ISTANBUL MEDENIYET UNIVERSITY
INSTITUE OF GRADUATE STUDIES
DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

Artificial Potential Function-based Water Quality Monitoring Using Unmanned Surface Vehicle

Master Thesis

Emre Fikri BALTACI

October 2023

ISTANBUL MEDENIYET UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

Artificial Potential Function-based Water Quality Monitoring Using Unmanned Surface Vehicle

Master Thesis

Emre Fikri Baltacı

Supervisor

Asst. Prof. Haluk Bayram

October 2023

THESIS JURY APPROVAL

This Master thesis titled "Suüstü Robotu Kullanarak Yapay Potansiyel Fonksiyon Tabanlı Su Kalitesi İzleme (Artificial Potential Function-based Water Quality Monitoring Using Unmanned Surface Vehicle)" written by Emre Fikri Baltacı at the Department of Electrical and Electronics Engineering was accepted by our jury.

JURY MEMBERS	**SIGNATURE**
Supervisor:	
Asst. Prof. Haluk Bayram	....
Institution: İstanbul Medeniyet University	
Other Jury Members:	
Prof. Zehra Saraç	....
Institution: İstanbul Medeniyet University	
Asst. Prof. Hüseyin Şerif Savcı	....
Institution: İstanbul Medipol University	

Date of Deffence: 03.10.2023

STATEMENTS

Style and Reference Manual Statement

Having reviewed this thesis written under my supervision, I confirm that it has been written in accordance with APA Manual of Style and used its [footnote/in text] reference format consistently throughout the entire text.

Asst. Prof. Haluk Bayram

Declaration of Originality

I hereby declare that all information in this dissertation has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conducts, I have fully cited and referenced all material and results that are not original to this work.

Emre Fikri Baltacı

ACKNOWLEDGEMENTS

I would like to express my greatest gratitude to my advisor, Asst. Prof. Haluk Bayram, for his motivation, guidance in research, and contributions to the thesis.

I also extend my sincere appreciation to Prof. Zehra Saraç and Asst. Prof. Hüseyin Şerif Savcı for serving as my thesis committee members.

Furthermore, I would like to thank Assoc. Prof. Elif Soyer for her valuable advice on water quality monitoring.

The collaborative spirit and shared passion of the members of the Field Robotics Laboratory at Istanbul Medeniyet University have provided an inspiring and nurturing environment.

I would like to thank my family for their emotional encouragement and support.

Finally, I would like to thank my teammates for taking over my workload when I took leave to work on my thesis.

GENİŞ ÖZET

Suüstü Robotu Kullanarak Yapay Potansiyel Fonksiyon Tabanlı Su Kalitesi İzleme

Emre Fikri Baltacı

Yüksek Lisans Tezi, Elektrik-Elektronik Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Haluk BAYRAM

Ekim, 2023. 61 sayfa.

Bu tez çalışmasında insansız suüstü robotu kullanarak yapay potansiyel fonksiyonu tabanlı su kalitesi izleme işlemi gerçekleştirilmiştir. Bu kapsamda yapay potansiyel fonksiyonu hareket için kullanılırken, Gauss süreci regresyon modeli su kalitesi haritalama için kullanılmaktadır. Ayrıca bu araç enerji farkındalığına sahip olacak, ihtiyaca göre var olan gezinti işlemini durdurup en yakın kıyı noktasına gidip şarj edilmeyi bekler.

Temizlik, gıda üretimi, içme gibi alanlarda kullanılan, insan hayatında büyük bir öneme sahip olan su, küresel ısınma, kirlilik, müsilaj gibi zarar veren etkenler sebebiyle değeri giderek artmaktadır. Göl, deniz, nehir, baraj gibi su kaynaklarını izlemek için sabit izleme istasyonlarından ölçümler yapılarak su kalitesi haritalandırılması yapılabildiği gibi, su kalitesi hareket kabiliyetine sahip su kalitesi izleme araçları ile de yapılabilmektedir. Bu çalışmada hareket kabiliyetine sahip bir araç için su kalitesi izleme operasyonu kendi güzergahını belirleyebilecek şekilde otonom olarak gerçekleştirilmesi hedeflenmektedir. Bu operasyondaki problem su kalitesinin nasıl haritalandırılacağı ve otonom güzergah oluşturma işleminin nasıl yapılacağıdır.

