DYNAMIC COMPENSATION AND SENSOR FUSION FOR A GSM-BASED WATER QUALITY MONITORING NETWORK

by

ZHUO CHEN

B.Sc., Dalian University of Technology, 2013

A THESIS SUBMITTED IN PARTIAL FUFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Mechanical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2016

Abstract

With the increasing demand for water, access to clean water is becoming a more challenging problem for people, in both rural and urban communities. The quantity and quality of fresh water resources, both surface water and ground water, are of major concern worldwide. A continuous water quality monitoring system with access to accurate real-time data can play an important role in water quality tracking and environmental protection. However, evaluation of water quality is complicated; on the one hand, a great number of physical, chemical and biological parameters are usually involved. Hence, multi-sensors network is often deployed for collecting a variety of useful water quality information, such as pH value, ammonia concentration, oxidation-reduction potential, temperature, electrical conductivity, turbidity, and the concentration of dissolved oxygen. On the other hand, objectives of in situ testing are complex and dynamic, and the testing environment in the field is also dynamic and harsh. This thesis develops a wireless data transmission platform to solve the communication problem between the monitoring sensor nodes in the field and the base station. What's more, an individual sensor is only able to make a judgment using a single parameter as evidence. Simplex information is neither sufficient nor reliable, and some parameters also have mutual interference with each other to some extent. Specifically, there should be a systematic way to integrate information from multiple sensors to obtain more accurate and reliable water quality information. Furthermore, allowance has to be made for the variation in the conditions of a sensor, which will affect the sensor accuracy. Therefore, compensation and fusion of sensory data from disparate sources are very necessary to secure a reliable, accurate, and comprehensive monitoring result. By applying Dempster-Shafer theory and Euclidean Distance, this thesis presents a method of assigning four different parameters in the same scale, and combining them into an integrated and reliable quality evaluation result. The necessary methodologies are systematically presented. They are applied to realistic sensory data to illustrate their application and effectiveness.

Preface

The entire work presented in this thesis was conducted at Industrial Automation Laboratory of the University of British Columbia(Vancouver), under the supervision of Dr. Clarence W. de Silva. The system architecture mentioned in Chapter 2 was proposed and designed by me and my lab mate Mr. Teng Li together. Development of the sensor node was inspired by Arduino open-source forum. In this whole work, I was responsible for related literature review, hardware design and installation, open-source software development, data processing algorithm development and implementation, and experiments, which are the majority of the work. Also, Dr. Clarence W. de Silva helped me revise and edit my manuscripts.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADF	Average Data Fusion
ANN	Artificial Neural Network
BPA	Basic Probability Assignment
BS	Base Station
CAU	Central Assessment Unit
CDMA	Code Division Multiple Access
CVDF	Coefficient of Variance Data Fusion
DAQ	Data Acquisition
D.O.	Dissolved Oxygen
FSDF	Fuzzy Set Data Fusion
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HDPE	High-Density Polyethylene
ORP	Oxidation-Reduction Potential
SADF	Self-Adaptive Data Fusion
TCVCXO	Temperature-Compensated Voltage – Controlled Oscillator
VCO	Voltage – Controlled Oscillator
WHO	World Health Organization
WQI	Water Quality Index
WQM	Water Quality Monitoring
WSN	Wireless Sensor Network

Acknowledgments

Studying for my Master's degree is just the beginning of my academic career journey. In the past three years of study and research, not only did I gain a great deal of knowledge, but also I learned to think as an academic researcher: dialectically and rigorously. Besides such achievement, the experience of facing challenges, bottlenecks, hardship, pressure and tears helped me understand what gratitude is.

First, I wish to express my sincere appreciation and gratitude to my supervisor, Dr. Clarence W. de Silva. Thanks to the opportunity he offered me to pursue a MASc. degree and his constant mentorship and supervision, I was able to overcome many difficulties and obstacles in the past three years. To me, more than as an academic supervisor, Dr. de Silva is also my life mentor, like a kind grandfather, being very nice and generous, always encouraging and understanding me. Therefore, the particular acknowledgment should be made here to Dr. de Silva, being his student is my proudest achievement so far.

Furthermore, my appreciation should be given to another mentor of mine, Dr. Min-Fan Ricky Lee for his recommendation and suggestion. Though I am not studying in his lab anymore, he is always supporting my life and research in UBC.

Moreover, I wish to thanks all my colleagues in the Industrial Automation Laboratory (IAL), Dr. Haoxiang Lang, Mr. Teng Li, Ms. Yu Du, Mr. Shan Xiao, Mr. Shujun Gao, Mr. Min Xia and so on for their friendship and help.

Finally, I would like to express my best gratitude to my mother and father who have given me life and always supported me in my goals.

Chapter 1: Introduction

1.1 Motivation

Water is one of the most essential components for sustaining human life. The quality problem of water has increasingly attracted attention not only from experts of the area, but also from the general public. Maintaining a good water quality level in lakes, streams and rivers benefits both humans and aquatic ecosystems, and failing that, any imbalance in the water quality will severely affect human health and the aquatic ecosystems. To properly determine the condition of water, a monitoring system is necessary. Water sources are found everywhere, but are usually located far from urban areas, which presents many difficulties for water quality monitoring. Nonetheless, monitoring is a very important way to detect the water quality level and trends, and requires an excellent monitoring system that will respond fast and accurately[1].

Three primary means for water quality monitoring are available: discrete, mechanical and automated. Discrete monitoring is a conventional technique in which experts have to travel to the field where the water source is located, bring samples back to the laboratory, and then determine the water quality by analyzing the samples, as shown, for example, in Figure 1.1.

Figure 1.1 Sampling in the field[2]

An advanced method of water quality monitoring is to deploy measurement devices at the site or portable devices are applied[2], so scientists do not have to travel to the field to collect samples. They only go to the field to record data from devices, as shown for example in Figure 1.2.

Figure 1.2 Reading of device data in the field[3]

Both these two methods are inconvenient; they consume considerable time and require unnecessary labour due to the need for frequent travel between a laboratory and the field sample sites, sometimes carrying massive equipment. Furthermore, experts risk personal safety, travelling to a wild area like a rainforest.

Thanks to the development of automation technologies, automated sensors can be placed in a body of water, for continuous measurement of water quality parameters, so groups of data collected on-site can be easily sent back to the Base Station (BS) by a Wireless Sensor Network (WSN)[4]-[8]. Figure 1.3 illustrates the structure of a WSN: sensor nodes are expressed by dots. In the broken line circle, central dots are primary nodes, which execute duties such as transmitting data to the gateway, while surrounding dots are secondary nodes, which provide continuous measurement of parameters. All the data are transmitted through a gateway to the BS via a GSM Network. The data are processed separately at the BS, and then the refined data are sent to the laboratory - Central Assessment Unit (CAU) as packages for further analysis.

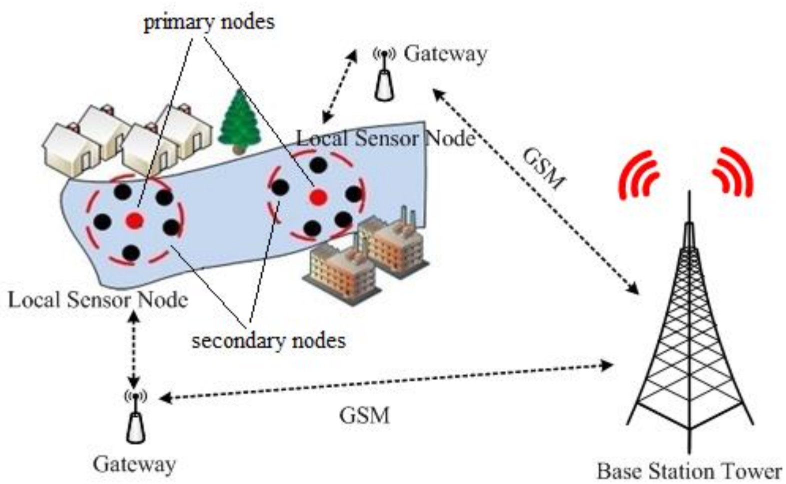

Figure 1.3 Architecture of WSN

A cheap, effective and efficient WSN for a water quality monitoring system was proposed[9] as a solution to replace sparse and manual sampling for underground and surface water in 2009. In this system, Code Division Multiple Access (CDMA) communication technology was applied to handle data transmission[10]. This is a real-time system and it has a simple structure, but plenty of sensors are used. Readings are taken by these sensors and transmitted through a CDMA wireless network.

However, a comprehensive monitoring system is complicated, and hardware development alone is not enough. Even when implemented with high-level equipment, the work will not be complete without data processing. Most water quality sensors in the market provide readings for one single parameter, and an individual sensor is not as reliable as multiple sensors. Therefore, sensor fusion (multi-sensor data fusion) should improve the reliability and accuracy of the end result. A multi-sensor collaboration would be able to enhance the system fault tolerance and increase its accuracy and plausibility[11][12]. There are a number of approaches to combine sensors or sensory data together, most of which are derived from probability theory. Sensor fault diagnosis using self-adaptive data fusion can help decrease interference from faulty sensors[13]. This is an application that can be achieved by integrating parameters that represent same properties, like fusing pH data and D.O. data together. A comparison of pH sensor fusion among Average data fusion (ADF), Self-adaptive data fusion (SADF), Fuzzy set data fusion (FSDF), and Coefficient of variance data fusion (CVDF) is proposed[14], which may provide data fusion results that are superior to a single failed sensor and the average data fusion.

Some researchers have found another type of fusion: combining two or more data sets from different sensing sources with different properties. Based on Dempster-Shafer (D-S) evidence theory, a multi-sensor information fusion method is used for integrated circuit diagnosis[15]. Through the judgement regarding individual temperature sensor or voltage sensor and the judgement depending on temperature-voltage combined value, the comparison indicates fusion data gives a more accurate and dependable fault component diagnosis, limiting the possibility of wrong judgment to a highly significant level. D-S theory has also been applied widely in other areas[16][17]. Also, to evaluate water quality using D-S theory has long been discussed and proposed[18], which presents this thesis with significant inspiration in the sensor fusion area.

1.2 Research Scope and Problem Specification

1.2.1 Water quality parameters

Water covers about 70% of the Earth's surface. In our daily lives, precipitation, household and industrial discharge, and ground-water are main sources of the rivers from which we obtain water for daily use. Figure 1.4 is a simple water system diagram.

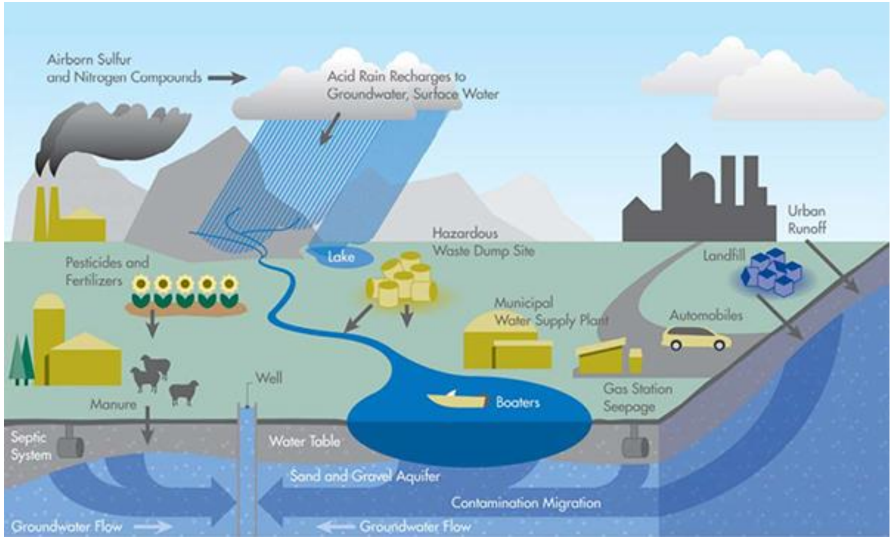

Figure 1.4 A schematic view of a general water system[19]

When measuring or evaluating the quality of a water source, many parameters may be involved. Each parameter describes the quality aspect of a specific measurement, which could be physical, chemical or biological. Temperature, turbidity, and conductivity are physical characteristics. pH value, dissolved oxygen level, oxidation-reduction potential, and nitrate and phosphate concentrations are usually measured chemical properties. Biological factors generally describe the presence of some microorganisms like E.coli, some algae or phytoplankton contained in the water.

