02427 Computer Exercise 03

Part 1: Simulation and discretization of diffusion processes

The Bonhoeffer-Van der Pol equations give a two-dimensional simplification of a famous four-dimensional system of ordinary differential equations proposed by Hodgkin and Huxley, in order to describe the firing of a single neuron. The equations are

$$\frac{dx_t^1}{dt} = \theta_3 \left(x_t^1 + x_t^2 - \frac{1}{3}\left(x_t^1\right)^3 + \theta_4 \right) \tag{1a}$$

$$\frac{dx_t^2}{dt} = -\frac{1}{\theta_3}\left(x_t^1 + \theta_2 x_t^2 - \theta_1\right) \tag{1b}$$

where x_t^1 is the negative potential over the membrane, x_t^2 is the permeability of the membrane, and $\theta_1, \theta_2, \theta_3$ and θ_4 denote the physical parameters. When $(\theta_1, \theta_2, \theta_3, \theta_4) = (0.7, 0.8, 3.0, -0.34)$ the system has a limit cycle, which describes the periodic slow charging and fast decharging that have been observed through experiments.

The effects of imperfections in the membrane and firing of the surrounding neurons can be simulated by incorporating additive noise with a small standard deviation in (1a). This leads to a two-dimensional system of Itô stochastic diffential equation

$$dX_t^1 = \theta_3 \left(X_t^1 + X_t^2 - \frac{1}{3}\left(X_t^1\right)^3 + \theta_4 \right) dt + \sigma dW_t \tag{2a}$$

$$dX_t^2 = -\frac{1}{\theta_3}\left(X_t^1 + \theta_2 X_t^2 - \theta_1\right) dt \tag{2b}$$

where W_t is a standard Wiener process and $\sigma > 0$ is the incremental standard deviation of the noise.

In order to be able to simulate (2), it is necessary to discretize it, i.e. to consider finite steps in the Wiener process ΔW_n instead of the infinitesimal dW_t. It can be shown that the so called Euler-Marayama approximation to (2) is given by the two-dimensional stochastic difference equation

$$Y_{n+1}^1 = Y_n^1 + \theta_3 \left(Y_n^1 + Y_n^2 - \frac{1}{3}\left(Y_n^1\right)^3 + \theta_4 \right) \Delta + \sigma \Delta W_{n+1}^1 \tag{3a}$$

$$Y_{n+1}^2 = Y_n^2 - \frac{1}{\theta_3}\left(Y_n^1 + \theta_2 Y_n^2 - \theta_1\right) \Delta \tag{3b}$$

where Δ is a suitably chosen small time interval, $\Delta W_{n+1} \in N(0, \Delta)$ and $\left(Y_n^1, Y_n^2\right), n = 1, \ldots, N$ is a discrete approximation to $\left(X_t^1, X_t^2\right)$ in the time interval $0 \leq t \leq T$.

Question 1.1

Let $\Delta = 2^{-9}, \theta_1 = 0.7, \theta_2 = 0.8, \theta_3 = 3.0, \theta_4 = -0.34$ og $\sigma = 0$. Simulate (3) in the time interval $0 \leq t \leq T = 100$ with initial conditions $Y_0^1 = -1.9$ and $Y_0^2 = 1.2$.

Plot the realizations of Y_k^1 and Y_k^2 and make a phaseplot of (Y_k^1, Y_k^2). Repeat for $\sigma = 0.10, 0.20, 0.30$ and 0.40 . Comment on the effect of adding noise to the equations.

Question 1.2

The Euler-Marayama method is an example of a weak method, implying that it only gives an approximation of (functionals of) the moments of X_t^1, X_t^2, whereas a strong method approximates the whole distribution. However, a weak method can give some essential visual information about the performance of the stochastic dynamic system in question. A histogram might for example indicate the form and support of the density function of the asymptotically stable, stationary solution.

Let $\sigma = 0.10$ and simulate (2) using the approximation (3) with the same parameter values as given above (you may reuse the results from question 1 if you like). Partition the phase plane in 100×100 equal cells. Count the number of trajectories that passes through each cell, and make a threedimensional plot (the units on the axes have no essential meaning in this case).

Which extra information does the plot contain, compared to the standard phase-plot?

Repeat for $\sigma = 0.20, 0.30$ and 0.40 (again using the results from question 1 if you like).

Hints

You can download a Matlab and R script for making the 2D histogram plot.

Part 2: Modeling of Rainfall-runoff

Introduction

In this exercise, you will explore how we can model rainfall-runoff using stochastic differential equations and estimate the models with ctsmTMB.

