

Parameter Inference for a Simple Stochastic Hydrological Model

C. Albert*, S. Ulzega , P. Reichert

*Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf,
Switzerland*

February 18, 2015

In this paper, we study the feasibility of a full Bayesian inference, for simple conceptual rainfall-runoff models with scale-invariant noise. We start with a single linear reservoir and describe the dynamics of its water content, $S(t)$, by means of the stochastic differential equation

$$\dot{S}(t) = r(t) - \frac{1}{K} \left(1 + \frac{\gamma}{2}\right) S(t) + \sqrt{\frac{\gamma}{K}} S(t) \eta(t), \quad (1)$$

where $r(t)$ describes the rain input and $\eta(t)$ denotes white noise, i.e.,

$$\langle \eta(t) \eta(t') \rangle = \delta(t - t'). \quad (2)$$

Given rainfall and runoff time-series, we want to infer the two system parameters K and γ , measuring, respectively, the *retention time*, and the noise. The noise parameter is dimensionless, which renders the noise term scale-invariant. Due to the state-dependence of the noise term, eq. (1) is ill-defined unless we specify the discretization convention. For computational reasons, we work with the *Stratonovich* convention. Our parametrization is such that, for constant rain input $r(t) = r$, the mean of S converges to Kr , in the long-time limit (see below).

According to ?, the probability $P(S)$ of finding the system in a state S at time $t = T$ given that it was in an initial state S_0 at time $t = 0$, can be expressed in the form of a path integral as,

$$P(S) = \int_{S_0}^S e^{-\tilde{S}[\dot{S}(t), S(t)]} \mathcal{D}S, \quad (3)$$

where the integral extends over all paths starting at S_0 and ending at S . The path-measure $\mathcal{D}S$ is the formal limit $\Delta t \rightarrow 0$ of the discretized measure

$$\mathcal{D}S = \prod_i \frac{dS_i}{\sqrt{2\pi \Delta t \gamma / K S_i}}. \quad (4)$$

The *action* is a functional defined on paths $\tilde{S} : [0, T] \rightarrow \mathbb{R}_+$ and reads as

$$\tilde{S}[\dot{S}(t), S(t)] = \int_0^T dt \left\{ \frac{K}{2\gamma S^2(t)} \left(\dot{S}(t) - r(t) + \frac{1+\gamma}{K} S(t) \right)^2 - \frac{2+\gamma}{4K} \right\}. \quad (5)$$

*Email: carlo.albert@eawag.ch

Note that the action includes the Jacobian that arises when changing coordinates from $\eta(t)$ to $S(t)$ and ensures the normalization condition $\int P(S) dS = 1$.

It is possible to calculate the equilibrium distribution $P_{\text{eq}}(S) = \lim_{T \rightarrow \infty} P(S(T))$ in the simple case of a constant rain input r_0 . Systems at thermal equilibrium must fulfill detailed balance, i.e.,

$$P(S_1 t_1 | S_0 t_0) P_{\text{eq}}(S_0) = P(S_0 t_1 | S_1 t_0) P_{\text{eq}}(S_1), \quad (6)$$

where $P(S_1 t_1 | S_0 t_0)$ is the conditional probability for the state S_1 at time t_1 given the initial state S_0 at time t_0 and $P(S_0 t_1 | S_1 t_0)$ is the conditional probability for the reverse path $\bar{S}(t) = S(-t)$. Assuming a solution in the form

$$P_{\text{eq}}(S) \propto e^{-\mathcal{H}(S)}, \quad (7)$$

equation (6) can be rewritten as

$$P(S_1 t_1 | S_0 t_0) = P(S_0 t_1 | S_1 t_0) \exp^{-\int_{t_0}^{t_1} dt \frac{dS}{dt} \frac{\mathcal{H}}{S}}. \quad (8)$$

After plugging in (3) for the conditional probabilities, and using the action (5), equation (8) yields

$$-r_0 + \frac{1+\gamma}{K} S = \frac{\gamma S^2}{2K} \frac{d\mathcal{H}(S)}{dS}. \quad (9)$$