This thesis mainly focuses on the development of a remote monitoring system and application of sensor fusion for water quality monitoring, from a viewpoint of mechatronics. It does not undertake to provide an extensive description of the underlying chemical and biological principles. In present research, five basic water quality parameters are considered in realizing the monitoring system: temperature, pH value, dissolved oxygen level, electrical conductivity, and oxidation-reduction potential. These are outlined next.

Temperature

In a biological system or an electro-mechanical system, temperature is a key indicative parameters of proper functioning. It is a very basic and common measurement. Similarly, temperature plays an important role in determining water quality. For example, it determines the kinds of biological species that can survive in a water body. Some species will die outside restricted temperature limits. Aquatic animals thrive only in a suitable temperature, which is very important in aquaculture. Temperature can also affect the chemical component ratio in water. For example, warm water holds less oxygen than cold water. The physical structure of some electrical sensors will alter with temperature and this will lead to sensing error.

pH Value

pH value is a chemical parameter that is measured to determine a solution's acidity. This determination depends on the concentration of hydrogen ions (H+) and hydroxyl ions (OH-). In water, a small number of water molecules (H_2O) will break into hydrogen ions (H+) and hydroxyl ions (OH-), but these two ions still stay in a balanced amount. When other compounds are dissolved in the water, they may react with the hydrogen (H+) or hydroxyl ions (OH-), causing an imbalance between these two ions. If more hydrogen ions (H+) react with the compounds, then more hydroxide ions (OH-) are left in solution and the water is basic (alkaline). Similarly, if more hydroxide ions (OH-) react, more hydrogen ions are left and the water is acidic. As the “potential hydrogen” or pH value is a measure the molar concentration of hydrogen ions (H+), and it is a numerical value of acidity[20] [21].

pH value is defined as the negative logarithm of hydrogen ion (H+) concentration. If the molar concentration of H+ in a solution is 10^{-5} mol, the pH value of this solution is 5 (-log 10^{-5}). The total amount of ions in the water is 10^{-14}[22]. Because of this, pH value ranges from 0 to 14, and the higher the concentration of (H+), the lower the pH value, and the more acidic the solution is. On the contrary, the higher the OH- concentration, the higher the pH, and the solution is more basic. Figure 1.5

illustrates a pH scale and presents some common solutions and their pH values. Neither high nor low pH value in water is good, and slightly basic water is better for humans.

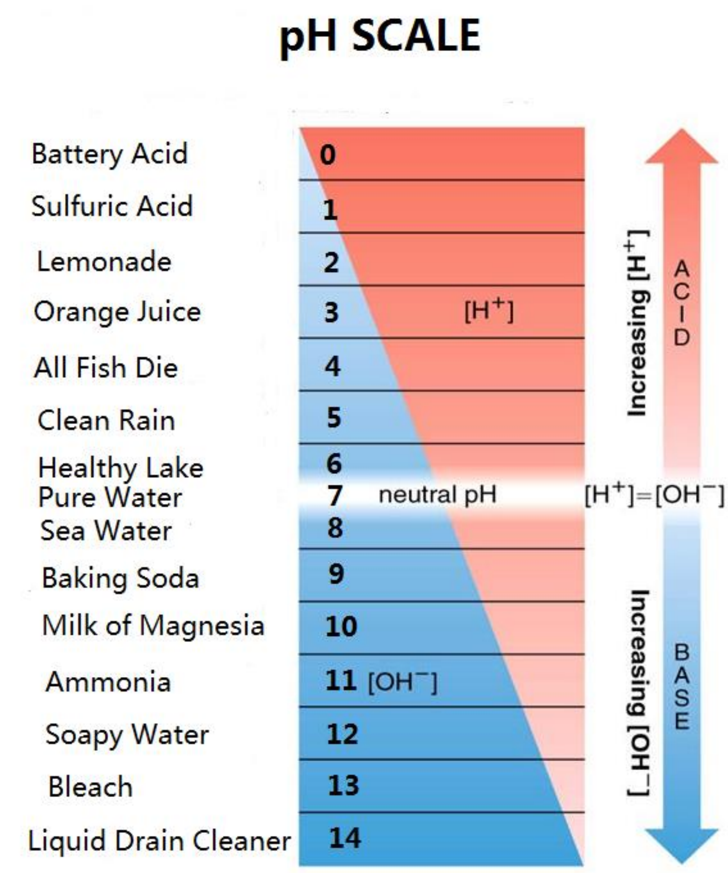

Figure 1.5 pH scale and pH value of some common liquids

Many man-made or natural factors can affect the pH value in water. Natural factors like precipitation and even some minerals in surrounding rocks will make the pH value to fluctuate. Some man-made issues such as industrial wastewater and daily discharge from human activities are main factors for this.

Electrical Conductivity

Electrical conductivity (commonly called "conductivity" when there is no confusion with thermal conductivity) is a measure of the capability of a solution such as water to pass an electric current. This ability is directly related to the concentration of any dissolved electrolyte ions in the water, although it does not identify the specific ions in the water[20]. The more ions the solution contains, the higher the conductivity, so any external compounds dissolved in water may lead to a corresponding increase in conductivity, which could be the result of polluting discharges entering the water. A creek will have a baseline conductivity depending on the local geology and soils. Higher conductivity will result from the presence of various ions including nitrate, phosphate, and sodium.

Conductivity is usually measured in units of µS/cm. Distilled or deionized water has a very low conductivity value, usually ranging from 0.5 to 3 µS/cm, and most streams or rivers have a range from 50 to 1500 µS/cm[20]; however, a suitable value is in the range 50 - 500uS/cm for humans and aquatic animals. As one of the early indicators of water quality change, most water bodies keep a relatively stable conductivity value. Any significant change in conductivity could be due to detrimental ingestion into the water body.

Dissolved Oxygen (D.O.) Level

Dissolved oxygen refers to oxygen gas molecules (O_2) present in the water. Plants and animals cannot directly use the oxygen that is part of the water molecule (H_2O), but instead depend on the dissolved oxygen for respiration. Oxygen enters streams from the surrounding air and as a product of photosynthesis from aquatic plants. Consistently high levels of dissolved oxygen are desirable for a healthy ecosystem. That means healthy waters generally have high levels of D.O. (areas like a swamp have a low-level D.O.).

The level of dissolved oxygen varies depending on such factors as water temperature, time of day, season, depth, altitude, and rate of flow. Water at higher temperatures and altitudes will have less dissolved oxygen. Dissolved oxygen reaches its peak during the day. At night, it decreases, as photosynthesis has stopped while oxygen consuming processes such as respiration, oxidation, and respiration continue, until shortly before dawn.

Human factors that affect dissolved oxygen in streams include addition of oxygen-consuming organic waste such as sewage, addition of nutrients, changing the flow of water, raising the water temperature, and the addition of chemicals. The level of dissolved oxygen is measured in mg/L.

Oxidation-Reduction Potential (ORP)

ORP stands for Oxidation-Reduction Potential (also known as "Redox Potential"). It is a measure of water's ability to oxidize contaminants. ORP also can be described as a chemical or electrical "potential" energy stored in water that is ready to be put to work. In electrical terms, the potential can be measured and expressed in mV. Actually, it measures the presence of oxidizing or reducing agents through their specific electrical charge.

ORP has both positive and negative values. The higher the ORP value, the more oxidizing agents there are. A substance with a positive ORP is considered an oxidizer, and high ORP damages cells and increases the aging process. A solution with a negative ORP is considered as an anti-oxidant or reducing agent; it inhibits or slows the process of oxidation, and also slows the aging process.

Measuring ORP is a simple way to monitor the effectiveness of a sanitizer or the quantity of anti-oxidants in a water body. For humans, generally a higher ORP is better for outside of the body, while a lower ORP is preferred for consumption due to the high anti-oxidant value.

There are some specific optimal ORP values for different applications. For example, according to the World Health Organization (WHO), the minimum ORP for pool & spa disinfection is +650mV and generally a swimming pool is maintained at a value around +450mV. For drinking water, anything below -550mV is considered too strong and not recommended, although the WHO has not set a standard. In North America, the ORP of most tap water is from +200 to +600 mV (i.e., these waters are oxidizing agents). Typical aquaculture requires an ORP value between 150 and 250mV[23]. High pH value ionized water has a negative ORP and so it is a reducing agent. Most bottled water in the market is acidic at low pH value, and some of them are even quite acidic and they also have a higher ORP (over +400mV).

1.2.2 Problems in Water Quality Monitoring (WQM)

For environmental reasons, some WQM activities are conducted in suburban areas, and sensor nodes are deployed in scattered locations[24]. Data transmission then becomes a crucial part of the monitoring system. However, conditions in remote places may not be conducive to a good communication network. Difficulties like these have impeded the development of remote monitoring, and most wireless monitoring platforms use GPRS (General Packet Radio Service) to transmit data in the urban part of a river or lake in city, in order to find a strong signal.

Multiple sensor nodes with wireless transmission function can be rather costly. Applying numerous costly sensing devices to realize multiple location monitoring becomes a challenge.

Evaluating the quality of water always poses a complex calculation with plenty of data and parameters. Unlike a temperature value, it is not practical to have the public understand the terminology of various sensory parameters and assess the water quality by themselves. The British Columbia Ministry of Environment, Lands and Parks and the Alberta Environment of Canada have developed a method to present an integrated water quality index for public convenience since 2001. There are three functions involved in this Index:

(1) Scope (F1): the number of variables not meeting water quality objectives,
(2) Frequency (F2): the number of times these objectives are not met,

(3) Amplitude (F3): the amount by which the objectives are not met.

Assessment results are presented in grades from 0 to 100, and also categorized into 5 descriptive quality levels: Excellent, Good, Fair, Marginal, and Poor[25]. This method statistically considers the number of testing times and the parameter amounts which are not satisfied. It is a compromising method that does not take the difference of each sensor's contribution into account. Although failed data is very limited, much information may be lost, thus a direct way to integrate data properly is a challenge that has to be overcome.

1.3 Related Work

1.3.1 Data acquisition (DAQ)

Data acquisition is a critical process of a monitoring system. Physical variables are measured and converted into digital numeric values. This process includes sensors, signal conditioning circuitries, and analog-digital converters[26]. Figure 1.6 presents the key processes of DAQ.

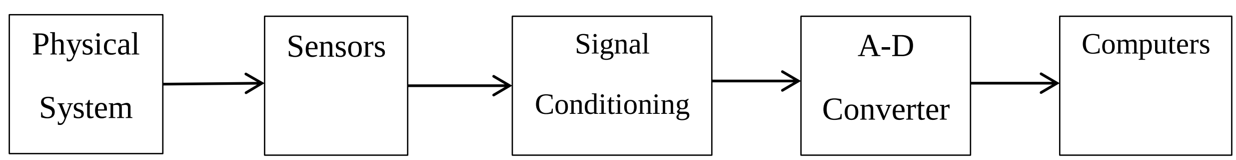

Figure 1.6 Data acquisition process

As an input segment of a monitoring system, data acquisition requires both hardware and software to perform the necessary functions.

1.3.2 Data compensation

Basically, electrical sensors are developed using physical principles and express various physical quantities with voltage or current as an output. Function of sensors may be disturbed by the external environment especially temperature, humidity and pressure. Errors appear and the accuracy drops when sensors do not function in a proper way. Compensation is an approach to reduce or remove undesirable effect of a changing environment in a sensor reading[27].