Geçmişte yapılan çalışmalara göre makine öğrenmesi yöntemlerinden biri olan regresyon alanında, su kalitesi izleme işlemi için Gauss süreci regresyon modeli kullanılmaktadır. Gauss süreci regresyon modelinde belirli miktardaki eğitim verisi girdisi ile model eğitildiğinde test data girdisi için sadece değer sonucu vermemekle birlikte ilgili test datası için hata oranını da alınabilmektedir. Su kalitesi uygulaması için Gauss süreci regresyon modelinde eğitim verisi ölçüm lokasyonu ve ilgili lokasyondaki su kalitesi değeri olup belirli bir arazideki tüm lokasyon bilgileri test verisi olarak kullanıldığında tüm lokasyon için tahmin edilen su kalitesi değerleri ve bunların standart sapma/varyans bilgisini de vermektedir. Regressörün bu iş için bir kovaryans (Kernel) fonksiyonuna ihtiyacı vardır. Kernel fonksiyonları üzerinde toplama, çıkartma, ve çarpma işlemi yapılabilme-

siyle farklı varyasyonlar türetilebilmektedir. Mekansal bazlı uygulamalarla sıklıkla kullanılan radyal bazlı fonksiyon kerneli, sabit kernel, rasyonel ikinci dereceden kernel tipleri ve ayrıca beyaz gürültü tanımı için beyaz gürültü kerneli bu tez çalışmasında simülasyon için farklı kombinasyonlar denenip karşılaştırma yapılacaktır. Regresyon işleminin yapılmasında sonra ortaya çıkan varyans bilgileri suüstü robotun bir sonraki en optimal ölçüm işleminin belirlenmesi probleminde önemli bir rol oynamaktadır.

Sistem bir yandan su kalitesi haritalandırması yapıp bir sonraki ölçüm yeri için hareket ederek online eğitimle işlemini gerçekleştirecektir. Buradaki ölçüm yeri hesaplama işi bilgilendirici rota planlama problemini işaret eder. Arazilerdeki kıyı bölgelerden, engellerden kaçarken öte yandan maksimum belirsizliğe gitme işlemi için yol planlama yöntemlerinden potansiyel fonksiyonu yöntemi bu tez çalışmasında kullanılacaktır.

Uygulama yapılacak arazinin büyük olma durumu, sürekli uzak noktalara yapılacak zig-zag işlemi gibi verimsiz işlemler için bu tez çalışmasında yeni bir metot önerilmektedir. Bütün arazi, belirli sayıdaki alt bloklara ayrıştırılıp, bu bloklar da kendi içerisinde ayrıklaştırma ile her biri bir tane ölçüm noktası içeren belirli sayıdaki hücrelere yatay ve düşey olarak bölünmüştür. Bir blok içerisinde hücreler arası hareket ve su kalitesi ölçüm işlemi lokal yaklaşım, bloklar arası gezme işlemi global yaklaşım olarak sınıflandırılmaktadır. Bloklar arası hareketi içeren global yaklaşımda robot yapay potansiyel fonksiyonundan türetilmiş yeni yaklaşımla hareket etmektedir. Bu yaklaşım için bloklar ölçüm alınmış ve ölçüm alınmamış olarak gruplandırılmaktadır. Ölçüm alınmamış bloklar çekici potansiyel, ölçüm alınmış bloklar itici potansiyele sahip olup ayrıca sahil kıyı gibi engeller de itici potansiyele sahiptir. Ölçüm alınmamış blokun oluşturduğu potansiyelde ilgili bloktaki su kalitesi ortalama standart sapma değerleriyle doğru, blok ile robot arasındaki uzaklığın karesi ile ters orantılıdır. Ölçüm alınmış blokun oluşturduğu itici potansiyel, blok ile robot arasındaki mesafenin karesi ile ters orantılıdır. Belirli bir mesafedeki kıyı bölgesi noktalarının yaptığı itici potansiyel, robot ile kıyı bölgesi noktası arası mesafe ile ters orantılıdır İlgili potansiyellerin degradeleri x ve y yönünde toplanarak bu toplamın ark-tanjantı robota açı girdisini sağlar. Global harekette robot maksimum hızına gittiği düşünülür. Global yaklaşımda robot içinde ölçüm aldığı bloktan itici potansiyel ile itilerek, degrade ile edilmiş yönde ölçüm alınmamış bloklara doğru hareket eder. Giderek artan çekim etkisiyle yakınlarından biri olan ölçüm alınmamış bloğun ağırlık merkezine robot vardığında, çekici potansiyellerin maksimum olanındaki değer sıkışması, tekrar durumunun oluşması global yaklaşımı durdurur ve robot lokal yaklaşıma geçer.

