} \log_2 \left(\frac{S_1}{S_2} \right), \quad (9)$$

$$\hat{\sigma} = \sqrt{\frac{S_2}{(N-2)(4-2^{2\hat{H}})\delta^{2\hat{H}}}}, \quad (10)$$

$$\hat{\theta} = \frac{1}{N} \sum_{k=1}^N X(k), \quad (11)$$

$$\hat{\kappa} = \left(\frac{\frac{1}{N} \sum_{k=1}^N (X(k) - \hat{\theta})^2}{\hat{\sigma}^2 \hat{H} \Gamma(2\hat{H})} \right)^{-1/(2\hat{H})}. \quad (12)$$

4. Application to Karstic Spring Discharge Modeling

4.1. Geological and Hydrological Settings

The Aggtelek Karst, part of the Gömör-Torna Karst region on the Hungarian-Slovakian border, is predominantly composed of Triassic carbonates, mainly Wetterstein limestone and dolomites [27]. The Hungarian section comprises approximately 200 km² of east-west-oriented karst plateaus at elevations of 400–600 m a.s.l., dissected by valleys as low as 150–260 m. Karstification began in the Miocene under subtropical conditions, with significant Pliocene uplift that forms the present-day north-south topographic gradient [9,11]. The mean annual precipitation is about 600 mm, most of which infiltrates the plateau and resurfaces at karst springs. The area has a humid continental climate with a mean annual temperature of 9.1 °C, about 120–130 frost days, and 40 snow-covered days annually [27].

These geological and climatic data originate from Telbisz et al. [27]. The dataset, and descriptions therein are based on decades of multidisciplinary fieldwork and geological mapping [8,18], and are considered reliable and high-quality within the regional geoscientific community.

4.2. Data Description

The measurement of the daily spring discharge of the most significant karstic springs in the Aggtelek Karst was carried out by the Hungarian Water Resources Research Plc. (VITUKI) from the early 1960s to the 1990s [22]. We selected four springs (Csörgő, Kecsekút, Komlós, Lófej) with non-missing daily observations for 6200 days in the period from 1 March 1974 to 18 February 1993. The basic descriptive statistics of these decades of uninterrupted hydrological data series are shown in Table 1.

Table 1. Characteristics of the four selected springs: Aquifer size, daily average discharge (Q), ratio of maximum to median discharge (Q_{max}/Q_{med}), and water temperature statistics.

Spring		Aquifer Size (km ²)	Discharge (Q)		Water Temperature (°C)		
Name	Code		Mean (m ³ /day)	Q_{max}/Q_{med}	Min	Max	Mean
Csörgő	Cso	2.7	1121.4	61.2	8.6	11.8	11.1
Jósva	Jos	34.4	14508.5	28.9	7.5	14.7	10.1
Kecsekút	Kec	0.8	286.5	45.0	8.4	10.6	9.5
Komlós	Kom	2.5	846.3	90.0	9.3	12.1	10.3

The springs are located within a relatively small area, close to each other (Figure 1). However, significant scale differences can be observed in their daily discharges: the multiannual averages vary between 287 and 14508.5 m³/day, see Table 1. Cave systems of varying lengths, are known to be connected with the springs [15,25]. In Jósva spring water comes to the surface from the 25 km long Baradla-Domica cave system. Komlós spring (Kom) is connected to Béke cave (approx. 7.2 km long). Szabadság cave (approx. 3.2 km) belongs to the aquifer of Kecsekút spring (Kec). Caves of much shorter confirmed lengths – typically a few hundred meters, as cave researchers put it to current knowledge - lie behind the Csörgő spring (Cso) [15]. The cave systems direct quick flows to the mouth of the springs, taking essential role in shaping the surges in the hydrographs.

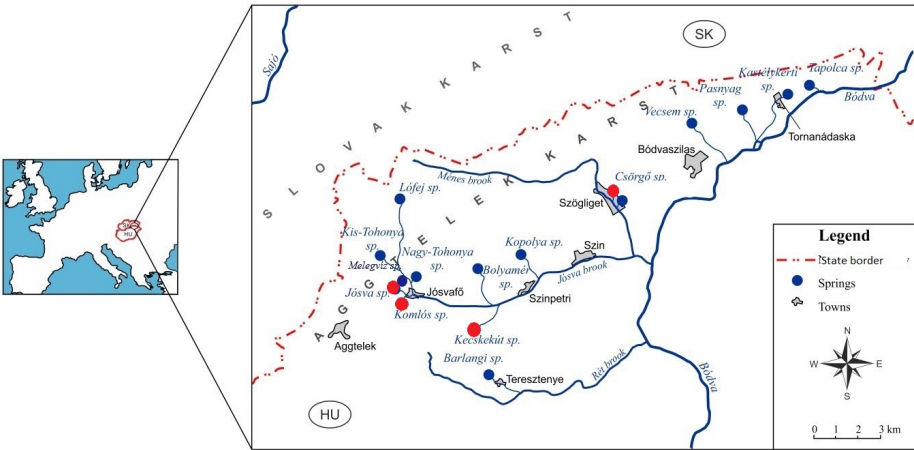


Figure 1. A schematic map displaying the location of the four selected springs in the Aggtelek Karst region, NE Hungary.

5. Methodology: Parameter Estimation and Simulation

The objective is to model the daily discharge time series $X(t)$ of several karst springs. The model assumes that the discharge dynamics results from the superposition of a continuous, persistent background process and discrete, sudden jump events, often triggered by precipitation. A jump fOUp as described in section 2 is employed. The analysis proceeds spring by spring; the following sections outline the methodology for a single representative time series $X(t)$, $t = 1, \dots, N$. The modeling process involves two main phases: estimating model parameters from the observed data and simulating the process using these parameters.

5.1. Phase 1: Parameter Estimation from Observed Data

5.1.1. Data Preparation and Initial Analysis

Let $X(t)$ denote the observed discharge at day t . The differenced series, $\Delta X(t) = X(t) - X(t-1)$ for $t = 2, \dots, N$, is calculated to highlight rapid changes or jumps (Algorithm A1, line 1).