Compensation has been applied in a wide range of electrical, mechanical, and civil engineering applications. A voltage-controlled oscillator (VCO) is an electronic oscillator controlled by voltage input. However, the associated piezoelectric effect is affected to some extent by temperature. A temperature-compensation version (TCVCXO) has been developed, and frequency drift is corrected[28]. In cameras, motion compensation is used to keep a picture or video steady by predicting and correcting a frame of a picture or video.

Temperature is a common factor that causes error in an engineering device. Generically, temperature compensation is the adjustment in performance of a system to compensate for changes in temperature to keep the system operating normally[29]. This kind of compensation is a broad subject and is accomplished in many ways. In electronics, it usually takes the form of a temperature-sensitive component (e.g., thermometer) either connected as part of the active circuit or as part of a feedback loop.

Ideally for a strain gauge, the resistance changes only in response to the strain applied. However, due to reasons like the material, or the lead wiring, it also responds to changes of temperature. In order to reduce the material's thermal sensitivity, the gage material is processed to adapt to thermal expansion, even though the temperature shift still persists. Additionally, by using half-bridge and full-bridge configurations in the strain gage, temperature changes become identical in the gages, and the ratio of resistances stays constant. Errors caused by temperature effects can be easily removed by this method.

1.3.3 Wireless transmission

Technologies of wireless communication have made remote monitoring practical[30]. Data achieved by nodes located at a long distance, even worldwide, is transmitted back to the central unit through global networks. GPRS is a very popular method to send data, and it takes place through 3G and 4G in advanced monitoring systems that require high data upload and download speeds.

However, high-speed transmission needs a high-quality wireless signal to be secured. Basic plain text transmission works well in poor signal strength areas, with reduced speed as a compromise.

1.3.4 Sensor fusion

Development of a multi-sensor system involves not only data acquisition, but also transmission and particularly applying an appropriate sensor fusion technology. The integration of different types of information about sensing objectives, some of them even from different dimensions, usually requires complex mathematical methods.

Sensor fusion involves integrating data from multiple sensors to generate a more reliable and accurate result. Since it was first proposed in the literature in the 1960s, there have been many approaches to realize sensor fusion. Bayesian networks, Kalman filter, fuzzy logic, artificial neural network (ANN) and also Dempster-Shafer evidence inference theory are the popular algorithms of sensor fusion. Achieving an optimizing value by utilizing each sensor's advantages and avoiding its shortcomings is the ultimate target of sensor/data fusion[31].

Practical applications of sensor fusion have merged into many areas, for example, robotic obstacle avoidance and path planning, pattern recognition, and fault diagnosis in maintenance engineering.

1.4 Contributions and Organization of the Thesis

The main contributions of the present thesis are listed below.

1. A wireless measuring platform with five water quality transmitters is proposed and developed as one of the nodes in a water quality monitoring network. With sensors (probes) for temperature, pH value, dissolved oxygen (D.O.) level,

electrical conductivity, and oxidation-reduction potential, the platform is able to measure five basic water quality parameters simultaneously. As a wireless node, readings from each transmitter are conveyed to a base station (BS) through a GSM network in the plain text form. Further data processing work is conducted at the BS.

2. Dynamic temperature compensation algorithms for pH and D.O. probes are implemented into the proposed sensing platform. Every time the pH and D.O. sensors take readings, first an instant temperature value from the thermometer is held as a reference temperature and will be substituted into the pH and D.O. value calculations. Reading errors caused by temperature changes are corrected with such compensation steps and sensing accuracy of the platform is improved.
3. By using Dempster-Shafer evidence inference theory, a data fusion method for four different water quality parameters is posed. Water quality levels are divided into six categories from A-F of which A indicates the best quality. Each sensor has its own recognition of the quality level regarding its readings, and a plausibility of sensors is calculated by applying Euclidean Distance theory. Finally, through the comparison of probabilities, an integrated evaluation of water quality using four sensors is secured.

The organization and contents of each chapter of rest of the thesis are summarized below.

Chapter 2 proposes the overall monitoring system architecture and introduces the associated hardware. Sensors applied in the system are introduced individually by explaining their working principles and also specifications. The microcontroller for data acquisition is introduced and the mounting methods are presented.

Chapter 3 demonstrates the way by which sensors take readings from a viewpoint of mathematics. The related equations are listed to express the relationship between a water quality parameter value and a measured value (current/voltage). Then, the temperature compensation methods for the pH and D.O. probes are presented.

Chapter 4 introduces the wireless transmission segment of the system. Analysis and comparison of several popular wireless transmission approaches are presented. Moreover, explanation of SMS data transmission, which is applied in the system is presented.

Chapter 5 mainly presents the sensor/data fusion theory. It explains the Dempster-Shafer evidence theory and calculation for belief function using Euclidean distance. The combination of data is detailed step by step mathematically. Experimental results and judgement criteria are discussed, which verify the reliability of fusion.

Chapter 6 concludes the thesis by summarizing the overall research. The main contributions of this work are pointed out. Limitations of the work are indicated. Possible directions for future work to solve the existing problems are suggested as well.

Chapter 2: System Architecture

2.1 System Overview

Water quality monitoring takes places in a geographically distributed manner. Sensor nodes are distributed in different locations and connected to a base station through a wireless network as Figure 2.1 shows.

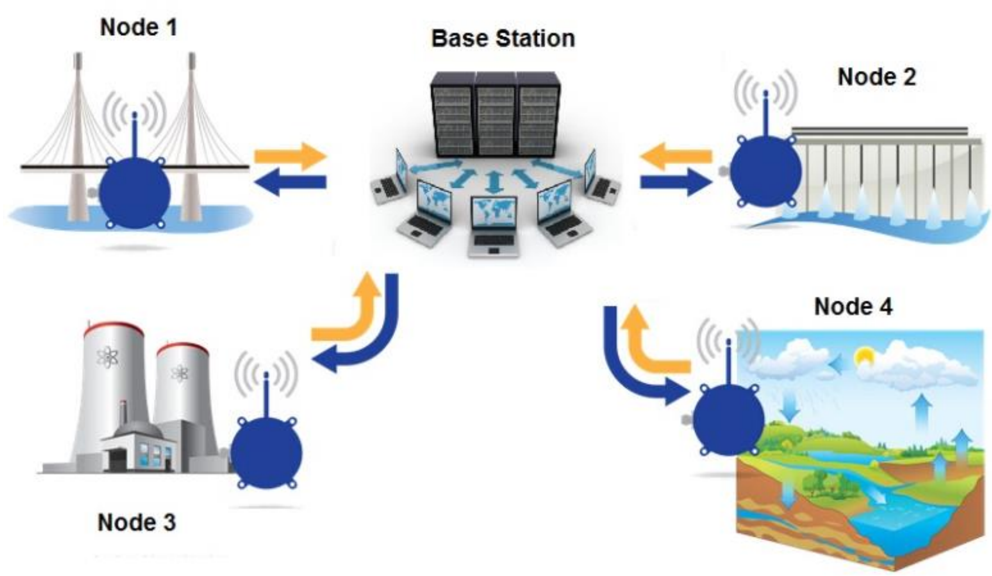

Figure 2.1 Wireless Water Quality Monitoring Network

The Base Station (BS) is a place where all data from each node in its control region is gathered. Sensor nodes measure water quality parameter values and handle a part of preliminary data processing work like signal processing and compensation. Further information processing like data fusion and analysis is carried out at the base station in order to reduce the power consumption of the sensor nodes. A GSM network is applied to carry out communication between the nodes and the base station. This thesis mainly focuses on sensor node development and data fusion.

2.2 Sensor Characterization

A total of five sensors are employed in the present work, constituting the data input layer of the monitoring platform. A general specification summary of these sensors is given in Table 2.1.

Table 2.1: Specifications of sensors

Spec. \ Sensor	Temp	pH	D.O.	Conductivity	ORP
Time Constant	750ms	378 ms	650 ms	1 s	1 s
Impedance	2000ohm	1650 ohm	702 ohm	165 ohm	227.6 ohm
Resolution	0.0625 ℃	0.01	0.01 mg/l	0.1 μs	0.1 mV
Accurate Reading Range	-55 ℃ to +125 °C	0.01 – 14.00	0ml – 20ml	10μs to 1s	-1019.9 mV to +1019.9 mV
Compensation	--	Temperature	Temperature	--	--
Pin Number	D8	D14(TX), D15(RX)	D16(TX), D17(RX)	D18(TX), D19(RX)	D10(TX), D9(RX)

2.2.1 Temperature probe

A One-Wire Waterproof Digital Temperature Sensor (Figure 2.2) is a direct-to-digital temperature sensor. The default resolution at power-up is 12-bits, corresponding to 0.0625 ℃. It powers up in a low-power idle state. To initiate a temperature measurement and A to D conversion, the master must issue a Covert T [44h] command. Following the conversion, the resulting thermal data is stored in the 2–byte temperature register in the scratchpad memory and then the sensor returns to the idle state, waiting for commands.

Figure 2.2 One-Wire Waterproof Digital Temperature Sensor

2.2.2 pH probe

An Atlas Scientific pH probe (Figure 2.3) is a passive device that detects a current generated from hydrogen ion activity. This current (which can be positive or negative) is very weak and cannot be detected with a multimeter or an analog to digital converter. The current that is generated from the hydrogen ion activity is the reciprocal of that activity and can be predicted using:

$$E = E^0 + \frac{RT}{F}\ln(\alpha_{H+}) = E^0 - \frac{2.303RT}{F}pH \tag{1}$$

where E^0 is a reference voltage, R is the ideal gas constant, T is the temperature in Kelvin, and F is the Faraday constant.

Figure 2.3 Atlas Scientific pH probe

2.2.3 Dissolved oxygen probe

An Atlas Scientific D.O. probe (Figure 2.4) is a galvanic dissolved oxygen sensor and is a passive device that generates a small voltage from 0mV to 47mV depending on the oxygen saturation of the high-density polyethylene (HDPE) sensing membrane. This voltage can easily be read by a multimeter or an analog to digital converter. The probe is a tube with a zinc rod (anode) submerged in an electrolyte. The sensing element is the HDPE sensing membrane compressed against a silver disk (cathode). There are many factors that must be taken into account when reading dissolved oxygen, such as salinity and temperature. Therefore, there is no simple linear equation that describes the dissolved oxygen from the probe's output voltage, and dissolved oxygen can be expressed in mg/L or in % as a saturation percentage.

Figure 2.4 Atlas Scientific D.O. probe

2.2.4 Oxidation-reduction potential probe

An Atlas Scientific ORP probe (Figure 2.5) is a passive device that can detect very small voltages generated in ozonized water. The electrode is made of a material like gold or platinum, and electrons are released from this material to the oxidizer. A very weak voltage is generated during the electron migration and it is compared to the reference electrode in a silver salt solution. The more the available oxidizer, the greater the voltage difference between the solutions.

Figure 2.5 Atlas Scientific ORP probe

2.2.5 Conductivity probe

An Atlas Conductivity probe (Figure 2.6) is a simple measuring device of electrical conductivity. It has two conductors with a fixed surface area at a fixed distance from each other. This distance and surface area are known as the conductivity cell. The cell's distance and surface area are quantified as the conductivity cell *K* constant. Probes with different *K* constants have different measuring ranges. According to the testing objective of this thesis, a conductivity probe with K=10 is used.

Figure 2.6 Atlas Scientific Conductivity Probe

2.3 Data Acquisition

An Arduino UNO board is a 16 MHz open-source microcontroller for an embedded system, as shown in Figure 2.7. It runs on the ATmega328 core with 32kB memory, which is a low energy consumption chip. There are 14 digital I/O pins and 6 analog inputs on the board to provide a total of 20 available pins for the 5 sensors that are used. In addition to the property of enduring a severe environment, it also has to meet the needs of the field measurements. Figure 2.8 illustrates the pin assignment of sensors and other development modules.

