NATIONAL UNIVERSITY OF SINGAPORE

CE5310 HYDROINFORMATICS

(Semester I: AY2015/2016)

Time Allowed: 2.5 Hours

INSTRUCTIONS TO CANDIDATES

1. Please write your student number only. **Do not write your name**.
2. This assessment paper contains **FIVE** questions and comprises **FIVE** printed pages.
3. Answer **ALL** questions. All questions do not carry equal marks.
4. This is a "CLOSED BOOK" assessment.

Question 1 [25 marks]

Meteorological Services Division in Singapore has deployed system of sensors at various locations to collect rainfall data and keep records at every 5 minutes. We have obtained two files containing the information of the sensors and the rainfall data for our project. A database should be developed to manage the data. Below are examples of the files content.

C:\data\sensor.txt
Sensor Id, location name, latitude, longitude
1, Jurong West, 1.339340, 103.725407
2, Jurong East, 1.326850, 103.744204
3, Clementi, 1.314267, 103.764528
.....

C:\data\rain.txt
Sensor Id, record_dt, rain
1, 2015-09-07 13:00:00, 5
2, 2015-09-07 13:00:00, 7.5
3, 2015-09-07 13:00:00, 8
.....
1, 2015-09-07 13:00:05, 5.6
2, 2015-09-07 13:00:05, 7.2
3, 2015-09-07 13:00:05, 8.3
.....
1, 2015-09-07 13:00:10, 5.8
2, 2015-09-07 13:00:10, 6.9
3, 2015-09-07 13:00:10, 8.5
.....

(a) List five advantages of using database system to manage data over using files (e.g., text file or Excel files) to manage data.

[5 marks]

(b) Write SQL statements to create two database tables for the data mentioned in files above.

[4 marks]

(c) Write SQL statements to load data from these two files to the two database tables

[4 marks]

(d) Write a SQL statement to find the id and location name of the rain sensors in the following geographical region: latitude in range [1.280104, 1.339560], longitude in range [103.805108, 103.866312].

[5 marks]

(e) Write a SQL statement to find out the minimum, average, and maximum rain data for each sensor each day; the output should be sorted by date first then by sensor id.

[7 marks]

Question 2 [TOTAL: 5 marks]

Deterministic modelling of hydrodynamic flows for practical applications typically requires some assumptions to be introduced in order to simplify the basic conservation equations. List and describe these assumptions.

Question 3 [TOTAL: 25 marks]

Gardens by the Bay has a water system comprising two lakes, Kingfisher Lake and Dragonfly Lake, interconnected by Saraca Stream. Quality of the water in the lakes is studied by a combined monitoring and modelling study. As part of the monitoring study, the nitrogen and phosphorus concentration at a number of catchment inflow locations has been measured.

A GIS of Gardens by the Bay with the following layers is available:

- Water bodies;
- Sub-catchment delineation;
- Drainage infrastructure;
- Water quality monitoring station; and
- Land use.

A GIS of the garden area, water bodies and monitoring locations is shown in Figure 1.