In order to go from (8) to (9) we have exploited the property of the Stratonovich convention that the backward path is evaluated at the same points as the forward path. The solution to (9) is, up to an additive constant,

$$\mathcal{H}(S) = \frac{2Kr_0}{\gamma S} + \frac{2(1+\gamma)}{\gamma} \ln(S). \quad (10)$$

Thus, it turns out that $P_{\text{eq}}(S)$ is described by an inverse gamma distribution, $\Gamma^{-1}(\alpha, \beta)$ with scale parameter $\beta = 2Kr_0/\gamma$ and shape parameter $\alpha = (2+\gamma)/\gamma$, i.e.,

$$P_{\text{eq}}(S) \propto S^{-2(1+\gamma)/\gamma} e^{-2Kr_0/(\gamma S)}, \quad (11)$$

whose mean equals the equilibrium solution of the unperturbed system,

$$\langle S \rangle_{\text{eq}} = Kr_0, \quad (12)$$

and whose variance, for $\gamma < 2$, is given by

$$\langle (S - \langle S \rangle_{\text{eq}})^2 \rangle_{\text{eq}} = K^2 r_0^2 \frac{\gamma}{2-\gamma}. \quad (13)$$

Note that the variance diverges for $\gamma \geq 2$.

Let us consider now the general case of a time dependent rain input $r(t)$ with the following time-series of *observed* outputs,

$$y_i = \frac{1}{K} S(t_i) + \sigma \epsilon_i, \quad i = 1, \dots, n, \quad (14)$$

where the times t_i are chosen so that $0 = t_1 < t_2 < \dots < t_n = T$ and the ϵ_i are uncorrelated standard normal random variables. The *likelihood* function for the data (14) is, up to a factor that doesn't depend on the model parameters, given by the path-integral

$$f(\mathbf{y}|\boldsymbol{\theta}) \propto \int \exp \left[-\tilde{S}[\dot{S}, S] - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - S(t_i)/K)^2}{\sigma^2} \right] \mathcal{D}S(\boldsymbol{\theta}), \quad (15)$$

with the parameter vector $\boldsymbol{\theta} = (K, \gamma)^T$. Choosing for the parameters K and γ the improper prior given by

$$f(\boldsymbol{\theta}) = f(K)f(\gamma) = \frac{1}{K\gamma}, \quad (16)$$

parameter inference means generating a sufficiently large sample of parameter vectors from the posterior

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}). \quad (17)$$

We replace the variable $S(t)$ and the parameter K by the dimensionless quantities, $q(t)$ and β , respectively, by means of the transformations

$$\beta = \sqrt{\frac{T\gamma}{K}}, \quad S(t) = \frac{T\gamma r(t)}{\beta^2} e^{\beta q(t)}. \quad (18)$$

In terms of these dimensionless variables and parameters the posterior reads (without prior)

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto \int \exp \left[-\tilde{S}[\dot{q}(t), q(t)] - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - r_i e^{\beta q_i})^2}{\sigma^2} \right] \mathcal{D}q, \quad (19)$$

where r_i and q_i represent $r(t)$ and $q(t)$, respectively, at time t_i . In the transformed variables the path-measure takes the form, up to a parameter-independent multiplicative factor,

$$\mathcal{D}q = \prod_t dq(t), \quad (20)$$

while the action can be rewritten in the compact form,

$$\tilde{S}[\dot{q}, q] = \frac{1}{2T} \int_0^T dt \left\{ \left(T\dot{q}(t) + \rho(t) - \frac{\beta}{\gamma} e^{-\beta q(t)} \right)^2 - \frac{(2 + \gamma)\beta^2}{2\gamma} \right\}, \quad (21)$$

with

$$\rho(t) = \frac{T\dot{r}(t)}{\beta r(t)} + \frac{\beta(\gamma + 1)}{\gamma} = \frac{T}{\beta} \frac{d[\ln r(t)]}{dt} + \frac{\beta(\gamma + 1)}{\gamma}. \quad (22)$$