5.1.2. Jump Component Identification and Characterization

- **Jump Detection:** Jumps are identified where $\Delta X(t)$ exceeds a high quantile threshold, $q_{\Delta X} = \text{quantile}(\{\Delta X(t)\}, p)$, where p is a high probability, chosen as 0.975 (Algorithm A1, line 2). Let T_j be the time index of the j -th detected jump, $j = 1, \dots, M$ (Algorithm A1, line 3). (For heuristic explanation see Section 6.1).
- **Jump Size Analysis:** The jump sizes are $J_j = \Delta X(T_j) - q_{\Delta X}$ (Algorithm A1, line 4). Their distribution is analyzed by fitting candidate distributions (Weibull and Gamma) to J_j using Maximum Likelihood Estimation (MLE) (Algorithm A1, lines 6). Goodness-of-fit is assessed using Kolmogorov-Smirnov (KS) tests. Let $\hat{k}_J, \hat{\lambda}_J$ denote the estimated jump size parameters.
- **Jump Timing Analysis:** The inter-arrival times $\tau_j = T_j - T_{j-1}$ (with $T_0 = 0$) are calculated (Algorithm A1, line 5). Their distribution is analyzed similarly, fitting a Weibull distribution to a slightly perturbed data, via MLE (Algorithm A1, line 8). Perturbation is necessary because being a continuous distribution, integers and ties do not fit the Weibull. Perturbation is created by adding a Uniform[0.01,0.5] variable to the jump sizes. The effect of the perturbation is then analyzed using 100 perturbations. Let $\hat{k}_\tau, \hat{\lambda}_\tau$ denote the estimated jump time parameters.
- **Jump Dependence Structure:** The dependence between jump sizes J_j and preceding inter-arrival times τ_j is quantified by the sample correlation coefficient $\hat{\rho} = \text{Corr}(J_j, \tau_j)$ (Algorithm A1, line 9). Typically, $\hat{\rho}$ is significantly less than 0.

5.1.3. Background Process Characterization

- **Hurst Parameter Estimation:** The long-range dependence is quantified by the Hurst parameter $H \in (0, 1)$. For the background process, assuming the fOUp structure, it is estimated from (9) (Algorithm A1, line 10). However, for the complete $X(t)$ process with jumps – when (9) is not applicable – it is estimated using the madogram method via the fractal dimension: $\hat{H} = 2 - \hat{D}_{\text{madogram}}$. It also serves to remain compatible with the findings of [4].
- **Fractional Ornstein-Uhlenbeck (fOU) Parameter Estimation:** The background process is modeled as an fOUp, and its parameters are estimated from a truncated series $X'(t)$ (Algorithm A1, lines 11-12, see also section 6).
- **Volatility (σ):** Estimated from the second differences of $X'(t)$ using $\hat{H}$:

$$\hat{\sigma}^2 \approx \frac{\text{Var}(\Delta^2 X'(t))}{4 - 2^{2\hat{H}}}. \quad (13)$$

(Algorithm A1, line 13).

- **Mean Reversion Rate (κ):** Estimated from the variance of $X'(t)$ and $\hat{\sigma}$:

$$\hat{\kappa} \approx \left(\frac{\text{Var}(X'(t))}{\hat{\sigma}^2 \hat{H} \Gamma(2\hat{H})} \right)^{-1/(2\hat{H})}. \quad (14)$$

(Algorithm A1, line 14).

5.2. Phase 2: Simulation Setup

- **Jump Simulation via Gaussian Copula:** To preserve the correlation $\hat{\rho}$ between jump times and jump sizes, a Gaussian copula is used. To create a large sample from the Gaussian copula, $L \gg M$ pairs of correlated standard normal variables $(Z_{1,i}, Z_{2,i})$ with correlation ρ_G are generated and transformed to uniform (U_i, V_i) via the normal CDF $\Phi(\cdot)$ (Algorithm A1, line 15-16). Simulated

inter-arrival times τ'_i and raw jump sizes J'_i are generated using the inverse CDFs of the fitted Weibull distributions (Algorithm A1, lines 20-29). Since this transformation does not preserve the correlation, ρ_G is chosen slightly larger than $\hat{\rho}$, the target one. Subsampling is then performed on these L pairs until we obtain a final sequence of $M_{sim} \approx M$ pairs $(\tau_{sim,j}, J_{sim,j})$, so many that jump times cover the $[0,T]$ time interval, and their correlation reaches the target $\hat{\rho}$ (Algorithm A1, lines 19-25). An innovation array $J^*(t)$ of length N is built by filling it up with the final jump increments $J_{sim,j} + q_{\Delta X}$ at the simulated jump times $T_{sim,j} = \sum_{k=1}^j \tau_{sim,k}$, and with zeros elsewhere. (Algorithm A1, lines 31-34).

- **Fractional Gaussian Noise (fGn) Simulation:** A sequence of fGn increments $\epsilon(t)$ with Hurst parameter $\hat{H}$ is generated by differencing a simulated fBm (e.g., by using the somebm package of R) and scaled using the estimated volatility $\hat{\sigma}$ (Algorithm A1, lines 36-37).

5.3. Phase 3: Combined Process Simulation Loop

Writing $e^{-\kappa t}$ as $e^{-\kappa} \cdot e^{-\kappa(t-1)}$ and integrating in (16) first from 0 to $t - 1$ and then from $t - 1$ to t we obtain

$$\tilde{X}(t) = e^{-\kappa} X(t - 1) + \theta(1 - e^{-\kappa}) + \sigma \int_{t-1}^t e^{-\kappa(t-s)} dB_H(s) + e^{-\kappa t} \int_{t-1}^t e^{\kappa s} dJ(s).$$

(15)

The last integral is $J(T_n)$ if $t = T_n$ for a jump time T_n , or zero otherwise. Since κ is small in our case, we obtain an approximation

$$\tilde{X}(t) \approx e^{-\kappa} X(t - 1) + \theta(1 - e^{-\kappa}) + \sigma \Delta B_H(t) + J(T_n) \cdot \chi_{\{t=T_n\}}.$$

(16)

Using the approximation the final discharge series $\tilde{X}(t)$ is generated iteratively for $t = 1, \dots, N$:

$$\tilde{X}(t) = e^{-\hat{\kappa}} \cdot \tilde{X}(t - 1) + \theta(1 - e^{-\kappa}) + \text{Innov}(t)$$

(17)

where $\text{Innov}(t) = \sigma \cdot \epsilon(t) + J^*(t)$ is the total innovation ($\sigma \cdot$ fGn + jump), and $\tilde{X}(0)$ is the initial value (Algorithm A1, lines 38-42):

This procedure generates a synthetic time series $\tilde{X}(t)$ to replicate the statistical properties of the observed discharge.

6. Application: Jump Distribution Fitting and fOUp Parameter Estimation for Karstic Spring Discharges

6.1. Separating the Jumps

The surges in spring discharges are not merely isolated spikes; rather, they exert a dampening effect on the recession curve. Consequently, it is not sufficient to model these as simple jumps in the process itself — they must be incorporated into the driving force of the dynamics, specifically in the differential component of the stochastic differential equation (SDE). Therefore, to locate the jump positions and sizes, the differenced series has to be considered. The quantiles and maxima of the differenced spring discharge series are presented in Table 2.

