



IMMC 2020 Greater China Problem A (Autumn) (English 简体 繁體)  
**Smart Water Data Analysis**

### **Background**

Water is essential in our daily life. There is a strong motivation for engineers or researchers to explore the method of constructing a smart water system to use water efficiently. However, for a water supply system, the water leak caused by system malfunctions (such as water pipes vulnerability/valves flaw) is always a big problem. Accordingly, electromagnetic flowmeters are often used for flow measurement and leak monitoring. For example, the difference of the input water flow and the output water flow in a given region can be analyzed in details so that the flow status and leak potentials can be reviewed clearly.

### **Challenges**

Although many data analysis methods are available nowadays, some challenges still exist. Given the flowmeter data which are collected from different flowmeters at the same time, the first challenge is to design a proper general model to understand many flowmeters' data patterns simultaneously. The second challenge is to detect the abnormality of the flowing effectively and efficiently. Finally, the data may be influenced by various noises from the malfunction of the flowmeter itself to the system error mentioned before.

In the tasks of this problem, you need to design a data analysis model to tackle the challenges above. Your team is asked but not limited to do data cleaning on the given data, develop a model for abnormal detection, and optimize the model for final submission.

### **Given the following resources**

- A CSV file that contains the data of the difference of the input water flow and the output water flow in 8 different virtual regions. (Source data: IMMC2020 Autumn Problem A- Dataset.csv)

### **Tasks**

- To analyze the data pattern and figure out the criteria for identifying the abnormality from the given data.
- To develop a general mathematical model to do abnormal detection for the 8 virtual regions based on the result of data analysis.
- To test your model on the given dataset and give a logical explanation of your modeling and the result of abnormal detection.

### **Submission**

- Your solution paper should include a 1-page Summary Sheet. The body cannot exceed 20 pages for a maximum of 21 pages with the Summary Sheet inclusive. The appendices and references should appear at the end of the paper and do not count towards the 21 pages limit.
- Be sure to name the your paper and source code file(s) in your team's Control Number, and you can put your source code file as appendix to your paper. The source code can be written in any programming languages (e.g. Python, Java, and language for Graphic Calculator) but Graphic Calculator and Python are recommended.



IMMC 2020 中华赛 A 题（秋季赛）（English 简体 繁体）

## 智能水务数据分析

### 背景

水在我们的日常生活中至关重要。工程师或研究人员有强烈的动机来探索构建智能水务系统的方法，以提高水的使用效率。然而，对于供水系统而言，由于系统故障（例如水管易损性/阀门缺陷）引起的漏水始终是一个大问题。缘此，电磁流量计常常被用于流量测量和漏水监测。例如，详细分析特定区域的输入水流量和输出水流量的差值，便可清晰评价流量状态和潜在泄漏。

### 挑战

尽管当今有许多数据分析方法可供使用，但一些挑战仍然存在。给定从不同流量计同时收集的流量数据，第一个挑战是设计一个合适的通用模型来同时了解多个流量计的数据模式。第二个挑战是如何有效迅速地检测流量异常。最后，数据可能会受到各种噪声（包括流量计本身的故障以及前面提到的系统误差）影响。

在本问题的任务中，您需要设计一个数据分析模型来解决上述挑战。您的团队被要求（但不限于）对给定数据进行数据清理，开发流量异常检测模型，并在最终的提交中对模型进行优化。

### 给定以下资料

- 一份 CSV 文件，其中包含 8 个不同虚拟区域中输入水流量和输出水流量之差的数据。（数据资源：IMMC2020 Autumn Problem A- Dataset.csv）

### 任务

- 分析数据模式，并建立从给定数据中识别流量异常的标准。
- 根据数据分析结果，开发一个通用数学模型对 8 个虚拟区域进行异常值检测。
- 在给定的数据集上做模型测试，并对建模和异常值检测的结果给出合理的解释。

### 提交

- 您的解决方案论文应包括 1 页的摘要，正文不能超过 20 页，含摘要最多 21 页。附录和参考资料应出现在正文之后，不算在 21 页的限制之内。
- 请确保以团队的控制号命名您的论文和源代码文件，您可将源代码文件作为附录放在论文附录中。可以使用任何编程语言（例如 Python, Java 和用于图形计算器的语言）编写源代码，但建议使用图形计算器和 Python。



IMMC 2020 中華賽 A 題（秋季賽）（English 簡體 繁體）

## 智能水務數據分析

### 背景

水在我們的日常生活中至關重要。工程師或研究人員有強烈的動機來探索構建智能水務系統的方法，以提高水的使用效率。然而，對於供水系統而言，由於系統故障（例如水管易損性/閥門缺陷）引起的漏水始終是一個大問題。緣此，電磁流量計常常被用於流量測量和漏水監測。例如，詳細分析特定區域的輸入水流量和輸出水流量的差值，便可清晰評價流量狀態和潛在泄漏。

### 挑戰

儘管當今有許多數據分析方法可供使用，但一些挑戰仍然存在。給定從不同流量計同時收集的流量數據，第一個挑戰是設計一個合適的通用模型來同時了解多個流量計的數據模式。第二個挑戰是如何有效迅速地檢測流量異常。最後，數據可能會受到各種噪聲（包括流量計本身的故障以及前面提到的系統誤差）影響。

在本問題的任務中，您需要設計一個數據分析模型來解決上述挑戰。您的團隊被要求（但不限於）對給定數據進行數據清理，開發流量異常檢測模型，並在最終的提交中對模型進行優化。

### 給定以下資料源

- 一份 CSV 文件，其中包含 8 個不同虛擬區域中輸入水流量和輸出水流量之差的數據。（數據資源：IMMC2020 Autumn Problem A- Dataset.csv）

### 任務

- 分析數據模式，並建立從給定數據中識別流量異常的標準。
- 根據數據分析結果，開發一個通用數學模型對 8 個虛擬區域進行異常值檢測。
- 在給定的數據集上做模型測試，並對建模和異常值檢測的結果給出合理的解釋。

### 提交

- 您的解決方案論文應包括 1 頁的摘要，正文不能超過 20 頁，含摘要最多 21 頁。附錄和參考資料應出現在正文之後，不算在 21 頁的限制之內。
- 請確保以團隊的控制號命名您的論文和源代碼文件，您可將源代碼文件作為附錄放在論文附錄中。可以使用任何編程語言（例如 Python, Java 和用於圖形計算器的語言）編寫源代碼，但建議使用圖形計算器和 Python。