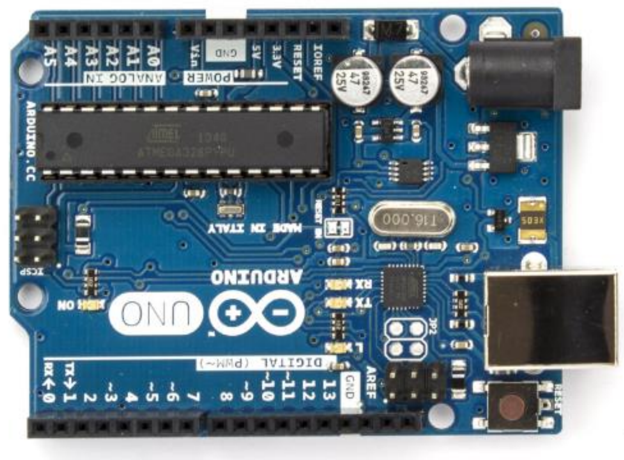

Figure 2.7 Arduino UNO board

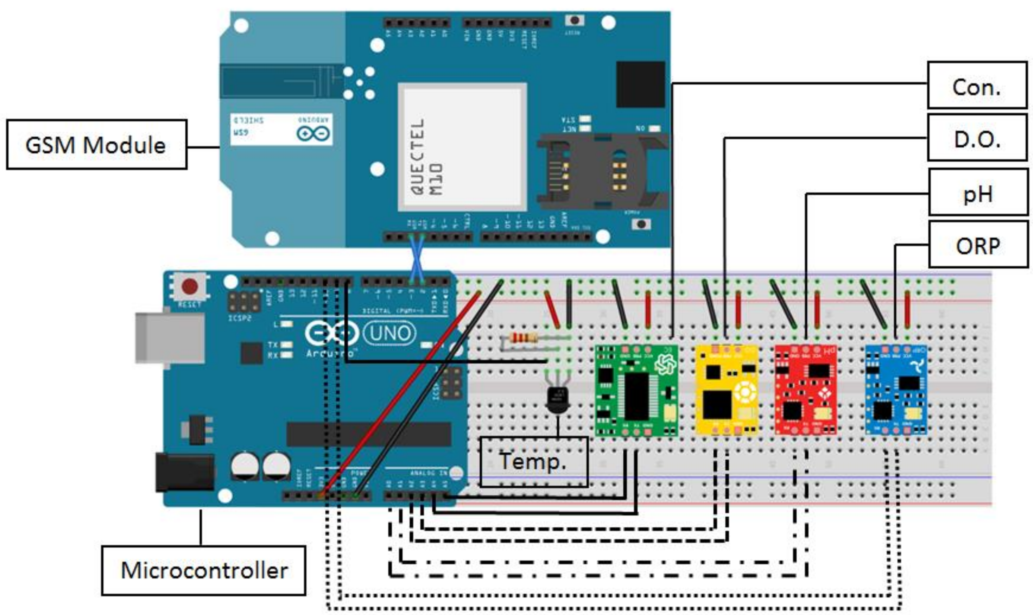

Figure 2.8 Layout of logical connections

Digital PIN2 (TX) PIN3 (RX): GSM: Quectel M10

Digital PIN8: One-Wire Digital Temperature Sensor (Black)

Digital PIN18 (TX) PIN19 (RX): Atlas Scientific Conductivity Sensor (Green)

Digital PIN16 (TX) PIN17 (RX): Atlas Scientific Dissolved Oxygen Sensor (Yellow)

Digital PIN14 (TX) PIN15 (RX): Atlas Scientific pH Sensor (Red)

Digital PIN9 (TX) PIN10 (RX): Atlas Scientific ORP Sensor (Blue)

POWER: Arduino UNO Board – 9V External Power

Sensors – 3.3V Arduino Power Supply PIN

GSM Module – 5V Arduino Power Supply PIN

Figure 2.9 shows the hardware layout and appearance of a sensor node. A signal strengthening antenna is fixed on the top of the case. The five probes are seen on the bottom of the unit.

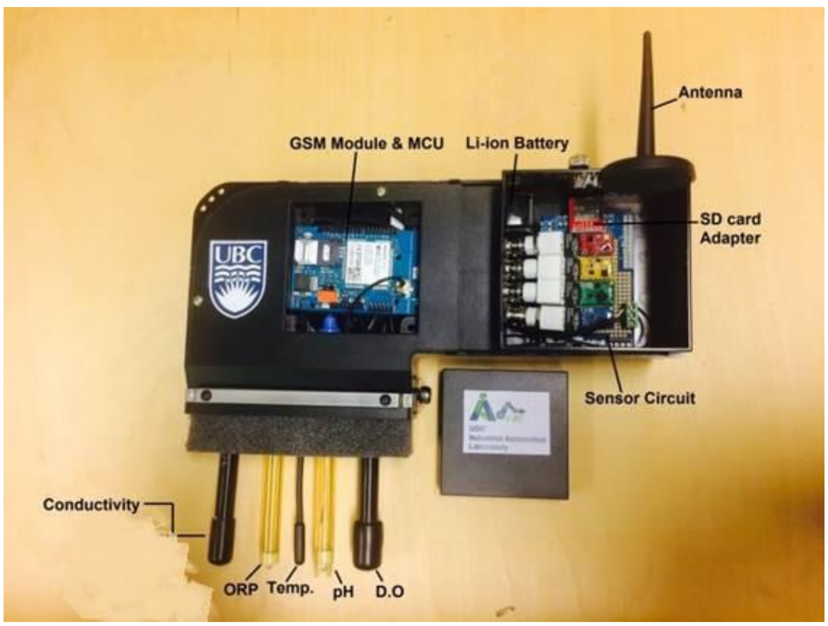

Figure 2.9 Sensor node

Chapter 3: Measurement and Compensation

3.1 Measurement

3.1.1 Temperature measurement

Each Dallas One-wired temperature sensor has its own physical address (hexadecimal number), with which data is transmitted between Arduino and the sensor. Through the “one_wire_address_finder” program, the address of the sensor is identified as “0x28, 0x91,0x95, 0x06, 0x06,0x00,0x00,0x25”. The following code is used for the sensor setup:

```
//Temperature sensor setup

#define ONE_WIRE_BUS 8    // set pin8 on the Arduino for temperature sensor

OneWire oneWire(ONE_WIRE_BUS); // Setup a oneWire instance to communicate with any OneWire device

DallasTemperature sensors(&oneWire);    // Pass our oneWire reference to Dallas Temperature

DeviceAddress insideThermometer = 0x28, 0x91,0x95, 0x06, 0x06,0x00,0x00,0x25}; // Assign the addresses of this sensor.
```

Two libraries must be added for the temperature sensing coding, one is “DallasTemperature.h”, and the other is “OneWire.h”. Sensor resolution is set to be 10 bit by “sensors. SetResolution (insideThermometer, 10).

3.1.2 pH value sensing

The pH value describes the amount of H+ cation $[H^{+}]$ in a solution. The concentration of these ions can change in a very wide range - most often it has values lying somewhere between 1 million and 1014 million, although sometimes even higher and lower concentrations can be observed. Equation (2) mathematically gives the definition of pH.

$$pH = -\log(H^{+}) \tag{2}$$

This definition is feasible in simple measurement of diluted solutions. In a general case, instead of the concentration of H+ cation $[H^{+}]$, the activity of H^{+} cation in the given solution a_{H+} (in mol/L) is used for more precise readings. The actual behavior of the ions in the solution depends not on their concentrations, but on activities, which cannot be ignored in a concentrated solution. Therefore, an improved form of equation (3) is applied in the pH sensing, which is called “thermodynamic pH” given by

$$pH = -\log(a_{H+}) \tag{3}$$

According to the Nernst equation for H^{+} we have,

$$E = E_O + \frac{RT}{F}\ln(a_{H+}) \tag{4}$$

which describes the relationship between electrochemical potential and concentrations of ions. A direct relationship between voltage and pH value is given by

$$E = E_O - 2.303\frac{RT}{F}pH \tag{5}$$

where 2.303 is the conversion factor of the natural logarithm to a decimal logarithm, R is the ideal gas constant, T is the Kelvin temperature of the solution, F is the Faraday constant, and E, E_0 express the output voltage and the reference voltage respectively.

Usually, for a rough sensing, the temperature T in the Nernst Equation is fixed at room temperature of 298K (25 ℃). In this case, the equation could be derived to:

$$E = E_O - 0.0592\,pH \tag{6}$$

This simplified equation presents a linear relationship between pH value and output voltage, and as Figure 3.1 shows, the slope of the pH electrode is 59.2mV/pH unit at

25 ℃. Fixing the temperature is simple, convenient and reasonable for an indoor test. However, when it comes to a complex and fluctuating natural environment, the temperature variable must be taken into consideration.

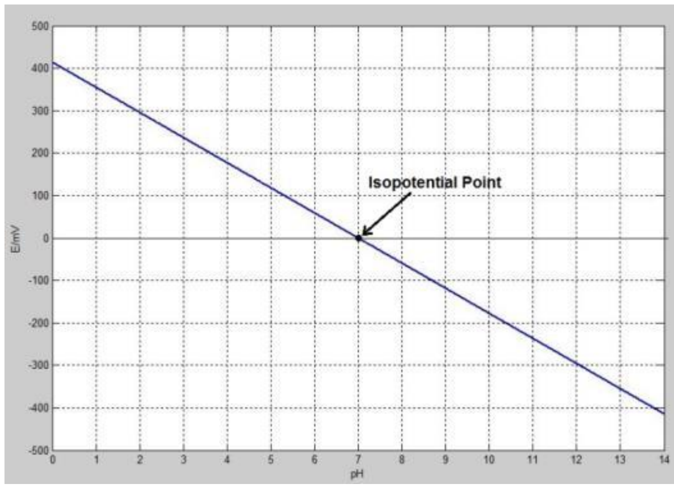

Figure 3.1 Ideal curve of pH value vs. output voltage

3.1.3 Dissolved oxygen measurement

Like the working principle of a pH sensor, though more complicated, when a D.O sensor takes a reading, the dissolved oxygen in the water that is being tested diffuses across the membrane at a rate proportional to the pressure of the oxygen in the water. The dissolved oxygen is then reduced and consumed at the cathode. This reaction produces an electrical current that is directly related to the oxygen concentration. This current is carried by the ions in the electrolyte and runs from the cathode to the anode. As this current is proportional to the partial pressure of oxygen in the test sample, it can be calculated by the following equation:

$$I_d = (4 \times F \times P_M(T) \times A \times P_{o2})/d \tag{7}$$

where: I_d is the current produced(μA), typically around 2μA,

F is Faraday's constant, which is 9.64×10^4 C/mol,

$P_M(T)$ is the permeability of the membrane, which is a function of temperature. Hence, temperature compensation is involved at this point (Barrer),

A is the surface area of cathode (m^2),

P_{o2} is the partial pressure of oxygen (mmHg),

d is the membrane thickness (m).

3.1.4 Oxidation-reduction potential measurement

Oxidation-reduction potential measures the electrical potential of a redox reaction. It also serves as a measure of how much oxidation or reduction takes place in the given conditions. A redox reaction refers to the exchange of electrons. Oxidation is the loss of electrons and therefore the solution is more positive. Reduction is the gaining of electrons which causes a negative charge. The two processes always occur together because the oxidizing agent gains electrons during reduction. The potential, measured by an ORP instrument is a measure of the power of a substance to gain electrons in solution. The stronger the reducing agent, the more likely it will readily lose electrons and give them to another substance. A strong reducing agent will have a high negative redox potential. A high positive redox potential is associated with a strong oxidizing agent. The ORP measurement is not a measurement of concentration; rather it measures the activity level, similar to a pH test. The basic thermodynamics for ORP reactions can be expressed using the Nernst equation:

$$E_{cel} = E_{Ox/\mathrm{Red}} + 2.303\frac{RT}{nF}[Ox/\mathrm{Re}\,d] \tag{8}$$

where: $E_{Ox/\mathrm{Red}}$ is the potential under standard conditions of unit activity referred to the standard hydrogen electrode,

R is the gas constant, 1.986 cal/(mol*K),

F is Faraday’s constant,

T is the kelvin temperature in K

n is the number of electrons exchanged in the reaction.