Table 2. Quantiles and maxima of differenced spring discharge series

Spring	0.9	0.95	0.975	0.99	Max	Max/q95	Max/q97.5
Csörgő	138	338.1	692.4	1700.48	28666	84.79	41.40
Jósva	288	687.6	1440	7200.00	313990	456.65	218.05
Kecskekút	9	22.0	54.0	252.16	3726	169.36	69.00
Komlós	108	357.2	634.05	1066.28	6698	18.75	10.56

The comparison of quantiles, maxima, and their relative proportions in Table 2 reveals the heavy-tailed nature of the underlying distribution. This observation rules out the adequacy of a pure fOUp

for modeling. Replacing fractional Brownian motion (fBm) with a Lévy flight in the SDE could address heavy-tail behavior, but it would fail to capture the correlation between extreme values and their timing. This limitation motivated the development of our proposed approach.

To corroborate the jump threshold used, Table 2 lists the empirical 0.95, 0.975 and 0.99 quantiles. Across all springs, the largest observed surge is roughly 20 to 450 times the 0.95-quantile and 10 to 200 times the 0.975-quantile, while the 0.975-quantile is only 2 to 3 times the 0.95-quantile. So, choosing the 0.975-quantile level as the cut-off effectively separates the extreme outliers, interpreted as hydrogeological ‘jumps’, from the long-memory background. 155 values are observed above the 0.975-quantile level, and it is still suitable to fit a distribution to them with sufficient accuracy. Hence, we choose this threshold to determine the jumps.

6.2. Jump Distributions and Waiting–Time Dependence

We analyze karst spring discharges through two complementary layers. First, we model the jump component by semi-Markov dynamics, whose inter-arrival times and magnitudes follow Weibull laws. Second, after excising these jumps, the residual series is represented by a fractional Ornstein–Uhlenbeck baseline whose parameters are inferred from second–order differences as set out in Section 3. All statistical decisions (choice of distribution, selection of the jump threshold, and verification of long–memory parameters) are supported by formal goodness-of-fit tests and by the scaling diagnostics summarized below.

6.3. fOU Background: Continuous Versus Truncated Records

After removing every surge that exceeds the 97.5th empirical quantile, a fractional–Ornstein–Uhlenbeck (fOU) model is fitted to the resulting *jump-filtered series*. Table 3 lists the estimated volatility $\hat{\sigma}$, damping coefficient $\hat{\kappa}$, memory scale $e^{-\hat{\kappa}}$ and Hurst exponent $\hat{H}$.

Table 3. Fractional OU parameter estimates for the jump-filtered baseline.

Spring	$\hat{\sigma}$	$\hat{\kappa}$	$e^{-\hat{\kappa}}$	$\hat{H}$
Csörgő	298	0.070	0.932	0.837
Jósza	758	0.076	0.927	0.675
Kecskekút	50	0.052	0.949	0.907
Komlós	102	0.061	0.941	0.848

Across all springs $\hat{\kappa}$ is small, so the background discharge reverts slowly to its mean; the associated memory $e^{-\hat{\kappa}} > 0.93$ indicates pronounced persistence, while the conditional variance is far lower than in the raw data. The Hurst exponents fall between 0.675 and 0.907, confirming long-range dependence in the smooth baseline component.

6.4. Scaling Check via Fractal Dimension

To verify that the combined model—semi-Markov jumps plus fOU baseline—captures the multi-scale roughness of the data, the fractal dimension D of each observed record is compared with the mean D from 100 simulated paths (Table 4). The close match confirms that the model reproduces the measured texture across scales.

Table 4. Fractal dimension (D) and implied Hurst exponent ($H = 2 - D$) for observed and simulated discharge records (mean of 100 simulations).

Spring	Observed		Simulated (mean)	
	D	H	D	H
Csörgő	1.309	0.691	1.263	0.737
Jósva	1.462	0.538	1.395	0.605
Kecskékút	1.313	0.687	1.275	0.725
Komlós	1.289	0.711	1.257	0.743

The $H = 2 - D$ values in Table 4 describe the full jump + fOU process, so they are systematically lower than the background $\hat{H}$ in Table 3: high-frequency jump spikes shorten the effective memory, whereas the jump-filtered baseline retains the longer-range dependence quantified in the first subsection.

Figure 2 displays the simulated and observed hydrographs of the four springs and demonstrates that simulations, based on the fitted model, exhibit strong visual agreement with the observed time series. This directly addresses a common criticism from hydrological and hydrogeological experts, who often contend that mathematical models—despite achieving good statistical fits in certain aspects—fail to produce realistic-looking hydrographs.