First, we introduce the concept of linear reservoir models (see e.g. (Breinholt et al., 2011)). In the linear reservoir model we approximate the runoff process as the discharge of water between a finite number of reservoirs (think of a series of basins). A simple illustration can be seen in Fig. 1. The model consists of a set of stochastic differential equations and is given as follows,

$$\begin{aligned} dX_{1,t} &= A \cdot U_t dt - \frac{n}{K} \cdot X_{1,t} dt + \sigma \cdot dw_{1,t} \\ dX_{2,t} &= \frac{n}{K} \cdot X_{1,t} dt - \frac{n}{K} \cdot X_{2,t} dt + \sigma \cdot dw_{2,t} \\ &\vdots \\ dX_{n+1,t} &= \frac{n}{K} \cdot X_{n,t} dt + \sigma \cdot dw_{n+1,t} \end{aligned} \tag{1}$$

Here, U_t is the rainfall intensity, treated as input data. The parameter A is the catchment area (the area within which the rainfall is relevant), while K is related to the time scale of the runoff process. The constant n is simply equal to the number of states (reservoirs) minus 1. The noise parameter σ can be a constant or a function, and in more complicated cases we may consider unique sigmas for each equation.

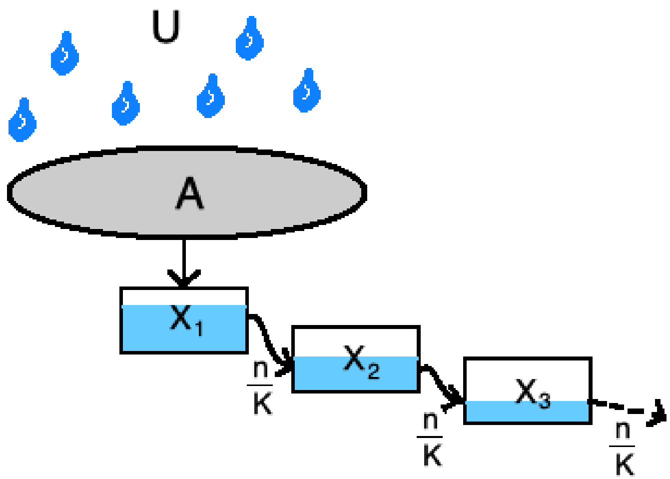

Figure 1: The concept of linear reservoirs.

Part 2.1 Linear rainfall-runoff

Consider the dataset found in `ex1_rainfallrunoff.csv`, which is illustrated in Fig. 2.

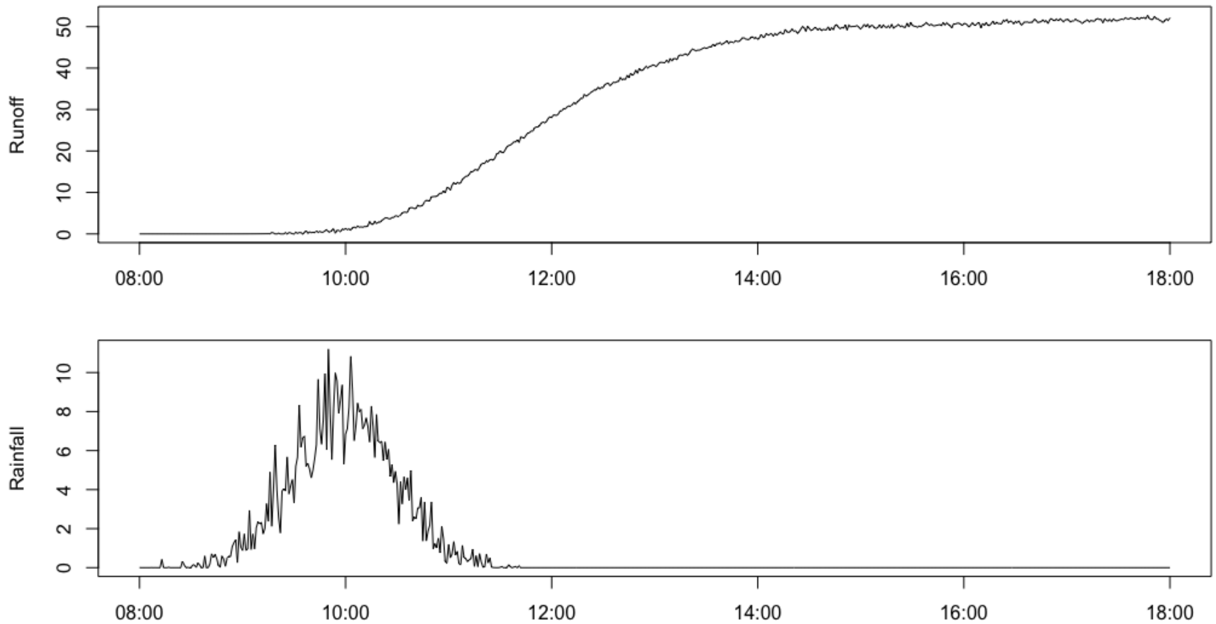

Figure 2: Data for Part 2.1.