$E_{Ox/\mathrm{Re}d}$ is found in tables in handbooks and is the value relative to the standard hydrogen electrode.

3.1.5 Conductivity measurement

Electrical conductivity represents the ability of a solution to conduct electrical current. Measuring a conductivity is simple. When a probe is immersed in a sample solution, a potential is generated between two cells of the probe, and the sensor measures the current. Conductivity is calculated as a reciprocal of resistance:

$$G = 1/R = I/E \tag{9}$$

In the equation(9): G is conductivity ($\mu S/cm$),

R is resistance (Ω),

I is current (A),

E is voltage (V).

The potential between two cells is usually takes the form of a sine wave, and conductivity of the solution is proportional to the volume of ions concentrated in the solution.

3.2 Data Compensation

3.2.1 Temperature and pH value

Equation (6) gives a simple linear relation between pH value and electrode voltage at room temperature (25 °C). However in a practical dynamic measuring environment, a fixed temperature could cause a bit of reading error. There is no perfect sensor; temperature change will also change the electrode's sensitivity. Since the temperature of a water body is usually much lower than room temperature, a reading taken by a pH sensor without correction will be lower than its actual pH value leading to a wrong quality evaluation. This temperature error is very close to 0.003pH/ °C /pH unit away from pH 7. Depending on equation (5), an instant temperature reading from a thermometer is taken into account. Figure 3.2 illustrates this entire procedure.

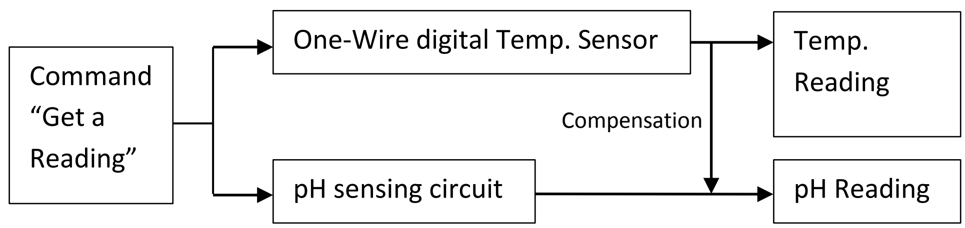

Figure 3.2 Data compensation procedure for pH sensor

When the command “Get a Reading” is sent out, the One-Wire digital temperature sensor takes a reading immediately and this reading passes down two pathways at the same time. One will be acquired by the microcontroller, while the other one will be merged into the pH sensing circuit for compensation. We deployed a test to compare reading results of the two pH values mentioned above. In the test, the pH value of a bottle of pH 4 standard solution is measured. Prior to the test, the standard solution was cooled down (11 ℃) in a refrigerator, as we need the solution temperature to differ from room temperature (25 ℃). The test result is shown in Figure 3.3, showing the error caused by temperature difference and the corrected data that is closer to the real value.

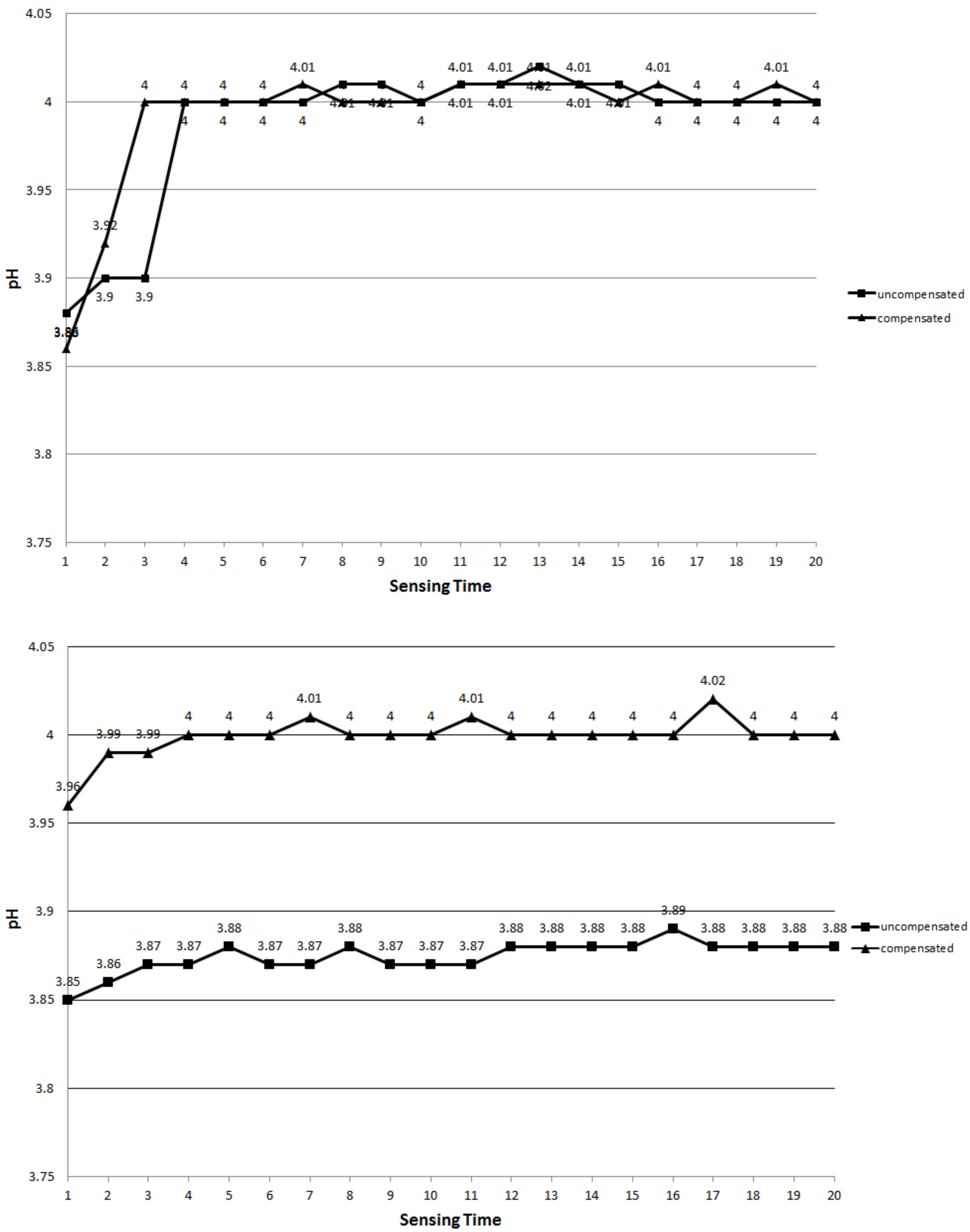

Figure 3.3 Comparison of compensated and uncompensated results

3.2.2 Temperature and dissolved oxygen sensing

The permeability of the membrane P_M can be calculated from equation (10):

$$P_M = SD \tag{10}$$

where S represents the sorption equilibrium parameter, which is the constant of proportionality between pressure (p) and concentration of the permeate (C) as the following equation:

$$S = C/p \tag{11}$$

and D is the diffusion coefficient (or mass diffusivity). The diffusion of solids at different temperatures is generally found to be well predicted by the Arrhenius equation:

$$D = D_o e^{-E_A/(kT)} \tag{12}$$

where: D is the diffusion coefficient (m^2/s),

D_o is maximum diffusion coefficient (at infinite temperature, m^2/s),

E_A is the activation energy for diffusion in dimensions of ($\mathrm{J/atom}$),

T is the absolute temperature (K),

k is the Boltzmann constant.

From equation (10), (11), and (12), the permeability of the membrane P_M can be easily derived as a function of temperature (T):

$$P_M = D_o e^{-E_A/(kT)} \cdot C/p \tag{13}$$

Then, substitute P_M into equation (7), we have:

$$I_d = \left[4 \cdot F \cdot D_o e^{-E_A/(kT)} \cdot C/p \cdot A \cdot P_{o2}\right]/d \tag{14}$$

Equation (14) expresses how the factor of temperature affects a galvanic D.O. sensor reading. For convenience, the default setting of temperature is usually fixed at 20 ℃ (293.15 K) for a rough measurement.

Like the temperature compensation procedure for the pH sensor, variable T is a changing quantity given by the One-Wire temperature sensor reading (Figure 3.4).

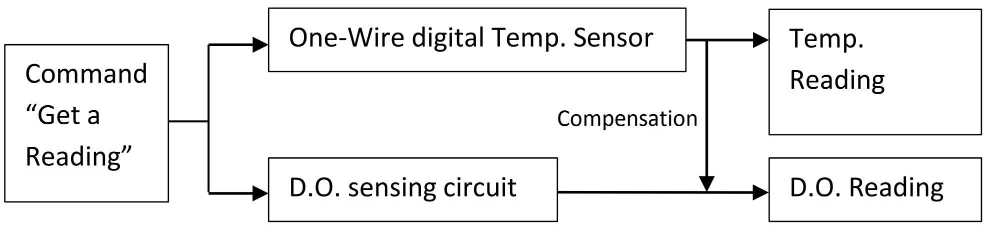

Figure 3.4 Data compensation procedure for D.O. sensor

The comparison of measuring 0 mg/L standard dissolved oxygen solution shows that the dynamic temperature compensated dissolved oxygen sensing provides a more accurate reading.

3.3 Summary

Principles on which water quality probe measurements were based are presented in this chapter while including mathematical representations. A thermometer is a simple sensor, while a conductivity probe relies on Ohm's law. Both pH and ORP probes utilize potential measurement, and the associated values are calculated by using the Nernst equation. The working principle of the D.O. probe is somewhat complex. It represents the ion permeability in a solution. Data compensations for pH value and D.O. probes were shown to be effective. by taking an instant temperature reading into account, the reading shift error generated by temperature is reduced significantly, which improves the accuracy and reliability of the monitoring platform. Data transmission will be discussed in the next chapter.

Chapter 4: Data Transmission

4.1 GSM Signal Coverage

GSM is a digital mobile communication system that is widely used around the world. GPSR is an extension to GSM in many wireless remote systems; GPRS is applied for data and command communication[32]. CDMA has a faster transmission speed, up to 150Kbps, while GPRS is limited to 80Kbps nominally; it is a particularly suitable for remote applications that require instant response[33]. However, some suburban places have a very weak GSM signal, the GPRS in such areas is not solid, and CDMA can be even worse, so sometimes they cannot be activated[34][35][36]. A wireless control based on GPRS cannot be realized in such a location. Figure 4.1 and 4.2 show the GSM signal coverage maps of Vancouver B.C. and Mumbai, India. It is seen that most urban places are fully covered. The red dots displayed along the banks of rivers and lakes indicate the weakness of GSM communication around that area.

Figure 4.1 GSM signal coverage in Vancouver B.C.[37]

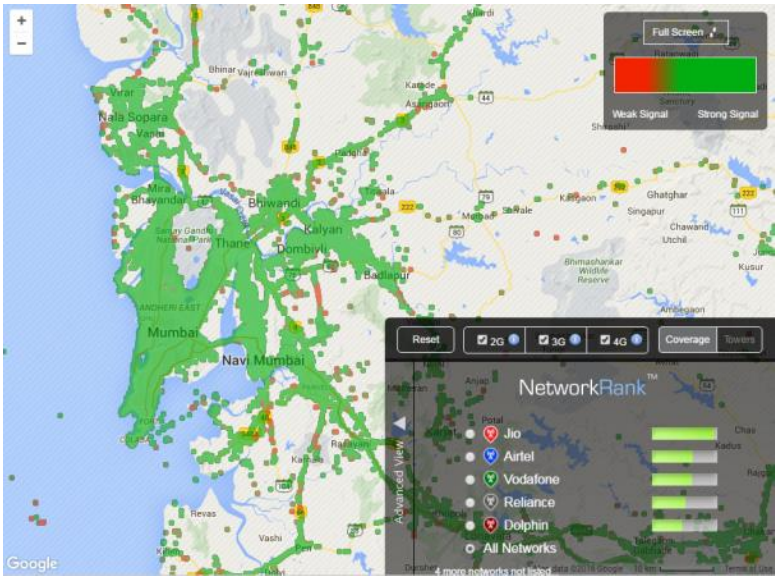

Figure 4.2 GSM signal coverage in India[37]

However, the Short Message Service, also known as SMS, functions well even with very poor signal strength; theoretically, it only requires a basic signal to work[38]. Because of this feature, it is implemented as a remote control facility in poor signal areas[39].