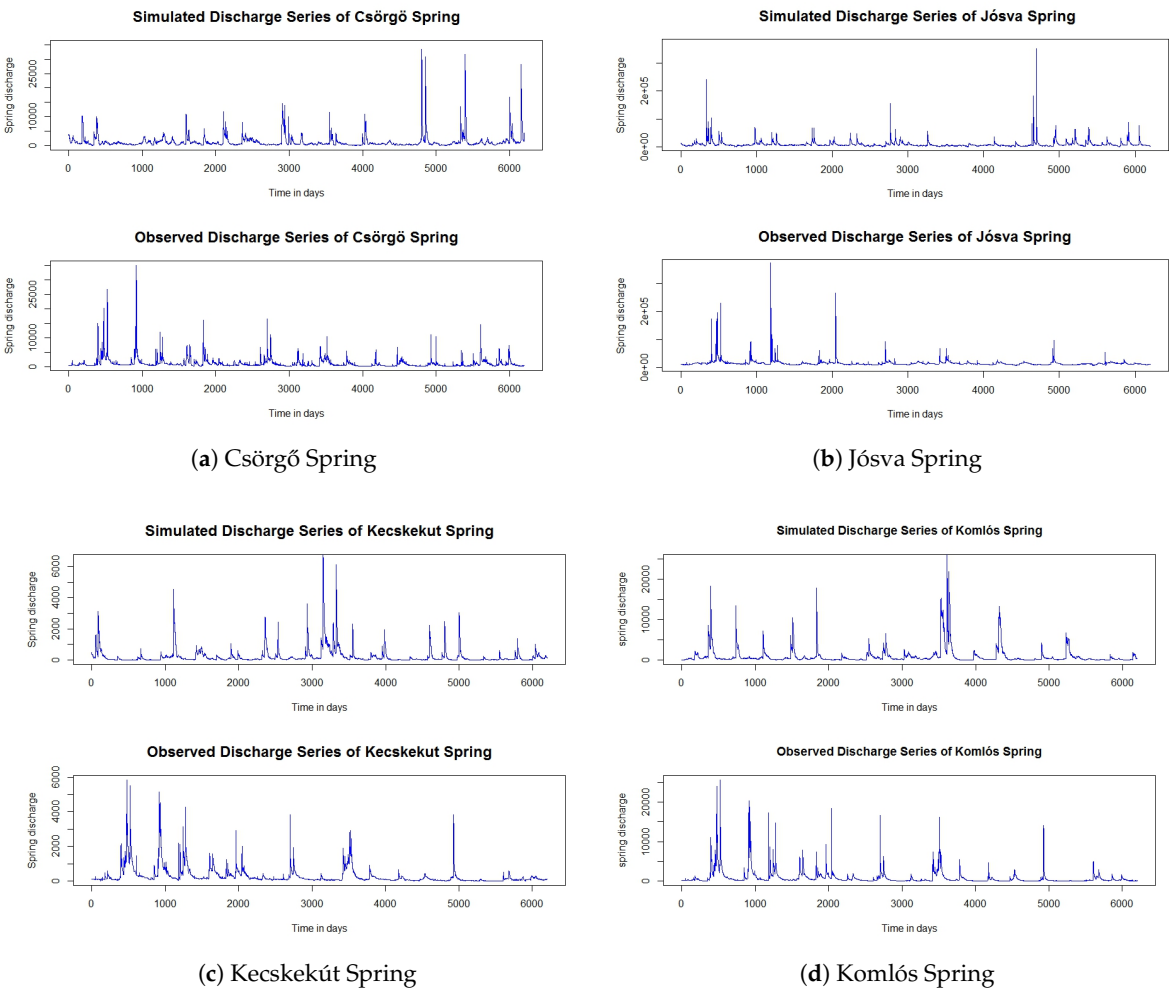


Figure 2. Simulated and observed hydrographs of Csörgő, Jósva, Kecskékút, and Komlós Springs.

From a water resources management perspective, cumulative discharge is of paramount importance, as it directly reflects the volume of water available for various uses over a given period.

Therefore, its accurate estimation is a key objective in hydrological modeling. Figure 3 demonstrates the performance of our model in this regard. To highlight the critical role of long-range dependence, we replace the background flow component with an autoregressive (AR) process, while retaining the same jump sequence used in the jump-fOUp simulation. The outcome is striking: the 100 cumulative discharge trajectories generated by the AR model (blue lines) significantly overshoot the observed discharge (red line), in some cases more than doubling it by the end of the period. In contrast, simulations based on the fOUp model closely track the empirical data, underscoring the model’s accuracy. Moreover, the variability among the jump-AR simulations is substantially higher than that of the jump-fOUp simulations, reflecting the latter’s smoother trajectories. These results illustrate that the jump-fOUp model yields more reliable cumulative discharge estimates, enhancing its utility for water resource planning and management.

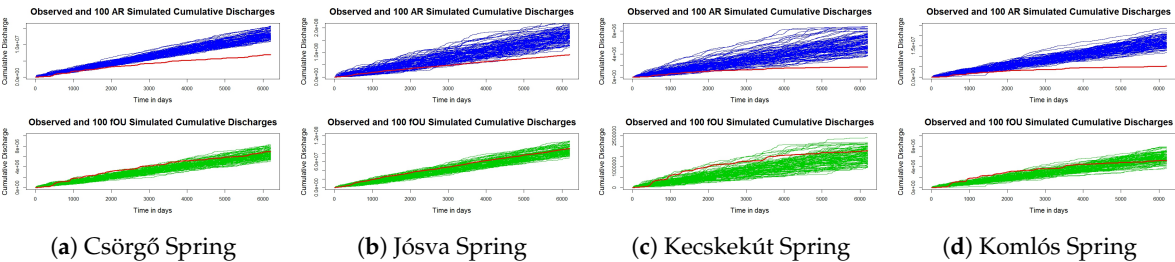


Figure 3. Accumulated Discharge Comparison: Observed (red line) vs. Jump-fOUp Model (100 simulations, green lines) and Jump-AR Model (100 simulations, blue lines) for Csörgő, Jósza, Kecskékút, and Komlós Springs.

7. Discussion

The jump-fOUp model integrates within a single stochastic framework the two primary controls of karst discharge: the *long-range dependent* background flow—comprising multiple hydrogeological components—captured via the fractional Ornstein-Uhlenbeck process, and the *heavy-tailed recharge surges*, modeled through a renewal-reward jump process with correlated, heavy-tailed inter-arrival times and magnitudes.

Figure 2 illustrates that simulated and observed hydrographs exhibit strong visual agreement, while Table 4 confirms that the model replicates the observed fractal dimension within the limits of sampling error. This demonstrates that the jump-fOUp model provides a generative mechanism capable of faithfully *reproducing* these key features.

In comparison to short-memory AR baselines [16], the jump-fOUp model offers significantly improved tracking of cumulative discharge volumes (Figure 3). This highlights that omitting long-range dependence—governed by the Hurst parameter H —can lead to potentially flawed assessments of water resources. When the background flow, represented by the truncated process, is instead modeled with a short-memory AR(1) process, a single AR parameter must account for all temporal dependencies. This results in systematically higher AR parameter estimates compared to the memory parameter in the fOUp model. Moreover, because this parameter acts not only on the driving noise but also on the jumps, it dampens the jump effects more slowly than in the fOUp case, thereby forecasting elevated discharges in the following days. The cumulative result—total discharged water volume—becomes unrealistically high, yielding an overly optimistic outlook for water resource management.

The improved accuracy in cumulative discharge estimation provided by our model arises from two key components:

- (i) a fractional kernel that propagates hydrological memory across months to years, and
- (ii) Weibull-Weibull jump pairs, whose negative correlation between size and waiting time reflects the natural depletion-recharge dynamics of the aquifer (Table 5).

Table 5. Weibull vs. Gamma fits for jump sizes and the mean of Weibull parameters fitted to the observed jump times with 100 uniform $U[0.01,0.5]$ perturbations, as described in 5.1.2. K-S p is the actual Kolmogorov-Smirnov p -value belonging to the mean parameters. M K-S p is the mean of the 100 K-S p values obtained from the perturbations. “Corr.” reports the Pearson correlation between jump size and waiting time in in the observed record.

Spring	Jump Size (Weibull)			Jump Size (Gamma)			Jump Time (Weibull)				Corr.
	Shape	Scale	K-S p	Shape	Scale	K-S p	Shape	Scale	K-S p	M K-S p	Obs.
Csörgő	0.685	1215.6	0.473	1.055	1159.9	0.0113	0.424	19.4	0.2806	0.273	−0.237
Jósva	0.475	7499.6	0.551	0.876	927.9	0.0002	0.425	16.4	0.4388	0.242	−0.219
Kecskékút	0.504	349.1	0.256	1.129	743.7	0.0019	0.375	13.1	0.0528	0.074	−0.303
Komlós	0.652	798.1	0.431	0.856	923.4	0.0036	0.423	15.3	0.0586	0.067	−0.268

Unlike Poisson jump-diffusions [28] or Lévy flights, the semi-Markov specification preserves both the heavy tail and the empirically observed dependence between surge size and lead time, a combination that classical multifractal surrogates describe but cannot generate mechanistically.

Parameter identification remains demanding, and further research into (e.g., iterative, likelihood-based, or Bayesian approaches for discretely observed jump-fOU processes would be beneficial. However, the current method yields parameters that produce simulations consistent with key empirical observations, including the crucial fractal dimension. Finally, extensions like time-varying κ or precipitation-driven jump intensities—offer an avenue to embed climate signals directly into the discharge generator.