In order to ease the numerical computations, it is useful to introduce a generalized time-dependent Hamiltonian,

$$\mathcal{H}_q(q, t) = \frac{1}{\gamma} e^{-\beta q} + q\rho(t), \quad (23)$$

and rewrite the action (21) as,

$$\begin{aligned}
\mathcal{S}[\dot{q}, q] &= \frac{1}{2T} \int_0^T dt \left\{ T^2 \dot{q}^2(t) + \left(\rho(t) - \frac{\beta}{\gamma} e^{-\beta q(t)} \right)^2 - 2T \frac{\partial \mathcal{H}_q(q, t)}{\partial t} - \frac{(2 + \gamma)\beta^2}{2\gamma} \right\} \\
&\quad + \mathcal{H}_q(q(T), T) - \mathcal{H}_q(q(0), 0) \\
&= \frac{1}{2T} \int_0^T dt \left\{ T^2 \dot{q}^2(t) + \left(\rho(t) - \frac{\beta}{\gamma} e^{-\beta q(t)} \right)^2 - 2T q(t) \dot{\rho}(t) - \frac{(2 + \gamma)\beta^2}{2\gamma} \right\} \\
&\quad + \frac{1}{\gamma} e^{-\beta q(T)} + q(T) \rho(T) - \frac{1}{\gamma} e^{-\beta q(0)} - q(0) \rho(0). \quad (24)
\end{aligned}$$

Since the integral on the r.h.s. of (19) is very expensive to calculate, we will sample simultaneously parameter vectors $\boldsymbol{\theta}$ and discretized system realizations $q(t)$ directly from an appropriate discretization of the kernel of (19). Since this distribution is very high dimensional and strongly correlated, we employ the method called *Hamiltonian Monte Carlo* (HMC) introduced by ?. This method interprets the exponent of the kernel of the posterior as the potential of a one-dimensional statistical mechanical system. Each degree of freedom, $q(t)$ and $\boldsymbol{\theta}$ in our case, is paired with a conjugate variable, $p(t)$ and $\boldsymbol{\pi}$, respectively, and the system is defined by the Hamiltonian

$$\mathcal{H}_{\text{HMC}}(q, \boldsymbol{\theta}; p, \boldsymbol{\pi}) = K(p, \boldsymbol{\pi}) + V(q, \boldsymbol{\theta}), \quad (25)$$

where

$$K(p, \boldsymbol{\pi}) = \int \frac{p^2(t)}{2m_q} dt + \sum_{\alpha=1}^s \frac{\pi_{\alpha}^2}{2m_{\alpha}}, \quad (26)$$

where in our case $s = 2$, and $V(q, \boldsymbol{\theta})$ is the logarithm of the kernel of (19). The HMC method, which is a combination of the *Metropolis algorithm* by ? and *molecular dynamics* methods [ref?], iterates the following steps

1. Momenta are sampled from the Gaussian distribution defined by (26).
2. A volume-preserving and time-reversible solution to a discretization of Hamilton's equations is calculated numerically (leap-frog algorithm).
3. The discretization error on the energy preservation due to the previous step is corrected by a Metropolis acceptance/rejection step.

The last step is the standard Metropolis algorithm, while the first two steps allow to make relatively large jumps in phase space while maintaining a relatively large acceptance rate.

In particular, step 2 requires the calculation of the gradient of (25), which is a somewhat tedious manual labour. Before we set out to do so, let us first write down the proper discretization of the path-integral (19), with the action (24). Therefore, let us assume that the measurement time points $\{y_i\}_{i=1, \dots, n}$ of the time series (14) are equidistantly distributed on the time interval $[0, T]$. We further partition this interval into N bins, with $N \gg 1$, according to the time series $0 = t_1 < t_2, \dots < t_N = T$, such that the time points $t_{a_{\kappa}}$ with indexes $a_{\kappa} = \kappa(N - 1)/n + 1$, where $\kappa = 1, \dots, n$, coincide with the data points y_i . Under these assumptions the discretized versions of $K(p, \boldsymbol{\pi})$ and $V(q, \boldsymbol{\theta})$ read

$$K(p, \boldsymbol{\pi}) = \sum_{i=1}^N \frac{p_i^2}{2m_q} \Delta t + \sum_{\alpha=1}^s \frac{\pi_\alpha^2}{2m_\alpha}, \quad (27)$$

