ECMM124

UNIVERSITY OF EXETER
FACULTY OF ENVIRONMENT, SCIENCE AND ECONOMY
ENGINEERING

January 2023

Hydroinformatics Tools

Module Convenor: Professor Guangtao Fu

Duration: TWO HOURS + 30 minutes upload time

No Word Count

Answer ALL FIVE questions in Section A and any TWO of the three questions in Section B.

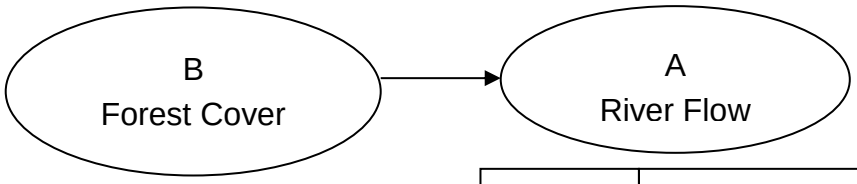
This is an **OPEN BOOK** examination.

SECTION A

Answer ALL FIVE questions in this section.

Question 1 (6 marks)

Consider the Bayesian network with the given conditional probabilities in Figure Q1.



P(B)

Good	Bad
0.7	0.3

A	B: Good	B: Bad
High	0.15	0.9
Low	0.85	0.1

Figure Q1. The Bayesian network and the conditional probabilities

What is the probability distribution for variable *A*? The joint probability of variables *A* and *B* is expressed as $P(A, B) = P(A|B)P(B)$.

(6 marks)

Question 2 (6 marks)

Water usage in a large apartment complex needs to consider localised treatment of greywater that will be re-used within the apartment for cleaning communal areas and watering the surrounding garden areas. To achieve this, two storage tanks will be installed: one for collected raw greywater and one for treated greywater. Assume that greywater is treated at a constant rate.

(a) Sketch a system dynamic model that captures: 1) the inflow of greywater to the raw greywater storage tank; 2) the flow between the two storage tanks; 3) the outflow of treated greywater to meet supply demands; and 4) the overflow from each tank if its storage capacity is exceeded.

(4 marks)

(b) Write a set of IF THEN rules for quantifying the volume of greywater overflow if the greywater storage tank capacity is exceeded.

(2 marks)

Question 3 (6 marks)

A genetic algorithm is used to solve a water distribution system design problem where pipe diameters are optimized. Binary representation is used for encoding decision variables. Pipe 1 has 8 potential diameters to choose from and Pipe 2 has 15 potential diameters.

(a) How many bits should be used to represent the two pipes in a chromosome? Explain how each diameter is represented using binary coding.

(4 marks)

(b) Can real-valued representation be used for pipe sizing in a genetic algorithm? Explain the main reason for your answer.

(2 marks)

Question 4 (6 marks)

Cellular automaton has found many applications in hydroinformatics.

(a) Identify an application of cellular automaton to a particular challenge in urban water management.

(2 marks)

(b) Describe one benefit and one drawback of applying a cellular automaton in the application identified in Q4(a).

(4 marks)

Question 5 (6 marks)

A water company is developing a solution to reduce Greenhouse Gas Emissions (GHG) from its operation. A set of control solutions is obtained using genetic algorithms considering two objectives: operational cost (*f*1) and GHG (*f*2), as shown in Figure Q5. The objective values are normalized into the range [0,1].

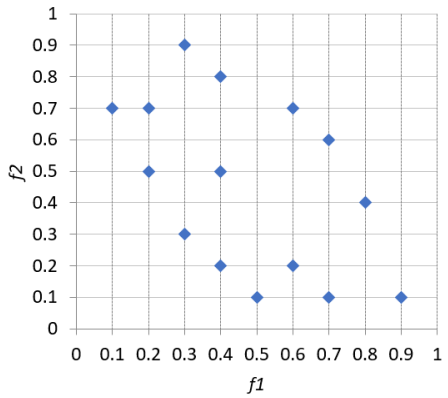


Figure Q5. Solutions from a two-objective optimisation problem

(a) List all the Pareto-optimal solutions using the format (*f1*, *f2*).

(4 marks)

(b) Which solution will you choose to implement if you are the manager and why?

(2 marks)

SECTION B

Answer any TWO of the three questions in this section.

Question 6 (35 marks)

(a) List eight different types of data that are essential to flood modelling, explain how they are used for model configurations, and what GIS format(s) can be used to represent each of them.

(24 marks)

(b) Describe the procedure of using GIS to integrate the above data to build a flood model.

(*11 marks)*

Question 7 (35 marks)

A water company is concerned with the Combined Sewer Overflows (CSOs) in a sewer system. Initial analysis found that the CSO discharge (q) is closely related to the water levels at three upstream manholes, denoted by d_1, d_2 and d_3. A machine learning model is developed to predict CSO events using the three water levels.

(a) Assume a three-layer Artificial Neural Network (ANN) is used as the machine learning model. Draw a diagram to show the architecture of the ANN with the input and output variables clearly explained.

(8 marks)

(b) Assume a single layer perceptron is used as the machine learning model. It has a linear activation function and is trained using the delta rule method. The initial network weights are all set to 0.1 and a learning rate $\alpha = 0.01$. It is observed that a CSO event occurred (i.e., $q > 0$) when the three upstream water levels are $d_1 = 0.4$ m, $d_2 = 0.3$ m and $d_3 = 0.45$ m. Calculate the new network weights using this observation.

(15 marks)

(c) After training with 1000 samples, the machine learning model is tested using a set of 100 samples. The test results are provided in the confusion matrix as shown in Table Q7. CSO discharging represents a positive event and no discharging is a negative event. Calculate the true positive rate, the false positive rate, the precision and the accuracy.

Table Q7. Confusion matrix

	Positive observed	Negative observed
Positive predicted	10	15
Negative predicted	5	80

(8 marks)

(d) If a regression model is developed for the CSO prediction problem, what data are required for training of the model? Explain what you would try to improve the predictive accuracy in the model development process.

(4 marks)

Question 8 (35 marks)

Bio-retention tanks are designed to solve the runoff pollution problem in a city. Three locations (L1, L2 and L3) are considered for constructing bio-retention tanks of varying sizes. The impacts of the three locations on pollution reduction are assessed quantitatively by experts using a scoring system. One unit tank area at L1, L2 and L3 brings a benefit of 20, 10, 12 scores, respectively. L1, L2 and L3 has a cost of £6000, £2000, £4000 per unit area, respectively. The total budget for the three bio-retention tanks is £1,000,000. The total area for L1 and L3 should not exceeds 60 units due to space limitation. A maximum of 30 units can be constructed at L2. The total area at L1 needs to be at least twice the total area at L2 due to inflow and water quality concerns. The areas for the three bio-retention tanks are optimized to achieve the maximum pollution reduction measured by the benefit scores.

(a) Formulate a Linear Programming problem for the optimisation problem described above.

(12 marks)

(b) Solve the problem using the simplex method and provide the detailed calculation steps.

(23 marks)

END OF QUESTION PAPER