8. Conclusions

In the presented analysis a jump-fOU with Weibull–Weibull semi-Markov surges reproduces the amplitude distribution, fractal dimension and cumulative discharge of four Hungarian karst springs. Long-range dependence ($H > 0.5$) is indispensable: replacing the fractional kernel with short-memory noise inflates the accumulated volume by up to 50 %. Surge size and waiting time are negatively correlated; modelling them jointly is crucial for realistic flood peaks.

The framework therefore bridges descriptive multifractal analysis and process-based simulation, and can be extended with time-varying parameters or rainfall covariates for predictive use.

Author Contributions: Conceptualization, L.M. and J.K.; Methodology, L.M. D.B. and E.B.; Software, D.B.; Validation, D.B., E.B. and A.D.; Formal Analysis, D.B. and L.M.; Investigation, E.B. and A.D.; Resources, J.K.; Data Curation, E.B. and J.K.; Writing—Original Draft Preparation, D.B. and L.M.; Writing—Review and Editing, L.M., J.K., D.B., E.B. and A.D.; Visualization, D.B. and L.M.; Supervision, L.M.; Project Administration, L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study involved the analysis of publicly available historical hydrological data and did not involve human subjects or live animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The daily discharge data for the Aggtelek karst springs used in this study were originally collected by the Hungarian Water Resources Research Plc. (VITUKI) and are described in [22]. The processed data used for the analysis are available from the corresponding author upon reasonable request.

Acknowledgments: We would like to thank the National Research, Development and Innovation Office for the support within the framework of the Thematic Excellence Program 2021 – National Research Sub programme: “Artificial intelligence, Large Networks, data security: mathematical foundation and applications” and the Artificial Intelligence National Laboratory Program (MILAB). We appreciate the support provided by neptune.ai. We are also incredibly thankful for the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, that took part in financing this research under the Eötvös Loránd University TKP 2021- NKTA-62 funding scheme.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AIC	Akaike Information Criterion
AR	Autoregressive
ARMA	Autoregressive Moving Average
Bm	Brownian motion
CDF	Cumulative Distribution Function
ET	Energy-Type (statistic)
fBm	Fractional Brownian motion
fGn	Fractional Gaussian noise
FD	Fractal Dimension
fOU	Fractional Ornstein–Uhlenbeck
fOUp	Fractional Ornstein–Uhlenbeck process
Jump-fOUp	Jump Fractional Ornstein–Uhlenbeck process
K–S	Kolmogorov–Smirnov
LF	Linear Filter
LSE	Least Squares Estimator
MLE	Maximum Likelihood Estimation
OUp	Ornstein–Uhlenbeck process
PDF	Probability Density Function
SD	Second Difference
SDE	Stochastic Differential Equation
VITUKI	Vízgazdálkodási Tudományos Kutató Intézet (Water Resources Research Institute)

Appendix A. Proof of Theorem 1

Problem Setup

We are given a stochastic process $\tilde{X}(t)$ defined in (16) as:

$$\tilde{X}(t) = e^{-\kappa t} X(0) + \theta(1 - e^{-\kappa t}) + \sigma \int_0^t e^{-\kappa(t-s)} dB_H(s) + \int_0^t e^{-\kappa(t-s)} dJ(s)$$

Here $(B_H(t))_{t \geq 0}$ is a fractional Brownian motion (fBm) with Hurst parameter $H \in (0, 1)$, $(J(t))_{t \geq 0}$ is a càdlàg, non-decreasing **renewal-reward process** (implying J has only finitely many jumps on any compact time interval $[0, T]$), and $X(0)$ is the initial condition, independent of B_H and J . The constants $\kappa > 0, \sigma > 0, \theta \in \mathbb{R}$ are given.

The integrals in (16) are understood as follows:

The term $I_H(t) := \sigma \int_0^t e^{-\kappa(t-s)} dB_H(s)$ is a **pathwise stochastic integral** with respect to fBm.

- If $H > 1/2$, B_H has sample paths that are almost surely Hölder continuous of any order $\delta < H$. Since δ can be chosen such that $\delta > 1/2$, and the kernel $s \mapsto k(t, s) := e^{-\kappa(t-s)}$ is Lipschitz continuous (hence, of bounded 1-variation) on $[0, t]$, the integral exists pathwise as a **Young integral** [30]. The condition for existence is $1/p + 1/q > 1$ where p is the variation exponent of the kernel and q is the variation exponent of the integrator. Here $p = 1$ (bounded variation) and B_H has finite q -variation for any $q \geq 1/H$. So $1/1 + 1/q > 1$ requires $1/q > 0$, which is always true. Thus, the Young integral exists.
- If $H \leq 1/2$, the paths of B_H are rougher (infinite q -variation for $q < 1/H$). Although the standard Young theory might not directly apply based solely on Hölder exponents, the integral is still well defined pathwise using extended frameworks like **fractional calculus integration** developed by Zähle [31]. This approach defines $\int f dg$ when f possesses a certain fractional smoothness related

to the roughness of g . Since our kernel $k(t, s) = e^{-\kappa(t-s)}$ is C^∞ in s , the necessary conditions are met for all $H \in (0, 1)$.

We denote this integral simply by $\int k(t, s) dB_H(s)$, understanding its pathwise nature (consistent with frameworks like Rough Paths [6]).

The term $I_J(t) := \int_0^t e^{-\kappa(t-s)} dJ(s)$ is a **Riemann-Stieltjes integral**. Since $J(s)$ is a renewal-reward process, its sample paths are càdlàg, non-decreasing, and piecewise constant (step functions). Consequently, $J(s)$ has almost surely **bounded variation** on any compact interval $[0, t]$. The kernel $s \mapsto e^{-\kappa(t-s)}$ is continuous on $[0, t]$. Therefore, the classical Riemann-Stieltjes integral exists pathwise for each realization ω . It can be explicitly written as a finite sum over the jump times T_n of J :

$$\int_0^t e^{-\kappa(t-s)} dJ(s) = \sum_{0 < T_n \leq t} e^{-\kappa(t-T_n)} \Delta J(T_n),$$

where $\Delta J(T_n) = J(T_n) - J(T_n - 1)$ is the jump size S_n at time T_n .

Our objective is to rigorously prove, using pathwise differentiation techniques, that $\tilde{X}(t)$ defined by (16) satisfies the stochastic differential equation (SDE):

$$d\tilde{X}(t) = \kappa(\theta - \tilde{X}(t)) dt + \sigma dB_H(t) + dJ(t).$$

The term $\sigma dB_H(t)$ should be interpreted in the context of the pathwise integral theory, and $dJ(t)$ represents the increments of the jump process.