$$\begin{aligned} V(q, \boldsymbol{\theta}) \approx & \sum_{i=2}^N \frac{K}{2\gamma\Delta t} (q_i - q_{i-1})^2 + \sum_{i=2}^{N-1} \frac{\Delta t}{\gamma} \left\{ \frac{(\rho_i - e^{-q_i})^2}{2K} - q_i \frac{(\rho_i - \rho_{i-1})}{\Delta t} \right\} \\ & - \frac{T(\gamma + 2)}{4K} + \frac{(e^{-q_N} + q_N \rho_{N-1} - e^{-q_1} - q_1 \rho_1)}{\gamma} \\ & + \sum_{\kappa=1}^n \frac{(y_\kappa - r_{a_\kappa} e^{q_{a_\kappa}})^2}{2\sigma^2} + \ln \left(K^{1-N/2} \gamma^{1+N/2} \right), \end{aligned} \quad (28)$$

where the last term comes from the path measure (20) and

$$\rho_i = K \frac{\ln(r_{i+1}/r_i)}{\Delta t} + \gamma + 1. \quad (29)$$

For strongly varying rain input or high-frequency output data, Δt needs to be chosen small. Thus, the harmonic part of the Hamiltonian becomes stiff and, consequently, the time-step of MD needs to be chosen small, which leads to slow convergence. To solve this problem, we change coordinates, such that the harmonic part of the Hamiltonian becomes (partly) diagonal and the particle interaction moves to the potential. One option would be to use the Fourier modes of $\eta(t)$ as configurational variables. Calculating the potential requires then an inversion of a Toeplitz matrix. This approach is interesting because, in our case, it allows for an approximate analytical integration of high-frequency modes.

Another approach is the so-called *staging*, described in Tuckerman et al. (1993). This approach is interesting if different time-scales are involved. We first note that the Hamiltonian $\mathcal{H}_{\text{HMC}}(q, \boldsymbol{\theta}; p, \boldsymbol{\pi})$ (25) can be seen as the classical Hamiltonian of a polymer chain with harmonic bonds between neighbouring beads, described by the first harmonic term of $V(q, \boldsymbol{\theta})$, in an external field described by the other non-harmonic terms of $V(q, \boldsymbol{\theta})$. Sampling system realizations, each defined by a set of coordinates $\{q_i\}$ and parameters $\boldsymbol{\theta}$, is equivalent to sampling states in the polymer dynamics. Thus, let us place $n+1$ equidistant (data) beads on the time domain, the first sitting on $t = 0$ and the last on $t = T$. In between, we place $n(j-1)$ equidistant staging beads. So that, in total, we have $nj+1 = N$ equidistant beads.

The discretized harmonic part of the action can be rewritten as

$$\begin{aligned} & \sum_{i=2}^N \frac{K}{2\gamma\Delta t} (q_i - q_{i-1})^2 \\ &= \frac{K}{2\gamma} \sum_{s=1}^n \left\{ \frac{(q_{(s-1)j+1} - q_{sj+1})^2}{j\Delta t} + \sum_{k=2}^j \frac{k}{(k-1)\Delta t} (q_{(s-1)j+k} - q_{(s-1)j+k}^*)^2 \right\}, \end{aligned} \quad (30)$$

with

$$q_{(s-1)j+k}^* = \frac{(k-1)q_{(s-1)j+k+1} + q_{(s-1)j+1}}{k}. \quad (31)$$

Following Tuckerman et al. (1993) we apply the coordinate transformation

$$u_{sj+1} = q_{sj+1}, \quad u_{sj+k} = q_{sj+k} - q_{sj+k}^*, \quad (32)$$

with its inverse given by

$$q_{sj+1} = u_{sj+1}, \quad q_{sj+k} = \sum_{l=k}^{j+1} \frac{k-1}{l-1} u_{sj+l} + \frac{j-k+1}{j} u_{sj+1}, \quad (33)$$

or equivalently by the recursive relation

$$q_{sj+k} = u_{sj+k} + \frac{k-1}{k} q_{s+k+1} + \frac{1}{k} q_{sj+1}. \quad (34)$$

Moreover, we introduce the following transformation for the parameters,

$$\beta = \ln(K/T), \quad \tau = \ln(\gamma) \quad (35)$$

and the corresponding transformed vector $\tilde{\boldsymbol{\theta}} = (\beta, \tau)^T$. This transformation ensures that the system parameters K and γ will be kept positive during the HMC algorithm.