Lokal yaklaşımda robot içerisinde bulunduğu bloktaki bütün hücreler içerisinde maksimum varyansın olduğu hücreye doğru hareketini potansiyel fonksiyonu ile sağlar. Robot blok içerisinde ölçüm alıp geçmişteki tüm ölçüm lokasyon ve su kalitesi değerleri ile regressörü eğitir. Her bir su kalitesi için bu regressör ayrı ayrı çalıştırılır. Birbirleri ile direkt olarak korelasyonu olmayan su kaliteleri bağımsız olay olarak düşünülüp varyansları toplanabilir. Buna göre ağırlık katsayıları ile çarpılan su kalitesi varyansları toplamı eş-değer varyansı ifade eder. Robot bulunduğu blok içerisinde lokal yaklaşımda eş-değer varyansın maksimum olduğu lokasyonu kendisine bir sonraki ölçüm noktası olarak hedefleyecektir. Bu hedef noktası robot için çekici potansiyel, arazi etrafındaki kıyı noktaları itici potansiyel oluşturacaktır. Hedef nokta ile robotun bulunduğu nokta arasındaki uzaklığın karesi ile çekici potansiyel doğru, robotun lokasyonu ile kıyı noktalarının arasındaki uzaklık ile itici potansiyel ters orantılıdır. Bu potansiyellerin degrade'i alınarak x ve y yönünde elde edilen degradeler toplamı ark-tanjant ile hedef yönün açı bilgisini sağlayacaktır. Lokal blok içerisinde hareket kabiliyeti problemi olabileceğinden robot maksimum hızda gitmeyecektir. Robotun hızı lokal yaklaşım başlangıcındaki çekici potansiyel ile ilgili andaki çekici potansiyel arasındaki orandan hesaplanıp, bu sayede robot hedefe yaklaştıkça hızı düşmektedir. Lokal yaklaşımdan global yaklaşıma geçmek için robot, ilgili bloktaki ortalama varyansın belirlenen eşik değerinden düşük olması durumunda gerçekleştirir. Bloklar arası gezinti işleminin bitirilip algoritmanın sonlandırılması için robot tüm blokların ortalama eş varyansların toplamının belirlenen global potansiyel eşik değeriyle kıyaslayarak kontrol eder. Robot ayrıca gidilen yol uzunluğunu hesaplar. Hareket yönü ve hızı hesapladığı bir sonraki gideceği lokasyon, gidilmiş yol uzunluğu ve bir sonraki lokasyon ile en yakındaki kıyı noktası arasındaki mesafelerin toplamı robotun gidebileceği maksimum menzilden büyükse robot kendisini en yakındaki kıyı şeridi lokasyonuna yönlendirir.