The arguments above establish that both stochastic integrals in the definition (16) of $\tilde{X}(t)$ are well defined pathwise for almost every realization ω , for all $H \in (0, 1)$. The integral $I_H(t)$ relies on the Young/Zähle theory, while $I_J(t)$ uses the classical Riemann-Stieltjes theory for integrators of finite variation.

Differentiating the Terms of $\tilde{X}(t)$

We differentiate $\tilde{X}(t)$ with respect to t by considering each term in (16). We employ appropriate Leibniz rules for pathwise Stieltjes integrals. Let $\tilde{X}(t) = I_D(t) + I_H(t) + I_J(t)$, where $I_D(t)$, $I_H(t)$, $I_J(t)$ are the deterministic, fractional, and jump parts respectively.

Deterministic Terms

The first two terms $I_D(t) = e^{-\kappa t} X(0) + \theta(1 - e^{-\kappa t})$ are deterministic and smooth. Therefore,

$$dI_D(t) = \left[\frac{d}{dt} (e^{-\kappa t} X(0)) + \frac{d}{dt} (\theta(1 - e^{-\kappa t})) \right] dt = \kappa e^{-\kappa t} (\theta - X(0)) dt. \quad (A1)$$

Fractional Brownian Motion Term

Let $I_H(t) = \sigma \int_0^t e^{-\kappa(t-s)} dB_H(s)$. We need to compute $dI_H(t)$. Since the integral is pathwise (Young/Zähle), we can apply a suitable Leibniz rule. For an integral $Y(t) = \int_a^t k(t, s) dg(s)$, where k is sufficiently smooth in t and the integral exists pathwise, the rule is formally:

$$dY(t) = k(t, t) dg(t) + \left(\int_a^t \frac{\partial k}{\partial t}(t, s) dg(s) \right) dt. \quad (A2)$$

This rule is rigorously justified in the context of Young integration [30] and its extensions [6, 31] when $g = B_H$.

We apply the rule to $I_H(t)/\sigma = \int_0^t e^{-\kappa(t-s)} dB_H(s)$. Here, $k(t,s) = e^{-\kappa(t-s)}$. Thus $k(t,t) = 1$ and $\frac{\partial k}{\partial t}(t,s) = -\kappa e^{-\kappa(t-s)}$. We have:

$$d\left(\int_0^t e^{-\kappa(t-s)} dB_H(s)\right) = 1 \cdot dB_H(t) + \left(\int_0^t (-\kappa) e^{-\kappa(t-s)} dB_H(s)\right) dt.$$

Multiplying by σ we arrive at:

$$\begin{aligned} dI_H(t) &= \sigma dB_H(t) - \kappa \left(\sigma \int_0^t e^{-\kappa(t-s)} dB_H(s) \right) dt \\ &= \sigma dB_H(t) - \kappa I_H(t) dt. \end{aligned} \quad (\text{A3})$$

Jump Term

Let $I_J(t) = \int_0^t e^{-\kappa(t-s)} dJ(s)$. Since $J(s)$ is a step function with bounded variation and $k(t,s) = e^{-\kappa(t-s)}$ is smooth, we can apply the Leibniz rule (A2) for Stieltjes integrals with $dJ(s)$ replacing $dg(s)$. Using $k(t,t) = 1$ and $\frac{\partial k}{\partial t}(t,s) = -\kappa e^{-\kappa(t-s)}$ again:

$$\begin{aligned} dI_J(t) &= dJ(t) + \left(\int_0^t (-\kappa) e^{-\kappa(t-s)} dJ(s) \right) dt \\ &= dJ(t) - \kappa \left(\int_0^t e^{-\kappa(t-s)} dJ(s) \right) dt \\ &= dJ(t) - \kappa I_J(t) dt. \end{aligned} \quad (\text{A4})$$

Here, $dJ(t)$ represents the Stieltjes differential of J . If t is not a jump time, $dJ(t) = 0$. If $t = T_n$ is a jump time, $dJ(t)$ corresponds to a measure with mass $\Delta J(T_n)$ at T_n . This formula correctly captures both the continuous evolution between jumps (where only the dt term is active) and the discrete jumps (where $dJ(t)$ is active and equals $\Delta J(t)$).

Combining the Differentials

Now we sum the differentials of all parts of $\tilde{X}(t)$:

$$d\tilde{X}(t) = dI_D(t) + dI_H(t) + dI_J(t).$$

Substituting from (A1), (A3), and (A4):

$$\begin{aligned} d\tilde{X}(t) &= \left[\kappa e^{-\kappa t} (\theta - X(0)) \right] dt + \left[\sigma dB_H(t) - \kappa I_H(t) dt \right] + \left[dJ(t) - \kappa I_J(t) dt \right] \\ &= \left[\kappa e^{-\kappa t} (\theta - X(0)) - \kappa I_H(t) - \kappa I_J(t) \right] dt + \sigma dB_H(t) + dJ(t). \end{aligned}$$

Simplify the coefficient of the dt term, by using (16) rearranged as:

$$I_H(t) + I_J(t) = \tilde{X}(t) - e^{-\kappa t} X(0) - \theta(1 - e^{-\kappa t}).$$

Substitute this into the dt coefficient:

$$\begin{aligned} &\kappa e^{-\kappa t} (\theta - X(0)) - \kappa \left(\tilde{X}(t) - e^{-\kappa t} X(0) - \theta(1 - e^{-\kappa t}) \right) \\ &= \kappa \theta e^{-\kappa t} - \kappa X(0) e^{-\kappa t} - \kappa \tilde{X}(t) + \kappa X(0) e^{-\kappa t} + \kappa \theta (1 - e^{-\kappa t}) \\ &= \kappa \theta e^{-\kappa t} - \kappa \tilde{X}(t) + \kappa \theta - \kappa \theta e^{-\kappa t} \\ &= \kappa (\theta - \tilde{X}(t)). \end{aligned}$$

Thus, the combined differential is:

$$d\tilde{X}(t) = \kappa(\theta - \tilde{X}(t)) dt + \sigma dB_H(t) + dJ(t).$$

This is precisely the target SDE (6).

In summary, we have rigorously shown that the process $\tilde{X}(t)$ defined by (16) satisfies the stochastic differential equation (6). The derivation relies on the pathwise interpretation of the stochastic integrals and standard rules of calculus applied to the deterministic part, combined with the Leibniz rule for pathwise Stieltjes integrals for the fractional and jump components. The calculation confirms that $\tilde{X}(t)$ is the unique strong solution to the SDE (6). $\square$

Appendix B. Algorithm Pseudocode

Algorithm A1 Simulation of Karst Spring Discharge using Jump-Diffusion fOU Model

Input: Observed discharge $X(t)$, $t = 1..N$; Quantile probabilities $p_{\Delta X}$ (e.g., 0.975), p_X (e.g., 0.90 for fOU truncation); Gaussian copula correlation ρ_G ; Number of copula pairs L ; Model parameters: δ (OU persistence adjustment), γ (damping strength), c (jump size scaling factor), c_v (fGn volatility scaling factor), C (max discharge multiplier for damping); Simulation length $N_{sim} = N$; Initial value $\tilde{X}(0)$.