The potential $V(q, \tilde{\boldsymbol{\theta}})$ can now be split into a fast component that scales linearly or quadratically with N or n ,

$$\begin{aligned} \mathcal{V}_f(\mathbf{u}, \tilde{\boldsymbol{\theta}}) = & \frac{N e^{\beta-\tau}}{2} \sum_{s=1}^n \left\{ \frac{(u_{(s-1)j+1} - u_{sj+1})^2}{j} + \sum_{k=2}^j \frac{k}{(k-1)} u_{(s-1)j+k}^2 \right\} \\ & + \frac{N}{2} (\beta - \tau) - \sum_{s=0}^n \frac{(y_s - r_{sj+1} e^{u_{sj+1}})^2}{2\sigma^2}, \end{aligned} \quad (36)$$

where we have used $T = N\Delta t$, and a slow N - and n -independent component,

$$\begin{aligned} \mathcal{V}_s(\mathbf{q}, \tilde{\boldsymbol{\theta}}) = & \sum_{i=2}^{N-1} \left\{ \frac{e^{-(\beta+\tau)}}{2N} (\rho_i - e^{-q_i})^2 - e^{-\tau} q_i (\rho_i - \rho_{i-1}) \right\} - \frac{(2 + e^\tau) e^{-\beta}}{4} \\ & + e^{-\tau} (e^{-q_N} + q_N \rho_{N-1} - e^{-q_1} - q_1 \rho_1). \end{aligned} \quad (37)$$

In the expressions above we have included the Jacobian associated with (35).

We can now fully exploit the two different time-scales that are involved in the problem. First, the harmonic part of the fast component $\mathcal{V}_f[\dot{\mathbf{u}}, \mathbf{u}]$ can be seen as describing the dynamics of heavy coupled *boundary* data-point beads and lighter uncoupled *staging* beads. The different weights of boundary and staging beads are defined by adjusting their *effective masses*, m_{bdy} and m_{st} , respectively. These values are used in (27) for the masses m_q . Second, the two different ranges of the forces associated with $\mathcal{V}_f(\mathbf{u}, \tilde{\boldsymbol{\theta}})$ and $\mathcal{V}_s(\mathbf{q}, \tilde{\boldsymbol{\theta}})$ can be efficiently taken into account by applying a reversible reference system propagator algorithm (RESPA), as outlined in Tuckerman et al. (1992). The RESPA algorithm, based on appropriate reversible molecular dynamics integrators, can significantly accelerate simulations of systems that are simultaneously subject to short- and long-range forces by separating the two different time-scales.

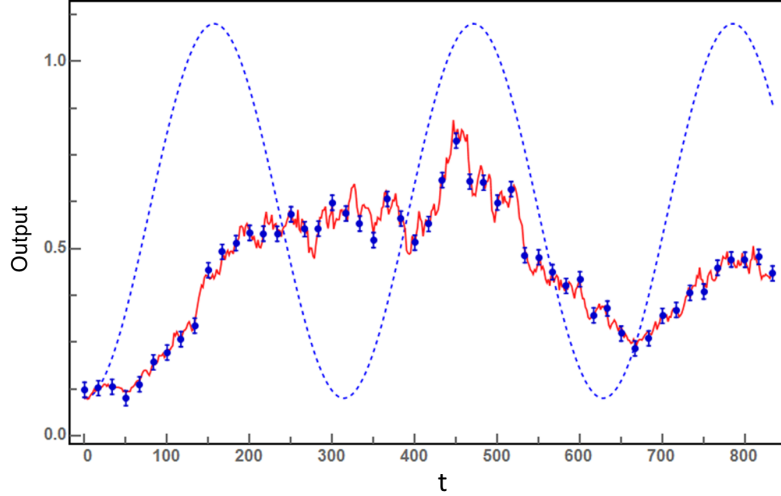


Figure 1: Rain input (dashed), system realization (solid line) and synthetic experimental data. The system output and the observed data were generated using $K = 200$ and $\gamma = 0.2$.