4.2 Transmission through SMS

When a group of readings is stored in the microcontroller memory, it is ready to be sent out in pure text format by the GSM module mounted on the Arduino Uno board (Figure 4.3).

Figure 4.3 GSM module

The following are the steps for activating GSM in Arduino:

(1) The GSM library should be input at the very beginning;

(2) Initialize the library instance of the classes that are about to be used,

```
{GSM gsmAccess
 GSM_SMS sms;}
```

(3) Open a serial connection to the computer with 9600 baud rate. Then send an initial message “Water Quality Monitor” as a system start mark,

```
{Serial. begin(9600);
 Serial. println(“ Water Quality Monitor”);}
```

(4) Create a local variable to track the connection to keep the sketch from starting until the SIM is connected to the network,

```
{boolean notConnected = true;}
```

(5) Connect to network by calling “gsmAccess.begin()”

```
{while (notConnected)
        {if(gsmAccess.begin()==GSM_READY)
```

```
            notConnected = false;
else {serial.println("not connected");
        delay (1000);}
}}
```

Now the GSM module is ready for data transmission. Then use the following code to send water parameter values denoted as variables “Temp, PH_data, DO_data, ORP_data, Cond_data” to the base station server in the text message form.

```
{ sms.beginSMS(remoteNumber);
  sms.print(Temp);
  sms.print(PH_data);
  sms.print(DO_data);
  sms.print(ORP_data);
  sms.print(Cont_data);
  sms.endSMS(); }
```

4.3 Summary

In this chapter, the wireless data transmission method applied in this field monitoring platform was presented. By comparing and analyzing different transmission modes, such as GPRS and CDMA, SMS was determined to be the most suitable means for water quality monitoring system deployed in a remote region. Also, it does not require an extremely instant response. In the second part of this chapter, software developed to activate Arduino GSM module was presented in steps.

Chapter 5: Data Fusion

5.1 Dempster-Shafer Theory

There are many methods for combination of data, knowledge and information. They include fuzzy rule-based inference[40], Bayesian Networks[41], Kalman filter theory[42], and Dempster-Shafer evidence method[43][44], which are the most common techniques.

Among these, Dempster-Shafer theory is a theory of evidence classically originated by Dempster (1968) and extended by Shafer (1976). It is also a generalization of the Bayesian theory of subjective probability complementing the missing of "No evidence" dealt with as equal non-informative priors in Bayesian theory.

5.2 Definitions

In the D-S evidence theory, the frame of discernment Θ is introduced, which is a finite set of mutually exclusive elements consisting of 2^{Θ} subsets. These elements can be any type of objects or hypothesis. For example, Θ = {A, B, C, D} has 16 (= 2^4) subsets. Those subsets are also called power sets, as defined. In the water quality problem, six quality levels $\{A, B, C, D, E, F\}$ are considered, and they are the elements of the frame of discernment. Another important concept is basic probability assignment (BPA), which is also called belief function *m*. Function *m* is defined as a mapping over the interval [0, 1] as follows:

$$m : 2^{\Theta} \longrightarrow [0,1]$$

$$m(\Phi) = 0$$

$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{15}$$

Here $m(\Phi)$ is a null set, the assignment $m(A)$ represents the probability (belief) of subset A, and any subset with a non-zero BPA ($m(A) \neq 0$) is called a focal element. The sum of all *m* is “1”. In the water quality problem, subset *A* expresses a certain level of water quality, and correspondingly $m(A)$ gives the probability of this level. When $m(A)$ is beyond a fixed value (threshold value), it is confident to say that the tested water is in level A.

5.3 Combination Rule

The Dempster-Shafer rule of combination presents an orthogonal combination of multiple belief functions from different sources (evidence) and produces one fused result taking every individual information source into account. Assuming that there are 2 sensors as data sources, then m_1 and m_2 are two belief functions of these two independent sensors. Specifically, m_1 indicates the probability of set A based on sensor 1, while m_2 is the belief function of sensor 2 to Level A. Then the D-S combination rule determines $m(A)$ which is a joint of m_1 and m_2 given by the following equations:

$$m(A) = \begin{cases} \dfrac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K} & A \neq \Phi \\ 0 & A = \Phi \end{cases}$$

$$\text{where } K = \sum_{B \cap C = \Phi} m_1(B) m_2(c) \tag{16}$$

Here *K* is the degree of conflict, measuring the conflict between evidence sources. (1-*K*) is a normalization factor, eliminating the interference from conflicting evidence.

In the water quality monitoring problem discussed here, 6 evaluation levels of water quality considered. They are denoted by A, B, C, D, E, F, F from high to low in sequence. Therefore, the frame of discernment $\Theta = \{A, B, C, D, E, F\}$. Four mutually independent sensors are applied and four different parameters are considered to determine the water quality. They are pH value, dissolved oxygen (D.O.) level, conductivity, and oxidation-reduction potential (ORP), denoted by s_1, s_2, s_3, s_4 respectively. Each sensor gives its own reading as evidence and then the belief function value towards each water quality level. All these belief functions are aggregated to a group of summarized values balancing judgement of every sensor according to the D-S rule. Figure 5.1 gives a simple representation of the fusion procedure.

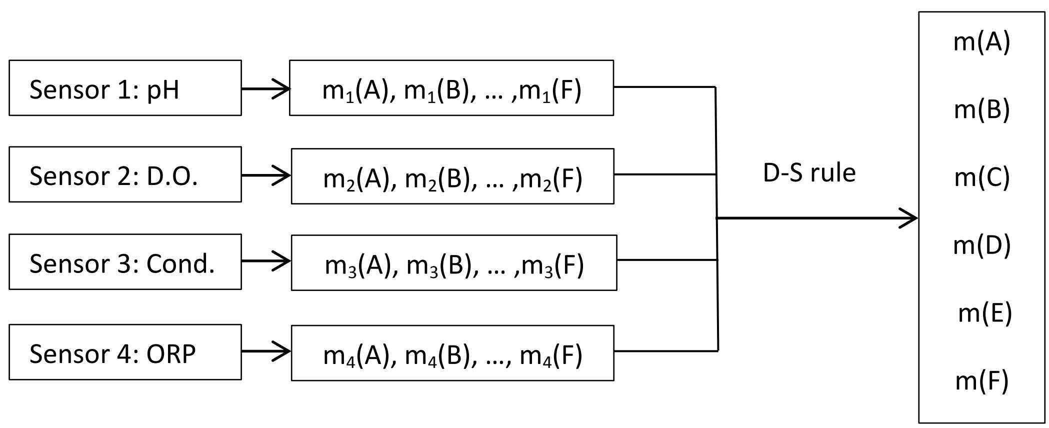

Figure 5.1 Fusion using D-S theory for water quality

Besides the belief value $m_i(A),\ m_i(B),\ m_i(C),\ m_i(D),\ m_i(E),\ m_i(F),$ of sensor i, the unidentified belief function, also called uncertain function of sensor i is $m_i(\theta)$:

$$m_i(\theta) = 1 - m_i(A) - m_i(B) - m_i(C) - m_i(D) - m_i(E) - m_i(F) \tag{17}$$

Clearly, a smaller $m_i(\theta)$ indicates a lower uncertainty, in other words, more confidence to judge the water quality level. On comparing the fusion belief function values, if the value of one specific belief function is significantly large than the rest (usually larger than the threshold value that is set), simultaneously the uncertain function value is very small, then we can confidently conclude the tested water quality is in the level with highest belief function value.

Each of the 4 parameters, pH, D.O., ORP, and conductivity has different scales and quality judgement standards. Simple aggregation of them is meaningless. According to environmental experts' experiences, water quality is classified into 6 degrees, denoted as letters A, B, C, D, E, F where A is the best. Features and parameter indicators for each quality class are given in Table 5.1.

Table 5.1: Parameter indicators of water quality classes

Para. / **Class**	**pH**	**D.O.**	**ORP**	**Conductivity**
A	8.5	10 mg/l	-100 mV	50 μS/cm
B	7	8 mg/l	100 mV	250 μS/cm
C	5.5	6 mg/l	300 mV	500 μS/cm
D	4	4.5 mg/l	500 mV	5,000 μS/cm
E	2.5	3 mg/l	700 mV	50,000 μS/cm
F	1	1 mg/l	900 mV	500,000 μS/cm

Mark R_x is the reading from sensor x, and I_y is the indicator of water quality level y.

Through calculating the distance between R_x and I_y, we can know how close the quality is to level y. Consequently, the probability of level y can be determined from the evidence, which is the sensor reading.

Usually similarity measurement are conducted by calculating the "Distance" among samples. A shorter distance indicates a higher similarity. Several mathematical methods can be used such as Euclidean Distance, Manhattan Distance, Chebyshev Distance and also Cosine similarity. Depending on the problem, application of these methods can be selected. In this thesis, data distributed in each dimension has different weights, so the Standardized Euclidean Distance measurement is applied. It is a modification of the original Euclidean Distance. If vector $\mathbf{p}(p_1, p_2, ...p_n)$ and $\mathbf{q}(q_1, q_2, ...q_n)$ are in a Euclidean n-space, the distance $d_{p,q}$ between them can be described as:

$$d_{p,q} = \sqrt{\sum_{k=1}^{n} \left(\frac{p_k - q_k}{s_k} \right)^2} \tag{18}$$

where, $d_{p,q}$ is the distance between p_k and q_k, s_k is the standard of p_k and q_k.

For the water quality problem in this thesis, readings from the x sensor taken every 75 minutes are gathered into one group and denoted as $\mathbf{R_{xk}}(R_{x1}, R_{x2}, ..., R_{xn})$, n = 15 since the sampling rate is 5 mins. Substituting p_k and q_k in equation (18) by $\mathbf{R_x}$ and I_y respectively, the distance d_{RxIy} between x sensor reading and quality y level indicator is calculated by

$$d_{R_X I_y} = \sqrt{\sum_{k=1}^{n} \left(\frac{R_{xk} - I_y}{s_k} \right)^2} \tag{19}$$

If $\mathbf{R_x}$ is closer to I_y, it shows a higher probability of quality level y. Define F_{xy} as a correlation function:

$$F_{xy} = \frac{1}{d_{RxIy}} \tag{20}$$

It represents the judgment of sensor x about the plausibility of water quality level y. Therefore, we can derive the quality belief function $m_x(y)$ as:

$$m_x(y) = \frac{F_{xy}}{F_{xA} + F_{xB} + F_{xC} + F_{xD} + F_{xE} + F_{xF}} \tag{21}$$

5.4 Verification

In the present experiment, we apply 5 sensors to monitor the water, but only 4 of them are counted for quality evaluation. Sensors take a reading every 5 minutes, so vector $\mathbf{R_{xk}}(R_{x1}, R_{x2}, ..., R_{xn})$ contains 15 continuous readings. For example, parameter readings measured by pH sensor (sensor 1) are denoted as $\mathbf{R}_1(R_1, R_{12}, ..., R_{15})$. The experimentally recorded parameter values are given in Table 5.2.