Output: Simulated discharge $\tilde{X}(t)$, $t = 1..N_{sim}$; Estimated parameters $\hat{H}, \hat{\kappa}, \hat{\sigma}, (\hat{k}_J, \hat{\lambda}_J), (\hat{k}_\tau, \hat{\lambda}_\tau), \hat{\rho}$.

Phase 1: Parameter Estimation from Observed Data

- 1: $\Delta X(t) \leftarrow X(t) - X(t-1)$ for $t = 2..N$.
- 2: $q_{\Delta X} \leftarrow \text{quantile}(\{\Delta X(t)\}_{t=2}^N, p_{\Delta X})$.
- 3: $\{T_j\}_{j=1}^M \leftarrow \{t \mid \Delta X(t) > q_{\Delta X}\}$. ▷ Indices of detected jumps
- 4: $J_j \leftarrow \Delta X(T_j) - q_{\Delta X}$ for $j = 1..M$. ▷ Jump sizes (excess over threshold)
- 5: $\tau_j \leftarrow T_j - T_{j-1}$ for $j = 1..M$ (with $T_0 = 0$). ▷ Inter-arrival times
- 6: Fit candidate distributions (e.g., Weibull, Gamma) to $\{J_j\}_{j=1}^M$ via MLE. Select best fit (e.g., Weibull with params $\hat{k}_J, \hat{\lambda}_J$).
- 7: $\tau_j^{pert} \leftarrow \tau_j + \text{Uniform}(0.01, 0.5)$ for $j = 1..M$. ▷ Perturb inter-arrival times
- 8: $(\hat{k}_\tau, \hat{\lambda}_\tau) \leftarrow \text{fitdistr_Weibull}(\{\tau_j^{pert}\}_{j=1}^M)$. ▷ MLE for perturbed inter-arrival times
- 9: $\hat{\rho} \leftarrow \text{Corr}(J, \tau)$ ▷ Estimate jump size, jump time correlation
- 10: $\hat{H} \leftarrow 2 - \text{fd_estimate_madogram}(X(t))$. ▷ Hurst parameter
- 11: $q_X \leftarrow \text{quantile}(\{X(t)\}_{t=1}^N, p_X)$.
- 12: $X'(t) \leftarrow \min(X(s), q_X)$. ▷ Truncated series for fOU parameters
- 13: $\hat{\sigma}^2 \leftarrow \frac{\text{Var}(\Delta^2 X'(t))}{4 - 2\hat{H}}$. ▷ Background volatility from $X'(t)$
- 14: $\hat{\kappa} \leftarrow \left(\frac{\text{Var}(X'(t))}{\hat{\sigma}^2 \hat{H} \Gamma(2\hat{H})} \right)^{-1/(2\hat{H})}$. ▷ Background mean reversion rate from $X'(t)$

Phase 2: Simulation Setup for Jumps and Baseline

- 15: Generate L pairs $(Z_{1,i}, Z_{2,i})$ from bivariate standard normal with correlation ρ_G .
- 16: $(U_i, V_i)_{i=1}^L \leftarrow (\Phi(Z_{1,i}), \Phi(Z_{2,i}))_{i=1}^L$. ▷ Transform to uniform via normal CDF to get an empirical copula
- 17: $\tau'_i \leftarrow F_\tau^{-1}(U_i; \hat{k}_\tau, \hat{\lambda}_\tau)$ for $i = 1..L$. ▷ Simulated perturbed inter-arrival times
- 18: $\tau'_i \leftarrow \text{round}(\tau'_{i,raw})$. Ensure $\tau'_i \geq 1$. ▷ Convert to discrete days, e.g., by rounding or truncation
- 19: $J'_i \leftarrow F_J^{-1}(V_i; \hat{k}_J, \hat{\lambda}_J)$ for $i = 1..L$. ▷ Simulated raw jump sizes (excess)
- 20: $\mathcal{T}_{sim} \leftarrow \{\}; \mathcal{J}_{sim} \leftarrow \{\}$. ▷ Initialize lists for selected thinned jumps
- 21: **Repeat** ▷ Thinning loop
- 22: Randomly sample a subset of indices $S = \{s_1, \dots, s_{|S|}\} \subset \{1..L\}$, where $|S| \ll L$ but still sufficiently large for the next step.
- 23: Form candidate inter-arrival times $\mathcal{T}_{cand} = \{\tau'_{s_i} \mid i \in 1..M\}$ by choosing $M \leq |S|$ so that $\sum_{i=1}^M \tau'_{s_i} \leq N_{sim} < \sum_{i=1}^{M+1} \tau'_{s_i}$.
- 24: Form the pairs (τ'_{s_i}, J'_{s_i}) , $i = 1..M$ to obtain the pair of vectors $(\mathcal{T}_{cand}, \mathcal{J}_{cand})$.
- 25: **If** $\hat{\rho} < \text{Corr}(\mathcal{T}_{cand}, \mathcal{J}_{cand}) < \rho_G$ **then**
- 26: $(\mathcal{T}_{sim}, \mathcal{J}_{sim}) \leftarrow (\mathcal{T}_{cand}, \mathcal{J}_{cand})$
- 27: **Break Repeat** ▷ Exit loop if criteria met
- 28: **End If**
- 29: **End Repeat**
- 30: Let $M_{final} = |\mathcal{T}_{sim}|$.
- 31: $T_{sim,j} \leftarrow \sum_{k=1}^j \tau_{sim,k}$ for $j = 1..M_{final}$. ▷ Final simulated jump times
- 32: $J^*(t) \leftarrow 0$ for $t = 1..N_{sim}$. ▷ Initialize jump innovation array
- 33: **For** $j = 1$ **to** M_{final} **do**
- 34: $J^*(T_{sim,j}) \leftarrow J_{sim,j} + q_{\Delta X}$. ▷ Populate with total jump magnitude
- 35: **End For**
- 36: $\epsilon'(t) \leftarrow \text{Generate_fGn}(\hat{H}, N_{sim})$. ▷ Generate raw fGn (e.g., diff of FBM)
- 37: $\epsilon(t) \leftarrow c_v \cdot \hat{\sigma} \cdot \text{scale_to_unit_variance}(\epsilon'(t))$ for $t = 1..N_{sim}$. ▷ Scale fGn

Phase 3: Combined Process Iterative Simulation

- 38: Initialize simulation array $\tilde{X}(t)$ for $t = 0..N_{sim}$; Set $\tilde{X}(0)$.
- 39: **For** $t = 1$ **to** N_{sim} **do**
- 40: $\text{Innov}(t) \leftarrow \epsilon(t) + J^*(t)$. ▷ Total innovation (fGn + jump)
- 41: $\tilde{X}(t) \leftarrow e^{-\hat{\kappa}} \cdot \tilde{X}(t-1) + \theta(1 - e^{-\kappa}) + \text{Innov}(t)$. ▷ Discretized fOU sample
- 42: **End For**
- 43: **return** $\tilde{X}(t)$ for $t = 1..N_{sim}$, and all estimated parameters.