The basic idea of the RESPA algorithm is to apply a Trotter decomposition to the system Liouville propagator \mathcal{L} to generate a discrete time propagator factorized into fast and slow contributions. Indeed, as shown in Tuckerman et al. (1992), the propagator $\exp[i\mathcal{L}\Delta\tau]$ can be written as

$$\mathcal{G}(\Delta\tau) = \exp\left\{\frac{\Delta\tau}{2}F_s(x)\frac{\partial}{\partial p}\right\} \exp\{i\mathcal{L}_f\Delta\tau\} \exp\left\{\frac{\Delta\tau}{2}F_s(x)\frac{\partial}{\partial p}\right\}, \quad (38)$$

where for the sake of clarity we have assumed a one-dimensional system. The argument can be easily extended to our multi-dimensional case. In (38) F_s is the force associated with the slowly varying long-range potential \mathcal{V}_s and $\mathcal{L}_f = \dot{x}(\partial/\partial x) + F_f(x)(\partial/\partial p)$ is the Liouville operator of the fast short-range potential \mathcal{V}_f . The propagator of the slow force F_s can be well approximated with a large time step $\Delta\tau$. On the other hand, in order to ensure stable dynamics, the short-range middle propagator is further factorized into n_f smaller time steps of duration $\delta\tau = \Delta\tau/n_f$, according to the Trotter expansion

$$\exp\{i\mathcal{L}_f\Delta\tau\} = \left[\exp\left\{\frac{\delta\tau}{2}F_f(x)\frac{\partial}{\partial p}\right\} \exp\left\{\delta\tau\dot{x}\frac{\partial}{\partial x}\right\} \exp\left\{\frac{\delta\tau}{2}F_f(x)\frac{\partial}{\partial p}\right\}\right]^{n_f}. \quad (39)$$

Coordinates and conjugate momenta, i.e., $(q, \theta; p, \pi)$ in our case, are propagated using (38) and (39) iteratively by applying a velocity Verlet algorithm.

Some preliminary results. The results shown in the figures refer to synthetic data obtained assuming a simple sinusoidal model for the rain input. The experimental data were then generated using the "true" parameter values $K = 200$ and $\gamma = 0.2$. We considered $n = 51$ data points, $j = 10$ staging points, and therefore $N = 501$ discretization points. The total time T is ≈ 833 [units of time], so that the discretization points are separated by a time interval $\Delta t \approx 1.6$ [units of time]. Following the RESPA method described above, we used a long time step $\Delta\tau = 0.02$ [units of time] for the evolution of the slow (long-range) potential, further subdivided in 16 smaller steps for the evolution of the fast (short-range) potential.