Önerilen yaklaşımın analizi için simülasyonlar bu tez çalışmasında yapılmıştır. Google Earth Pro'dan çekilmiş göl çevresine yol çizgileri çekilecek simülasyona eklenir. Yol çizgilerinden elde edilen poligon noktaları enlem ve boylam şeklinde olup bunlar doğu kuzey yönlü x ve y koordinatlarına çevrilerek metre cinsinden işlemlerin yapılabilmesi sağlanmıştır. Ayrıca tüm arazideki test hücrelerinin poligon içerisinde olup olmadığı Ray Casting Algoritması ile sağlanır. Burada her bir hücre için bir yatay bir doğru üzerinde olduğu düşünülür. Bu doğrunun kestiği poligon kenarını kesme sayısı hesaplanır. Tek sayıysa poligonun içerisinde, çift sayıysa poligonun dışarısında olduğu anlaşılır. Ayrıca göl, nehir gibi ortamlarda su eksilmesinden kaynaklı geri çekilmeler de sabit değerli bir uzaklık ile kontrol edilir. Burada sayı metre cinsinden olup poligon içerisindeki her bir hücre, belirlenen uzaklık içerisinde bir toprak parçası varsa o hücreyi yüzülebilir olarak kullanılamaz. Bu işlemler sonucunda regressörden tahmin edilen ölçüm noktaları çıktısı için verilecek olan test lokasyonları elde edilmektedir. Sentetik su verisini DO ve pH için elde etmek için ters uzaklık interpolasyon metodu simülasyon için kullanılmıştır. Yüzülebilir su hücrelerinden bazı noktalar belirlenmiş ve bu noktalara standartta var olan DO ve pH değerleri arasında değerler verilmiştir. Her bir hücre ile şeçilen noktalar arası uzakların karesi ile orantılı olacak şekilde değerler üzerinden sentetik su kalitesi haritası oluşturulmuştur. Suüstü robotun yaptığı ölçümlere gürültü faktörü eklenmiştir. Robot ilgili lokasyonu ölçüm noktaları listesine eklerken bu lokasyonlardan aldığı ölçüm değeri de gürültüye sahip olacaktır. Simülasyonlar ayrıca temel çizgi (baseline) algoritması ile kıyaslanmaktadır. Bu kapsamda bloklar arası global hakarette robot önerilen yaklaşımdan farklı olarak yerel yaklaşımda kullanılan maksimum varyansın hedef olarak uygulanmasına benzer bir yaklaşım kullanılmaktadır. Bu temel çizgi algoritmasında, bütün blokların ortalama maksimum varyans değeri hesaplanıp, robotun maksimum varyansın olduğu bloğun ağırlık merkezine kadar gitmesi global yaklaşım olarak baseline yaklaşımda gerçekleştirilir. Simülasyonlar süresinde çözünmüş oksijen (DO) ve pH su kalitesi sentetik haritası olarak kullanılacaktır. İki farklı su kalitesi haritaları, 4 farklı kernel ve Gauss süresi regresyon modelinin stokastik olması sebebiyle 5 tekrar test gölü üzerinde denenmiştir. Sonuçlar gidilen yol, yapılan ölçüm sayısı, DO ve pH'ın sentetik ve tahmin edilen değerlerinin mutlak farkının tüm hücreler için yapılarak ortalaması ve standart sapması üzerinden gerçekleştirilmektedir. Yapılan simülasyonlarda simülasyonun farklı tekrarlarına suüstü robotun farklı güzergahlar üzerinden gidebildiği görülmüştür. Bu Gauss süreci regresyon modelinin stokastik olmasından kaynaklıdır. Ayrıca farklı başlangıç noktalarından simülasyonlar gerçekleştirildiğinde, ilk ölçüm yapılan bloklar üzerinde yapılan genişlemeden dolayı farklı hareket noktaları, farklı ölçüm sayısı sonuçları ve farklı sayıda tahmin sonucu çıkabildiği gözlemlenmiştir. Farklı kerneller için yapılan simülasyonlarda, radyal bazlı fonksiyonun, rasyonel ikinci derecede kernelinden kötü sonuçlar verdiği gözlemlenmiştir. Radyal bazlı fonksiyonun seçilen hiperparametreleri ile birlikte kullanıldığında hataya fit olduğu, fazla eğitimdan dolayı görevi tamamlayamadığı gözlemlenmiştir. Ayrıca kerneller bir sabit kernel ile çarpıldığında elde edilen sonuçlarda, ilgili kernele kıyasla mesafe biraz daha artsa da daha sentetik veriye daha yakın tahmin sonuçları alınabildiği gözlemlenmiştir. Buna göre sabit kernel ile çarpılmış rasyonel ikinci derecede kernelin beyaz gürültü kerneli ile kullanımının en iyi sonuca vardığı kararlaştırılmıştır. Global yaklaşımda bloklar arası hareket için önerilen yapay potansiyel fonksiyonunun temel çizgi fonksiyonuna kıyasla ölçüm sayısı artsa da daha kısa toplam yol gidildiği görülmüş ve ayrıca mutlak fark sonuçlarında temel çizgi fonksiyonuna göre doğru sonuç verdiği gözlemlenmiştir. Bunun sebebi temel çizgi yaklaşımında çıkan maksimum varyans değeri genellikle ölçüm alınmış bloklara en uzak bloğu hedeflediğinden gidilen yolu uzatmaktadır. Öte yandan bizim önerdiğimiz yaklaşım bloklar arası gezerken çekici potansiyelin oluşturduğu akımla en yakın ziyaret edilmemiş bloğa doğru kayar. Bu durumda en uzak noktaya gidip ölçüm almak yerine blok blok gezerek lokal yaklaşımla ölçümler gerçekleştirmiş oluyor. Bu da gidilen yol konusunda tasarruf sağlamaktadır.