References

1. Addison, P. *Fractals and Chaos: An Illustrated Course*; Institute of Physics Publishing: Bristol, UK, 1997.
2. Bercu, B.; Courtin, S.; Savy, N. MLE for the drift parameter of the fractional Ornstein-Uhlenbeck process. *arXiv* **2016**, arXiv:1604.04030.
3. Biagini, F.; Hu, Y.; Øksendal, B.; Zhang, T. *Stochastic Calculus for Fractional Brownian Motion and Applications*; Springer Science and Business Media: London, UK, 2008.
4. Borbás, E.; Márkus, L.; Darougi, A.; Kovács, J. Characterization of karstic aquifer complexity using fractal dimensions. *GEM - International Journal on Geomathematics* **2021**, *12*, 4, 1–27. doi: 10.1007/s13137-020-00169-5
5. Fiorillo, F. The Recession of Spring Hydrographs, Focused on Karst Aquifers. *Water Resour. Manag.* **2014**, *28*, 1781–1805.
6. Friz, P.K.; Hairer, M. *A Course on Rough Paths: With an Introduction to Regularity Structures*; Springer: Cham, Switzerland, 2014.
7. Goldscheider, N.; Mádl-Szőnyi, J.; Erőss, A.; Schill, E. Thermal water resources in carbonate rock aquifers. *Hydrogeol. J.* **2010**, *18*, 1303–1318.
8. Grill, J.; Kovács, S.; Less, Gy.; Réti, Zs.; Róth, L.; Szentpétery, I. Az Aggtelek-Rudabányai-hegység főtektonikai felépítése és fejlődéstörténete. (Main tectonic structure and evolution of the Aggtelek-Rudabánya Mountains). *Földt. Kut.* **1984**, *27*, 49–56. (In Hungarian)
9. Haas, J. *Geology of Hungary*; Springer: Berlin/Heidelberg, Germany, 2012.
10. Hergarten, S.; Birk, S. A fractal approach to the recession of spring hydrographs. *Geophys. Res. Lett.* **2007**, *34*, L11401, doi:10.1029/2007GL030289.
11. Hevesi, A. Magyarország karsztvidékeinek kialakulása és formakincse I-II. (The origin and landforms of the karst terrains of Hungary). *Földr. Közl.* **1991**, *115*, 25–35, 99–120. (In Hungarian)
12. Hu, Y.; Nualart, D. Parameter estimation for fractional Ornstein-Uhlenbeck processes. *Stat. Probab. Lett.* **2010**, *80*, 1030–1038.
13. Hu, Y.; Nualart, D.; Zhou, H. Parameter estimation for fractional Ornstein-Uhlenbeck processes of general Hurst parameter. *Stat. Inference Stoch. Process.* **2017**, *22*, 111–142.
14. Kleptsina, M.L.; Le Breton, A. Statistical analysis of the fractional Ornstein-Uhlenbeck type process. *Stat. Inference Stoch. Process.* **2002**, *5*, 229–248.
15. Kordos, L. *Magyarország barlangjai.* (Caves of Hungary); Gondolat Kiadó: Budapest, Hungary, 1984. (In Hungarian)
16. Labat, D.; Mangin, A.; Ababou, R. Rainfall-runoff relations for karstic springs: multifractal analyses. *J. Hydrol.* **2002**, *256*, 176–195.
17. Labat, D.; Hoang, C.T.; Masbou, J.; Mangin, A.; Tchiguirinskaia, I.; Lovejoy, S.; Schertzer, D. Multifractal behaviour of long-term karstic discharge fluctuations. *Hydrol. Process.* **2012**, *26*, 1072–1083.
18. Less, G. Polyphase evolution of the structure of the Aggtelek-Rudabánya Mountains (NE Hungary), the southern element of the inner Western Carpathians a review. *Slovak Geol. Mag.* **2000**, *6*, 260–268.
19. Lohvinenko, T.; Ralchenko, K. Parameter estimation for the fractional Ornstein-Uhlenbeck process with periodic mean. *Mod. Stoch. Theory Appl.* **2019**, *6*, 337–355.
20. Majone, B.; Bellin, A.; Borsato, A. Runoff generation in karst catchments: multifractal analysis. *J. Hydrol.* **2004**, *294*, 176–195.
21. Mandelbrot, B.B. *Fractals: Form, Chance, and Dimension*, Revised ed.; W. H. Freeman and Co.: San Francisco, CA, USA, 1977.
22. Maucha, L. *Az Aggtelek-hegység karszthidrológiai kutatásai eredményei és zavartalan hidrológiai adatai 1958-1993.* (Results of the karst hydrological research in Aggtelek Mts and its unbiased hydrological data 1958-1993); Water Resources Research Institute VITUKI Rt., Hydrological Institute: Budapest, Hungary, 1998. (In Hungarian)
23. Pardo-Igúzquiza, E.; Dowd, P.A.; Durán, J.J.; Robledo-Ardila, P. A review of fractals in karst. *Int. J. Speleol.* **2019**, *48*, 11–20.
24. Ross, S.M. *Stochastic Processes*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1996.
25. Székely, K., Ed. *Magyarország fokozottan védett barlangjai.* (Strictly Protected Caves of Hungary); Mezőgazda Kiadó: Budapest, Hungary, 2003. (In Hungarian)
26. Tanaka, K.; Xiao, W.; Yu, J. Maximum likelihood estimation for the fractional Ornstein-Uhlenbeck process with discrete observations. *Econom. Theory* **2020**, *36*, 929–962.

27. Telbisz, T.; Gruber, P.; Mari, L.; Kőszeghi, M.; Bottlik, Zs.; Standovár, T. Geological Heritage, Geotourism and Local Development in Aggtelek National Park (NE Hungary). *Geoheritage* **2020**, *12*, 5, doi:10.1007/s12371-019-00423-z.
28. Veneziano, D.; Essiam, A.K. Flow through porous media with multifractal hydraulic conductivity. *Water Resour. Res.* **2003**, *39*, 1291, doi:10.1029/2001WR001018.
29. Xiao, W.; Wang, X.; Yu, J. Estimation and inference in fractional Ornstein–Uhlenbeck model with discrete-sampled data. Unpublished manuscript, 2018. Available online: <https://www.stern.nyu.edu/sites/default/files/assets/documents/TwoStageVm09.pdf> (accessed on 10 October 2024).
30. Young, L.C. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.* **1936**, *67*, 251–282.
31. Zähle, M. Integration with respect to fractal functions and stochastic calculus I. *Probab. Theory Relat. Fields* **1998**, *111*, 333–374.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.