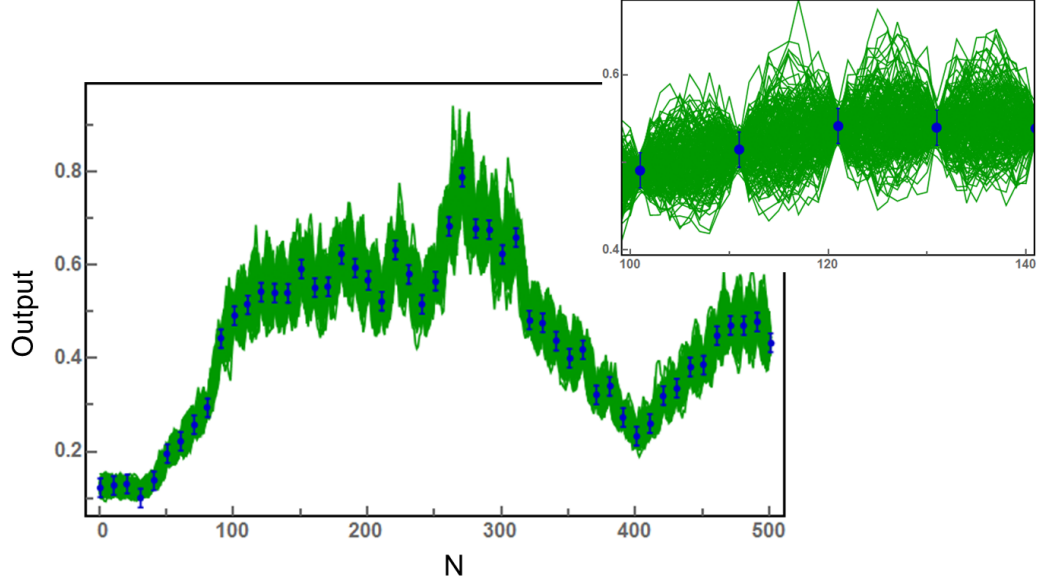


Figure 2: Simulated system realizations and synthetic experimental data obtained assuming a simple sinusoidal model for the rain input. In the inset one may appreciate the different dynamics of the heavy data-point beads and the light staging beads.

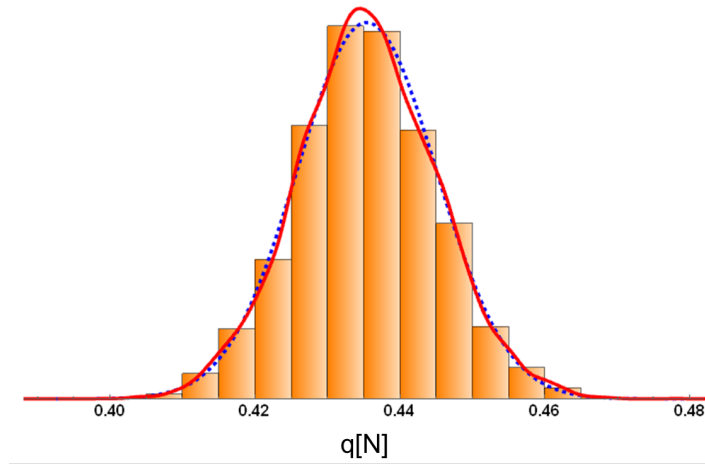


Figure 3: PDF for the system final state expressed via the coordinate set $\{q_N\}$. The dashed curve represents a normal distribution with the same mean and variance. The final state is basically normally distributed.

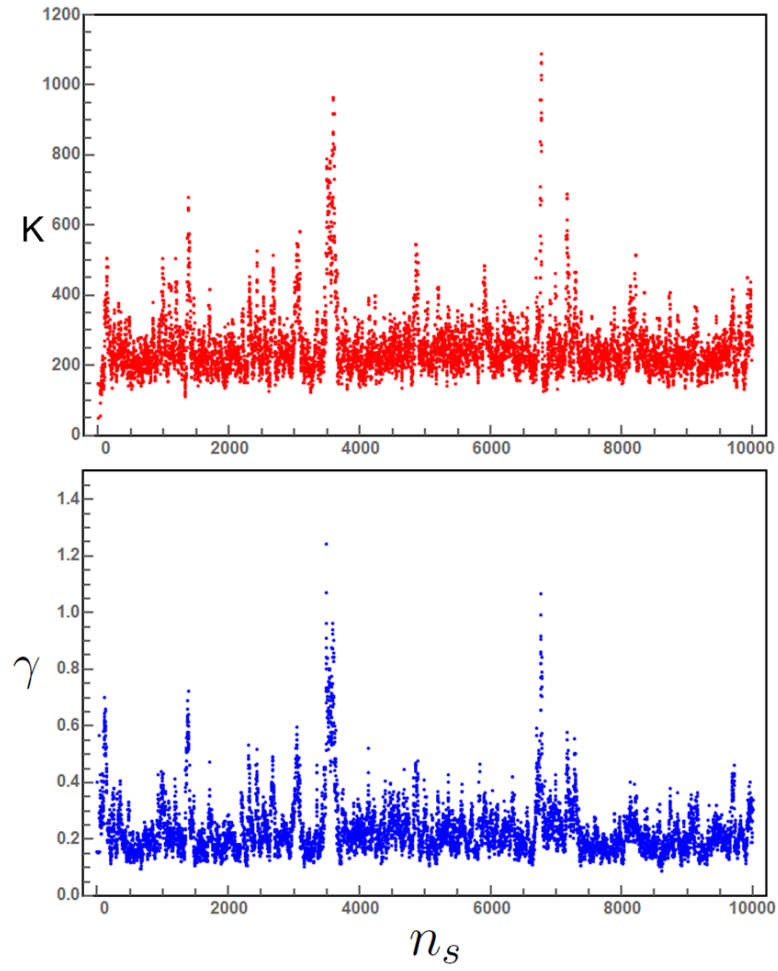


Figure 4: Markov chains for the two parameters K (top) and γ (bottom).