Table 5.2: Sensory measurements

Temperature(℃)	pH	D.O. (mg/L)	Conductivity (μS/cm)	ORP(mV)
8.62	7.56	15.23	94.31	256.40
8.62	7.55	15.23	94.31	257.00
8.62	7.58	15.25	95.00	257.51
8.65	7.58	15.26	94.50	257.33
8.66	7.58	15.26	94.30	258.69
8.68	7.59	15.25	94.35	258.61
8.68	7.59	15.27	94.48	258.63
8.67	7.58	15.28	97.21	258.60
8.68	7.60	15.28	95.56	258.70
8.68	7.59	15.28	96.12	258.71
8.68	7.58	15.28	96.60	258.70
8.67	7.58	15.29	96.52	258.34
8.66	7.56	15.28	95.44	258.50
8.68	7.57	15.29	95.60	258.63
8.68	7.58	15.30	94.20	258.99

In general, a water system is balanced no matter whether it is good or not, and its parameter values do not change too much unless there is external interference[45].

From the Euclidean distance equation (19), we can calculate how close each sensor reading is to the water quality level indicators. Then from equation (20) and

(21), the quality belief function $m_x(y)$ of each sensor can be achieved, with the result shown in the first 4 rows of Table 5.3.

To calculate the degree of conflict K value, first we need to find out all the empty set areas of sensor sources. In this case, there are 4 sensors and each sensor can give 6 kinds of quality evaluations. It follows that there are a total of 1296 ($=6^4$) combinations of belief functions, 6 of which are judgements for quality (common area), and the other 1290 are null sets. Then through the D-S rule of combination, equation (15), the fusion belief function value for every quality class is derived as given in the last row in Table 5.3. The last column of Table 5.3 gives $m\ (\theta)$, the uncertainty function supplementing the evaluation universe.

Table 5.3: Quality belief function value

BFV / Sensor	A	B	C	D	E	F	$m\ (\theta)$
pH	0.2764	0.4408	0.1226	0.0712	0.0502	0.0387	$1.11\text{x}10^{-16}$
D.O.	0.2792	0.2024	0.1587	0.1366	0.1199	0.1031	$1.25\text{X}10^{-16}$
Conductivity	0.7070	0.2067	0.0790	0.0065	0.0006	$6\text{x}10^{-5}$	$1.29\text{x}10^{-9}$
ORP	0.0681	0.1541	0.5837	0.1009	0.0052	0.0380	$1.94\text{x}10^{-16}$
Fusion	0.4979	0.3809	0.1203	0.0009	$2.85\text{x}10^{-5}$	$1.31\text{x}10^{-6}$	$1.33\text{x}10^{-16}$

It is easy to find in Table 5.3, according to the parameters measured as evidences, that the sensors have different judgments on the quality level. By the pH sensor, the belief function value of level B is obviously larger than the other levels. The D.O. sensor does not present a clear determination between level A and B though the belief function value of A is somewhat larger. Three sensors are in conflict, the pH sensor indicates a level B quality as just mentioned, the conductivity sensor strongly shows a belief in level A, while the ORP is more confident with level C. It is hard to see the actual water quality level by relying on an individual sensor only, but sensor fusion makes it more reliable and definitive.

According to the D-S theory, water quality evaluation monitoring principles follow these criteria:

(1) The level with the maximal belief function value may be considered as the water quality status. This value also should be larger than a certain threshold number. Usually, this threshold value is at least larger than 1/N (N is the number of levels we assumed), and larger the threshold value is, the more accurate this fusion evaluation will be. However, the fusion value will not meet a requirement of a threshold with an overly large value. In this thesis, this value is set as 0.33(2/N).
(2) The difference between the largest value and others should be larger than a certain value for a reliable decision. Here it is 0.1.
(3) The uncertain function value also reflects the reliability. A smaller uncertainty should be required to secure a stronger certain judgement.

In the last row of Table 5.3, the fusion belief function value of level A is 0.4979, which is the biggest number, and the uncertain function value is 1.33×10^{-16} which is a very small number. Conclusively, the water quality can be determined confidently at level A based on the parameters measured and fused.

Chapter 6: Conclusion

6.1 Main Contributions

In this thesis, a GSM wireless monitoring system based on the Dempster-Shafer evidence sensor fusion approach was developed for water quality monitoring. In the field implementation, five basic water quality related parameters: temperature, pH value, dissolved oxygen level, oxidation-reduction potential, and electrical conductivity value are measured by five types of sensors. To remove side effects of temperature on the pH and D.O. sensors, real-time temperature compensation is employed. Effects of this compensation are very significant and as a result, higher sensing accuracy obtained with increased reliability. An SMS data transmission protocol is applied in the developed system, which is able to work even in a poor communication environment, and helps to achieve data completeness.

Moreover, a water quality evaluation algorithm was proposed in the thesis. Each sensor acts as one individual data source, giving a single judgement for the quality level, and the quality judgement is expressed in probabilities. Like a voting mechanism, sensor judgements to various quality classes are combined (fused). As a result of this combination, consistency of the judgements is enhanced with a high belief function value, while inconsistency is reduced, with a low plausibility. An experimental result was presented, which verified the reliability of this fusion method.

6.2 Future Work

Although the water quality monitoring system developed and the sensor fusion method proposed in the present thesis have demonstrated good performance, there still exist several issues and aspects that need to be further investigated and explored. A water system is a very complex unit, and it contains a wide range of compositions, and factors that cause the quality change also vary. In this thesis, only 5 elements have been considered, and this may not be adequate to thoroughly assess water quality. Thus, it is expected that more parameters will be involved and more sensors will be incorporated into the monitoring system in due course.

This thesis has not considered any extreme situations, such as when one or two parameters change significantly and are out of the common range, such as in extremely acidic or basic water, while most of the rest of the readings do not change much, and their belief function values indicate a common level of quality. The fusion result does not correct for this in the present implementation. The algorithm for sensor fusion will be improved to solve such problems so that the performance of the whole monitoring system will become better.

Bibliography

[1] Verma, S., & P. (2012). Wireless Sensor Network application for water quality monitoring in India. *2012 National Conference On Computing And Communication Systems*. doi:10.1109/ncccs.2012.6412990

[2] Gimenez-Gomez, P., Escude-Pujol, R., Jimenez-Jorquera, C., & Gutierrez-Capitan, M. (2015). Multisensor Portable Meter for Environmental Applications. *IEEE Sensors J. IEEE Sensors Journal, 15*(11), 6517-6523. doi:10.1109/jsen.2015.2460011

[3] https://www.usgs.gov/, 2014/05/23

[4] http://hiyoshi-india.com/, 2014/05/23

[5] Lambrou, T. P., Anastasiou, C. C., Panayiotou, C. G., & Polycarpou, M. M. (2014). A Low-Cost Sensor Network for Real-Time Monitoring and Contamination Detection in Drinking Water Distribution Systems. *IEEE Sensors J. IEEE Sensors Journal, 14*(8), 2765-2772. doi:10.1109/jsen.2014.2316414

[6] Gimenez-Gomez, P., Escude-Pujol, R., Jimenez-Jorquera, C., & Gutierrez-Capitan, M. (2015). Multisensor Portable Meter for Environmental Applications. *IEEE Sensors J. IEEE Sensors Journal, 15*(11), 6517-6523. doi:10.1109/jsen.2015.2460011

[7] Lambrou, T. P., Panayiotou, C. G., & Polycarpou, M. M. (2015). Contamination detection in drinking water distribution systems using sensor networks. *2015 European Control Conference (ECC)*. doi:10.1109/ecc.2015.7331043

[8] Bennis, I., Fouchal, H., Zytoune, O., & Aboutajdine, D. (2015). Drip Irrigation System using Wireless Sensor Networks. *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*. doi:10.15439/2015f299

[9] Chaamwe, N. (2010). Wireless Sensor Networks for Water Quality Monitoring: A Case of Zambia. *2010 4th International Conference on Bioinformatics and Biomedical Engineering*. doi:10.1109/icbbe.2010.5515792

[10] Wang, J., Ren, X., Shen, Y., & Liu, S. (2010). A Remote Wireless Sensor Networks for Water Quality Monitoring. *2010 International Conference on Innovative Computing and Communication and 2010 Asia-Pacific Conference on Information Technology and Ocean Engineering*. doi:10.1109/cicc-itoe.2010.9

[11] Chang, N., Vannah, B. W., Yang, Y. J., & Elovitz, M. (2014). Integrated data fusion and mining techniques for monitoring total organic carbon concentrations in a lake.

International Journal of Remote Sensing, 35(3), 1064-1093. doi:10.1080/01431161.2013.875632

[12] Quan, H., Li, J., & Peng, D. (2014). Multisensor fault diagnosis based on data fusion using D-S theory. *Proceedings of the 33rd Chinese Control Conference*. doi:10.1109/chicc.2014.6896234

[13] Chou, J., Lin, C., Liao, Y., Chen, J., Tsai, Y., Chen, J., & Chou, H. (2014). Data Fusion and Fault Diagnosis for Flexible Arrayed pH Sensor Measurement System Based on LabVIEW. *IEEE Sensors J. IEEE Sensors Journal, 14*(5), 1405-1411. doi:10.1109/jsen.2013.2296148

[14] Liao, Y., & Chou, J. (2012). Comparison of pH Data Measured with a pH Sensor Array Using Different Data Fusion Methods. *Sensors, 12*(12), 12098-12109. doi:10.3390/s120912098

[15] Zhu, D., & Gu, W. (2008). Sensor Fusion in Integrated Circuit Fault Diagnosis Using a Belief Function Model. *International Journal of Distributed Sensor Networks, 4*(3), 247-261. doi:10.1080/15501320701260626

[16] Ding, Q., Peng, Z., Liu, T., & Tong, Q. (2014). Building Fire Alarm System with Multi-sensor and Information Fusion Technology Based on D-S Evidence Theory. *2014 International Symposium on Computer, Consumer and Control*. doi:10.1109/is3c.2014.238

[17] Yin, S., Huo, K., & Liu, Y. (2014). Multi-sensor fusion recognition method based on improved D-S evidence theory. *2014 International Conference on Information and Communications Technologies (ICT 2014)*. doi:10.1049/cp.2014.0595

[18] Sadiq, R., & Rodriguez, M. J. (2005). Interpreting drinking water quality in the distribution system using Dempster–Shafer theory of evidence. *Chemosphere, 59*(2), 177-188. doi:10.1016/j.chemosphere.2004.11.087

[19] Papa, B. (2004). *The source water protection primer*. Toronto: Pollution Probe

[20] Behar, S. (1996). *Testing the waters: Chemical and physical vital signs of a river*. Montpelier, VT: River Watch Network

[21] Langland, M., & Cronin, T.(2003), A Summary Report of Sediment Processes in Chesapeake Bay and Watershed. *In Water-Resources Investigations Report.*2003 03-4123

[22] Gray, J. R., Gylsson, G. D., Turcios, L. M., and Schwarz, G. E. Comparability of Suspended-Sediment Concentration and Total Suspended Solids Data, *USGS Water-Resources Investigations Report* 00-4191, 2000

[23] http://www.ozoneapplications.com/info/orp.htm, 2014/08/02

[24] Hu, C., Tian, D., & Yan, X. (2014). Research on placement of water quality sensor in water distribution systems. *Proceeding of the 11th World Congress on Intelligent Control and Automation*. doi:10.1109/wcica.2014.7053312

[25] Canadian Council of Ministers of the Environment, "CCME WATER QUALITY INDEX 1.0 Technical Report", Canadian Water Quality Guidelines for the Protection of Aquatic Life, 2001

[26] Chang, N., & Imen, S. (2015). Multi-sensor Acquisition, Data Fusion, Criteria Mining and Alarm Triggering for Decision Support in Urban Water Infrastructure Systems. *2015 IEEE International Conference on Systems, Man, and Cybernetics*. doi:10.1109/smc.2015.105

[27] Rivera-Mejia, J., Villafuerte-Arroyo, J. E., Vega-Pineda, J., & Sandoval-Rodriguez, R. (2015). Comparison of Compensation Algorithms for Smart Sensors With Approach to Real-Time or Dynamic Applications. *IEEE Sensors J. IEEE Sensors Journal, 15*(12), 7071-7080. doi:10.1109/jsen.2015.2469279