<u>Questions:</u>

2.1.1 Formulate a linear reservoir model with 2 states ($n = 1$) and estimate it with ctsmTMB. The observations (called "stormwater") should correspond to the last of the two states. It is recommended to model the system such that the time unit is hours. Report the log-likelihood, and report on K and A.

2.1.2 Extend the model with more states, to find the most suitable number of states. Explain your model selection procedure. How many states are needed for the best fit?

2.1.3 Report on K and A for the best model. Compare with the 2-state model from Question 1 and explain the differences of the parameter estimates.

2.1.4 Report the correlation matrix for the parameters. What can be learned from it?

2.1.5 Validate the model (residual analysis), and plot the simulated values. Are you satisfied with the result?

Part 2.2 - Non-linear rainfall-runoff with overflow

Consider the dataset found in `ex2_overflow.csv`, which is illustrated in Fig. 3. This case is similar to the case in Exercise 1, except there is an overflow structure somewhere in the system. This means, a certain amount of stormwater needs to accumulate in the system before it starts flowing into the observed state. Note that the overflow structure is a side facility of the system. This means, at the location of the overflow structure, the stormwater can discharge two ways: it is always discharging further downstream at the usual rate, and when the overflow threshold is met, it is also discharging into the side facility, where the measurements (the column "stormwater" in the dataset) are taken.

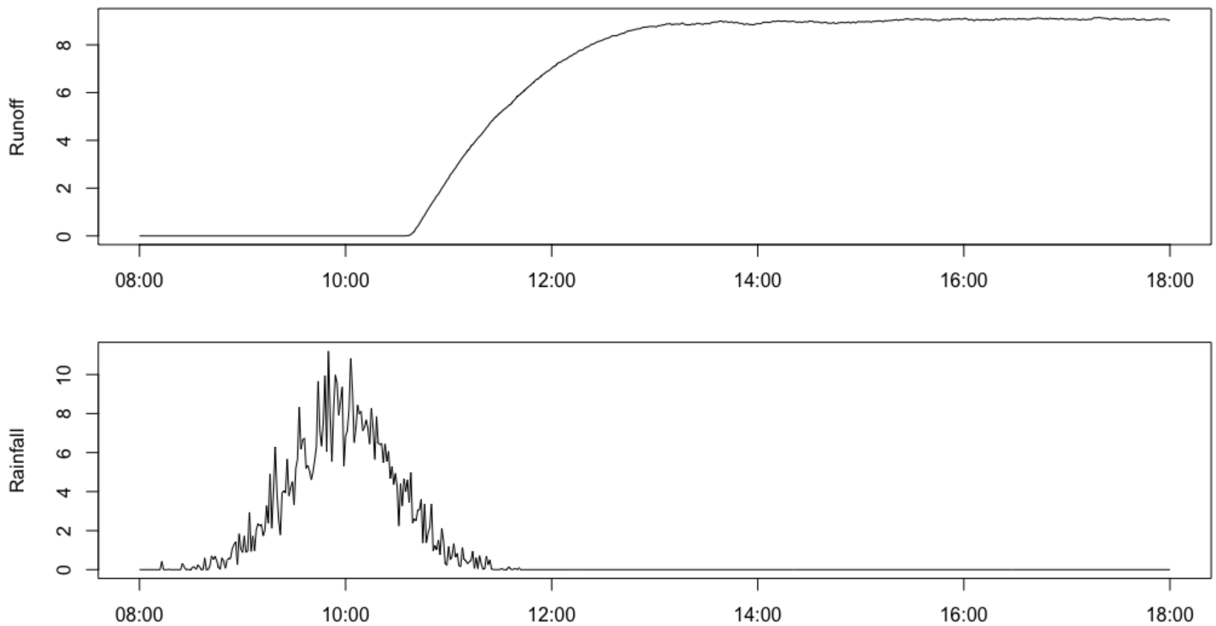

Figure 3: Data for Part 2.2.

Questions:

2.2.1 Formulate and estimate a linear reservoir model with a chosen number of states. Comment on the results. Is a linear model good enough to capture the dynamics of this case?

2.2.2 Find a way to introduce overflow into the system. Hint: use the sigmoid function

$$q(x) = \frac{1}{1 + \mathrm{e}^{-\alpha(x-\beta)}}, \tag{2}$$

where α and β are parameters. Estimate the extended model, and report the log-likelihood and all of the parameters.

