Spring 2015

Modelling rainfall-runoff processes (empirical models)

Stefano Galelli

Singapore University of Technology and Design, stefano_galelli@sutd.edu.sg

40.232 Water Resources Management
Term 7, 2015

Homework 3
Due date: March 23, 2015, at 11.55 pm

Download the file *data_canning_rainfall_runoff.txt*, which contains a complete record of daily precipitation and runoff for the period 1st January 1977 – 31st December 1980 for the Canning River catchment, a major tributary of the Swan River in South-Western Australia. The data columns are organized as follows: rainfall (mm/day) and runoff (m^3/s).

The task is to use these data to identify linear models for the one-day ahead prediction of the runoff in the Canning River. The classes of models that must be considered are Auto-Regressive (AR) and Auto-Regressive eXogenous (ARX) models[1]. The first three years (i.e., 1st January 1977 – 31st December 1979) must be used for calibration, while the last year (i.e., 1st January 1980 – 31st December 1980) for validation.

Part 1 – Preliminary analysis *[15 marks]*

(a) Produce a graph that shows the daily precipitation and runoff through the length of the historical sequence. Then, describe the characteristics of the rainfall-runoff process and explain these characteristics in terms of the relevant climate classification and the physical nature of the catchment.

Part 2 – AR models *[35 marks]*

(a) Produce an auto-correlogram of the runoff process and briefly comment on the result found.

(b) Identify an Auto-Regressive model of order 1 (i.e., AR(1)), and report: 1) the value of the model parameter, and 2) the value of the coefficient of determination (R^2) and Root Mean Squared Error (RMSE) on both calibration and validation period.

(c) Identify an Auto-Regressive model of order 2 (i.e., AR(2)), and report 1) the value of the model parameters, and 2) the value of the coefficient of determination (R^2) and Root Mean Squared Error (RMSE) on both calibration and validation period.

(d) What can we say about the AR(1) and AR(2) model accuracy? Comment on 1) the values obtained for R^2 and RMSE, and 2) a graph that shows the measured runoff against the prediction produced by the two models.

---

[1] For this exercise the models can be identified by working on the raw data (i.e., it is not necessary to normalize and standardize). A brief note on AR and ARX models is given in the last page of this document.

Part 3 – ARX models *[50 marks]*

(a) Produce a cross-correlogram of the rainfall vs. runoff process and briefly comment on the result found.

(b) Identify a proper Auto-Regressive eXogenous model of order (2,1) (i.e., ARX(2,1)), and report: 1) the value of the model parameter, and 2) the value of the coefficient of determination (R^2) and Root Mean Squared Error (RMSE) on both calibration and validation period.

(c) Suppose that you have been asked to improve the performance of the ARX(2,1) model. Which order would you recommend (Hint: you may consider excluding improper models and using the analysis of the cross-correlogram performed at point (a))? Identify the ARX model corresponding to the recommended order, and report: 1) the value of the model parameter, and 2) the value of the coefficient of determination (R^2) and Root Mean Squared Error (RMSE) on both calibration and validation period.

(d) What can we say about the accuracy of the two ARX models? Comment on: 1) the values obtained for R^2 and RMSE, and 2) a graph that shows the measured runoff against the prediction produced by the two models.

Notes on AR and ARX models

The simplest linear model of a runoff process is an Auto-Regressive model of order (n), AR(n), which is formulated as

$$y_t = \sum_{i=1}^{n} a_i \cdot y_{t-i} + \varepsilon_t$$

where y is the runoff process and ε a stochastic (Gaussian) process representing the model error. It follows that an AR(1) and AR(2) model can be written as $y_t = a_1 \cdot y_{t-1} + \varepsilon_t$ and $y_t = a_1 \cdot y_{t-1} + a_2 \cdot y_{t-2} + \varepsilon_t$, respectively.

An Auto-Regressive eXogenous model of order (n_a,n_b), ARX(n_a,n_b), is formulated as

$$y_t = \sum_{i=1}^{n_a} a_i \cdot y_{t-i} + \sum_{i=1}^{n_b} b_i \cdot u_{t-i} + \varepsilon_t$$

where u is the exogenous input (e.g., the precipitation). Note that there is a one-step delay between the exogenous input and the model output. ARX models with this characteristic are known as *proper models*. It follows that a proper ARX(2,1) model can be written as $y_t = a_1 \cdot y_{t-1} + a_2 \cdot y_{t-2} + b_1 \cdot u_{t-1} + \varepsilon_t$.