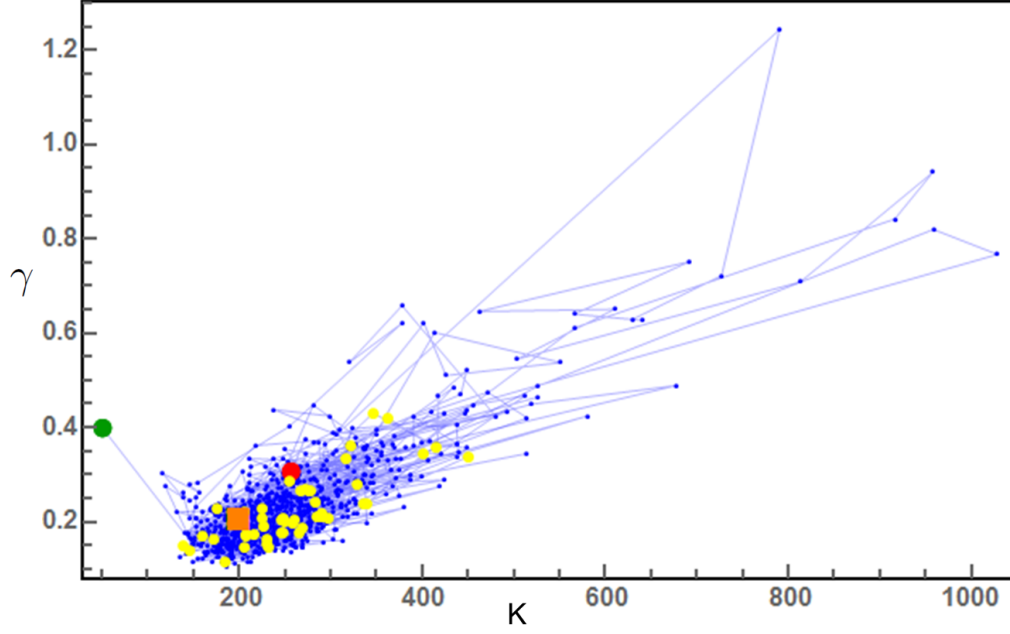


Figure 5: System dynamics in the phase space $K - \gamma$. The green dot is the initial state, the red dot is the final state, the yellow dots are the last 50 states of the HMC chain, and the orange square corresponds to the true parameter values used to generate the data.

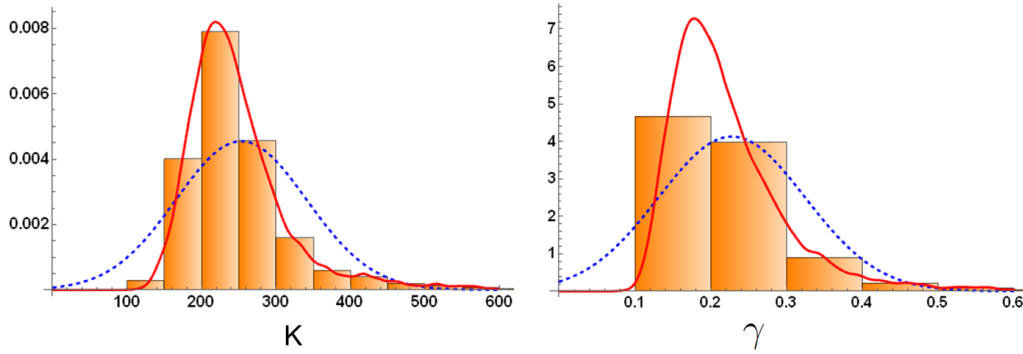


Figure 6: PDF for the parameters K (left) and γ (right). The true values used to generate the data are $K = 200$ and $\gamma = 0.2$. The initial values used in the simulations were $K = 50$ and $\gamma = 0.4$.

References