2.2.3 Report the correlation matrix for the parameters. What can be learned from it?

2.2.4 Validate the model (residual analysis), and plot the simulated values. Are you satisfied with the result?

Part 2.3 - Larger real-life case

Consider the dataset found in `ex3_largecase.csv`. It contains 6 different stormwater events from the year 2018. The measurements come from a storage tower at the end of a stormwater tunnel. In simple terms, the rainfall first has to travel over the ground surface, then flow into the combined sewer system below the ground surface, and then at a certain point in the system, it can overflow into the stormwater tunnel if the water level is high enough. See also Fig. 4. Each part of the system (ground surface, combined sewer system and stormwater tunnel) may be characterized with its own K parameter if desired. The dataset contains the following attributes:

- Timestamp (given as UTC time)
- Rainfall: rainfall given in µm/min.
- Pumpflow: the flow rate of which the stormwater is being pumped out of the storage, given in m^3/min.
- Volume: the current total stormwater volume in the storage facility, given in m^3.
- the event ID number. The numbers serve no other purpose than distinguishing between separate rainfall events.

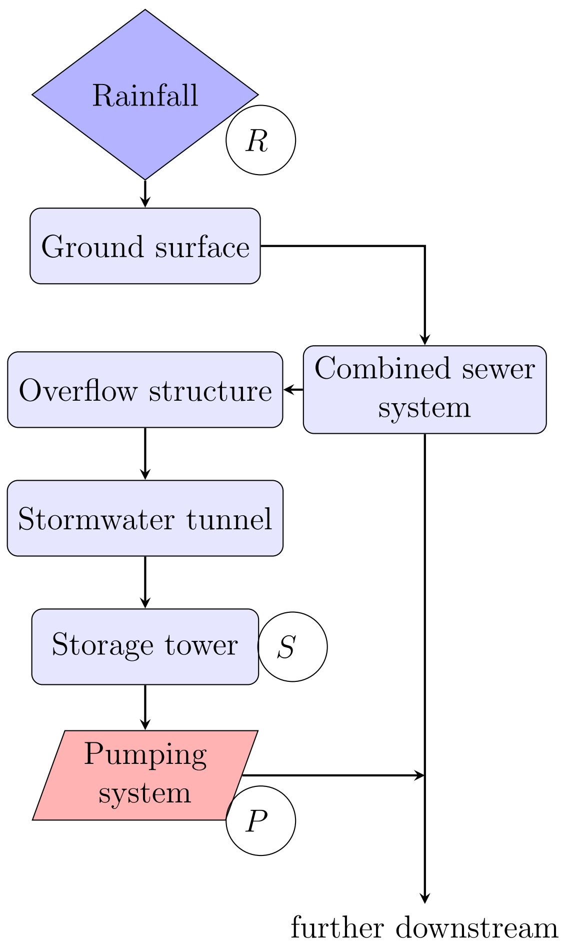

Figure 4: Simplified diagram of the stormwater system in Part 2.3. The presence of measurements are displayed as white circles (R = rainfall, S = stormwater volume, P = pumpflow).

All of the data is provided in 5-minute resolution. For the Volume attribute, the data points are the instantaneous measurements, while for the Rainfall and Pumpflow attributes are averaged values over 5 minutes. For example, at 2018-08-27 16:35 UTC, the rainfall is 114.5 µm/min. Hence, the total rainfall over the 5-minute interval 16:35:00-16:39:59 is estimated to 5 x 114.5 = 572.5 µm (or 0,57 mm).

The storage facility has a maximum capacity of roughly 29000 m^3.

The pumping operation follows a complex rule-based schedule which is not feasible to reverse engineer for this dataset.

Questions:

2.3.1 Plot the data from the 6 events, and give a description of the data as well as the possible challenges associated with modelling the case.

2.3.2 Use what you learned in Part 2.1 and Part 2.2 to formulate a model for the data. You can treat the pumpflow data as an input.

2.3.3 Implement the model in ctsmTMB. Choose just one of the 6 events and estimate the parameters. Report the results. Some tips:

- If the time unit of your model is hours, you should multiply the minutely pumpflow by 60.
- The water volumes are quite "large" numbers. Parameter estimation is easier if the numbers are on the same small order of magnitude. You may try to scale the water data by e.g. 1/1000, if you have difficulties getting ctsmTMB to converge. If you do so, make sure to scale the pumpflow by the same amount.
- If ctsmTMB crashes, it can be a sign that the optimization routine is stuck in an infeasible area. You can make the optimization more robust by forcing a large observation noise (set the lower bound of the observation noise to e.g. 1, if you have scaled the data by 1/1000).

2.3.4 Fit the model to the full dataset. You are free to do any tweaks to the data and to the model structure that you find are necessary to make it work. Report the results.

2.3.5 Discuss how the model could be improved, given that we had enough information and data.

References

Anders Breinholt, Fannar Örn Thordarson, Jan Kloppenborg Møller, Morten Grum, Peter Steen Mikkelsen, and Henrik Madsen. Grey-box modelling of flow in sewer systems with state-dependent diffusion. *Environmetrics*, 22(8):946–961, 2011.