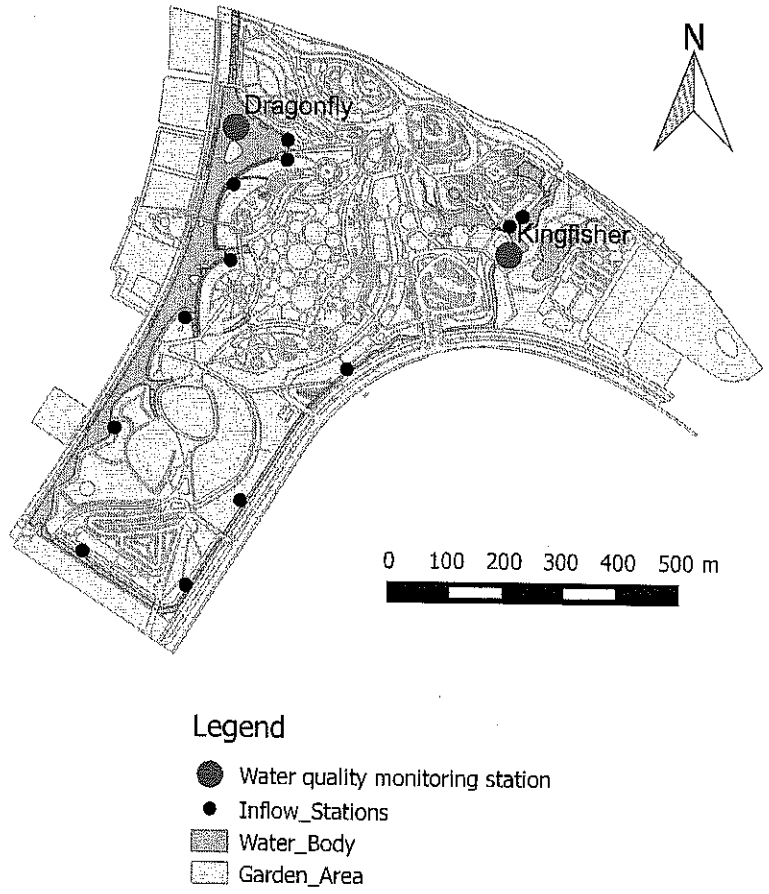

Figure 1: Representation of Gardens by the Bay in GIS.

a.) Except for land-use layer, all layers in this GIS are vector layers. For each of the vector layers, give the most appropriate feature type and possible attributes.

[8 marks]

b.) Gardens by the Bay wants to know which areas contribute most to nitrogen and phosphorus load to the water body. The nitrogen and phosphorus load [gram] of inflow locations indicated in Figure 1 is determined by multiplying runoff [m^3] by concentration nitrogen and phosphorus [gram/m^3] respectively. The nitrogen and phosphorus concentration is known from observations, but no runoff measurements are available. Describe how a GIS system with the above layers could be used to estimate runoff for each of these inflow locations. You may assume rainfall data for Gardens by the Bay is available and constant for the entire area and that each inflow location is associated with one sub-catchment.

Based on constant rainfall data, the total discharge along this area can be estimated and they are distributed by drainages. Dividing by cross-section area, we obtain discharge at inflow locations.

c.) Suppose you would not have sub-catchment delineation and drainage infrastructure available but instead a digital elevation model (DEM) of the Garden area. How would you use the DEM to address the estimation of runoff from the catchment in this case?

[5 marks]

d.) Water quality monitoring in the water body is carried out at only two locations, that is, in the Kingfisher and Dragonfly Lake, as indicated in Figure 1. No water quality data collected in the stream. A deterministic hydrodynamic and water quality model is used to describe water quality in the entire lake system, using the monitoring data for calibration and validation purpose. One of the parameters in water quality model is temperature. Explain how raster interpolation in GIS (e.g. Inverse Distance Weighting) could assist obtaining continuous temperature data in the lake system instead of a water quality model.

[4 marks]

Question 4: [TOTAL 20 marks]

Sketch artificial neural network consisting of:

- 4 neurons in input layer;
- 1 hidden layer with 3 neurons; and
- 1 neuron in output layer.

Assume full connectivity (all neurons in input layer connected to all neurons in hidden layer; all neurons in hidden layer connected to all neurons in output layer)

4.1 How many weights are involved in such articifial neural network? (4 marks)

4.2 Write pseudo algorithm for back-propagation algorithm. (6 marks)

4.3 Describe purpose of training, cross-validation and testing data sets in context of artificial neural networks (5 marks)

4.4 In some cases magnitudes of values in testing data set are larger than values used in training data set. What sort of problems, if any, would you expect in these situations? How would you deal with it?

(5 marks)

Question 5 [TOTAL: 25 marks]

Question 5.1
Write pseudo algorithm for evolutionary algorithm. Describe suitable representations which should be used for:
a) genetic algorithm; and
b) genetic programming.

(10 marks)

Question 5.2
If Genetic Programming is to be used for purpose of rainfall-runoff modelling, please discuss what input and output variables should be used to induce GP relationships.

(7 marks)

Question 5.3
When using GP for rainfall-runoff modelling (Question 5.2) please discuss differences between direct and iterative forecast. Which one of the two approaches would be more suitable and under which circumstances?

(8 marks)

- END OF PAPER -