[28] Ueno, Y., & Shimizu, H. (n.d.). Voltage controlled temperature compensated crystal oscillator using 2-port crystal resonator. *Proceedings of the 45th Annual Symposium on Frequency Control 1991*. doi:10.1109/freq.1991.145930

[29] Ali, M. (2016). Compensation of temperature and acceleration effects on MEMS gyroscope. *2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*. doi:10.1109/ibcast.2016.7429889

[30] Rana, R., Hu, W., & Chou, C. T. (2015). Optimal Sampling Strategy Enabling Energy-Neutral Operations at Rechargeable Wireless Sensor Networks. *IEEE Sensors J. IEEE Sensors Journal, 15*(1), 201-208. doi:10.1109/jsen.2014.2337334

[31] de Silva, C.W., Sensors and Actuators—Engineering System Instrumentation, 2nd edition, CRC Press/Taylor & Francis, Boca Raton, FL, 2016

[32] Gutierrez, J., Villa-Medina, J. F., Nieto-Garibay, A., & Porta-Gandara, M. A. (2014). Automated Irrigation System Using a Wireless Sensor Network and GPRS Module. *IEEE Trans. Instrum. Meas. IEEE Transactions on Instrumentation and Measurement*, *63*(1), 166-176. doi:10.1109/tim.2013.2276487

[33] Li, D., & Zhu, X. (2009). CDMA-Based Remote Wireless Water Quality Monitoring System for Intensive Fish Culture. *2009 WRI International Conference on Communications and Mobile Computing*. doi:10.1109/cmc.2009.241

[34] Ari, A., Anil, B., Syed, F. Y., Jarno, N., Mikko, V. (2015). Applicability of Frequency Selective Surfaces to Enhance Mobile Network Coverage in Future Energy-Efficient Built Environments. *21th European Wireless Conference*.2015. pp. 1-8

[35] Tranca, D. C., & Markovic, V. (2015). Energy consumption in periodical GSM/GPRS transmissions of small data chunks: An experimental study. *2015 12th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS)*. doi:10.1109/telsks.2015.7357777

[36] Purnima, P. S. (2014). Zigbee and GSM based patient health monitoring system. *2014 International Conference on Electronics and Communication Systems (ICECS)*. doi:10.1109/ecs.2014.6892762

[37] http://opensignal.com/, 2015/09/20

[38] Peersman, C., Cvetkovic, S., Griffiths, P., & Spear, H. (2000). The Global System for Mobile Communications Short Message Service. *IEEE Pers. Commun. IEEE Personal Communications, 7*(3), 15-23. doi:10.1109/98.847919

[39] Li, X., Yuan, Q., Wu, W., Peng, X., & Hou, L. (2010). Implementation of GSM SMS remote control system based on FPGA. *The 2nd International Conference on Information Science and Engineering*. doi:10.1109/icise.2010.5691440

[40] Wang, P., Wang, S., & Li, D. (2014). Research on the fuzzy set theory of evidence fusion algorithm with time-varying in multi-sensor detection network. *2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. doi:10.1109/icspcc.2014.6986286

[41] Wartzek, T., Bruser, C., Walter, M., & Leonhardt, S. (2014). Robust Sensor Fusion of Unobtrusively Measured Heart Rate. *IEEE Journal of Biomedical and Health Informatics IEEE J. Biomed. Health Inform., 18*(2), 654-660. doi:10.1109/jbhi.2013.2274211

[42] Huang, R., Tan, P., & Chen, B. M. (2015). Monocular vision-based autonomous navigation system on a toy quadcopter in unknown environments. *2015 International Conference on Unmanned Aircraft Systems (ICUAS)*. doi:10.1109/icuas.2015.7152419

[43] Yan, R., Li, G., & Liu, B. (2015). Knowledge fusion based on D-S theory and its application on Expert System for software fault diagnosis. *2015 Prognostics and System Health Management Conference (PHM)*. doi:10.1109/phm.2015.7380049

[44] Yang, K., Liu, S., & Li, X. (2015). A Novel Detection Scheme Based on D-S Evidence Theory in Wireless Sensor Networks. *2015 International Conference on Intelligent Networking and Collaborative Systems*. doi:10.1109/incos.2015.64

[45] Lemos, D., Dias, A. C., Gabarrell, X., & Arroja, L. (2013). Environmental assessment of an urban water system. *Journal of Cleaner Production*, *54*, 157-165. doi:10.1016/j.jclepro.2013.04.029

Appendix A

This Appendix gives the Arduino UNO code that was used for the present application.

```
//Libraries
#include "Arduino.h"
#include "GSM.h"
#include "OneWire.h"
#include "SoftwareSerial.h"

//PIN Number
#define PINNUMBER ""

// initialize the library instance
GSM gsmAccess; // include a 'true' parameter for debug enabled
GSM_SMS sms;

//Temperature sensor pin on digital 8
int DS18S20_Pin = 8;
//Temperature chip i/o
OneWire ds(DS18S20_Pin);
//PH Serial ports
#define PHtxpin 14
#define PHrxpin 15
SoftwareSerial PHserial(PHtxpin, PHrxpin);
//DO Serial ports
#define DOtxpin 16
#define DOrxpin 17
SoftwareSerial DOserial(DOtxpin, DOrxpin);
//ORP Serial ports
#define ORPtxpin 9
#define ORPrxpin 10
SoftwareSerial ORPserial(ORPtxpin, ORPrxpin);
```

```
//COND Serial ports
#define CDtxpin 18
#define CDrxpin 19
SoftwareSerial CDserial(CDtxpin, CDrxpin);

//Define PH variables
char PH_data[20];
byte PH_received_from_PC=0;
byte PH_received_from_sensor=0;
byte PH_startup=0;
float PH=0;
byte PH_string_received=0;
//Define DO variables
char DO_data[20];
byte DO_received_from_PC=0;
byte DO_received_from_sensor=0;
byte DO_startup=0;
float DO=0;
byte DO_string_received=0;
//Define ORP variables
char ORP_data[20];
byte ORP_received_from_PC=0;
byte ORP_received_from_sensor=0;
byte ORP_startup=0;
float ORP=0;
byte ORP_string_received=0;
//Define COND variables
char CD_data[20];
byte CD_received_from_PC=0;
byte CD_received_from_sensor=0;
byte CD_startup=0;
float CD=0;
byte CD_string_received=0;
```

```
//sensor set up
void setup(){
  // initialize serial communications
  Serial.begin(9600);
  PHserial.begin(9600);
  DOserial.begin(9600);
  ORPserial.begin(9600);
  CDserial.begin(9600);
  Serial.println("IAL WQ Remoter");

  // connection state
  boolean notConnected = true;

  // Start GSM shield
  // If your SIM has PIN, pass it as a parameter of begin() in quotes
  while(notConnected)
  {
    if(gsmAccess.begin(PINNUMBER)==GSM_READY)
      notConnected = false;
    else
    {
      Serial.println("Not connected");
      delay(1000);
    }
  }

  Serial.println("READY!");
}

void loop() {
//Read temperature

    float temperature = getTemp();
    delay(2000);
```

```
    char Temp[5]; // char Temperature reading
    dtostrf(temperature, 4, 2, Temp); //* float to char *

//Read pH
  PHserial.listen();
  delay(2000);
   if(PHserial.available() > 0) {
    PH_received_from_sensor=PHserial.readBytesUntil(13,PH_data,20);
    PH_data[PH_received_from_sensor]=0;
    PH_string_received=1; //a flag
  }

  PHserial.print("R\r");
   if(PH_string_received==1) {
    PH=atof(PH_data);
    PH_string_received=0;} //flag
//Read DO
  DOserial.listen();
  delay(2000);
   if(DOserial.available() > 0) {
    DO_received_from_sensor=DOserial.readBytesUntil(13,DO_data,20);
    DO_data[DO_received_from_sensor]=0;
    DO_string_received=1; //a flag
  }

  DOserial.print("R\r");
   if(DO_string_received==1) {
    DO=atof(DO_data);
    DO_string_received=0;} //flag
//Read ORP
  ORPserial.listen();
  delay(2000);
   if(ORPserial.available() > 0) {
```

```
    ORP_received_from_sensor=ORPserial.readBytesUntil(13,ORP_data,20);
    ORP_data[ORP_received_from_sensor]=0;
    ORP_string_received=1; //a flag
  }

  ORPserial.print("R\r");
   if(ORP_string_received==1) {
    ORP=atof(ORP_data);
    ORP_string_received=0;} //flag
//Read COND
  CDserial.listen();
  delay(2000);
   if(CDserial.available() > 0) {
    CD_received_from_sensor=CDserial.readBytesUntil(13,CD_data,20);
    CD_data[CD_received_from_sensor]=0;
    CD_string_received=1; //a flag
  }

  CDserial.print("R\r");
   if(CD_string_received==1) {
    CD=atof(CD_data);
    CD_string_received=0;} //flag

//Prepare to send data
  char remoteNumber[20]="+17788891234";  // telephone number to send sms
  Serial.println(remoteNumber);
  Serial.println("SENDING");
  Serial.print("Message:Temp=");

  Serial.println(Temp);
  Serial.println(PH_data);
  Serial.println(DO_data);
  Serial.println(ORP_data);
  Serial.println(CD_data);
```

```
// send the message
  sms.beginSMS(remoteNumber);
  sms.print(Temp);
  sms.print(PH);
  sms.print(DO);
  sms.print(ORP);
  sms.print(CD);
  sms.endSMS();
  Serial.println("\nCOMPLETE!\n");
  delay (6000);
}

//Returns the temperature from one DS18S20 in DEG Celsius
float getTemp(){

  byte data[12];
  byte addr[8];

  if ( !ds.search(addr)) {
      //no more sensors on chain, reset search
      ds.reset_search();
      return -1000;
  }

  if ( OneWire::crc8( addr, 7) != addr[7]) {
      Serial.println("CRC is not valid!");
      return -1000;
  }

  if ( addr[0] != 0x10 && addr[0] != 0x28) {
      Serial.print("Device is not recognized");
      return -1000;
  }
```

```
  ds.reset();
  ds.select(addr);
  ds.write(0x44,1); // start conversion, with parasite power on at the end

  byte present = ds.reset();
  ds.select(addr);
  ds.write(0xBE); // Read Scratchpad

  for (int i = 0; i < 9; i++) { // we need 9 bytes
    data[i] = ds.read();
  }

  ds.reset_search();
  byte MSB = data[1];
  byte LSB = data[0];
  float tempRead = ((MSB << 8) | LSB); //using two's compliment
  float TemperatureSum = tempRead / 16;

  return TemperatureSum;
}
```

Appendix B

This appendix presents the code for the Matlab Calculation for fusion.

```
//A: load matrix to be processed
Col=6;
Row=4;
Matrix_processed=[0.276386587 0.440814518 0.122641007 0.071227276 0.050187477
0.038743134;0.279232463   0.202401373 0.158727295 0.13661778  0.119914575
0.103106515;0.707049695 0.206691311 0.079031855 0.006522135 0.000641012 0.00006399;
0.068083232 0.154141878 0.583692768 0.100873368 0.055206488 0.038002267;];
//A: End
final_result=zeros(1,5);
result_flag=0;
for row_1=1:6
    for row_2=1:6
        for row_3=1:6
            for row_4=1:6
               if(not((row_1==row_2) && (row_2==row_3) && (row_3==row_4)))
                   result_flag=result_flag+1;
                   final_result(result_flag,1)=...
                       Matrix_processed(1,row_1)*...
                       Matrix_processed(2,row_2)*...
                       Matrix_processed(3,row_3)*...
                       Matrix_processed(4,row_4);
//                 pointer of each row is right the column number of each
//                 element that creates the final result
                   final_result(result_flag,2)=row_1;
                   final_result(result_flag,3)=row_2;
                   final_result(result_flag,4)=row_3;
                   final_result(result_flag,5)=row_4;
               end
            end
```

```
        end
    end
end
```