Bu tez çalışmasından yola çıkarak ileriye dönük yapılabilecek işlemlere bakıldığında bu tez çalışması mekânsal bazlı analiz yaparken zamansal bazlı da su kalitesi tahmini incelenebilir. Ayrıca başka insansız suüstü araçları ile çoklu robot sistemlerine giriş yapılarak belirli bloklarda kendi devriyesini oluşturan çoklu görev analizi eklenebilir veya çoklu suüstü robotları ile gerçekleştirilen su kalitesi izleme işlemi yapılabilir.

Anahtar Kelimeler: Su Kalitesi İzleme, Bilgilendirici Rota Planlama, Yapay Potensiyel Fonksiyon, Gauss Süreci Regresyon Modeli, İnsansız Suüstü Aracı.

ABSTRACT

Artificial Potential Function-based Water Quality Monitoring Using Unmanned Surface Vehicle

Emre Fikri Baltacı

Master Thesis, Electrical and Electronic Engineering Department

Supervisor: Asst. Prof. Haluk Bayram

October, 2023. 61 pages.

In this thesis, we consider the the informative path planning for water quality monitoring using artificial potential functions.

The area of terrain can be wide, so the whole area is discretized into blocks, and each block is also discretized into cells. On the other hand, the number of water quality values for monitoring is more than one, so the variances of uncorrelated water quality values are weighted summed to obtain an equivalent water variance.

In one block, a local mission is offered. After taking measurements and calculating Gaussian process regression in the local block, the robot aims to the following measurement location with the maximum uncertainty (maximum equivalent variance value) for each cell. The maximum uncertainty location has an attractive potential, and the borders of the lake have a repulsive potential for the unmanned surface vehicle (USV). Using attractive and repulsive potential and their gradients, the velocity and direction of the USV can be calculated for local missions.

Moreover, for transition operation between blocks, the robot uses a newly proposed artificial potential function for the global mission. Using attractive and repulsive potential and their gradients, the velocity and direction of the USV can be calculated for local missions. Moreover, for transition operation between blocks, the robot uses a newly proposed artificial potential function.

In the baseline approach, as in the local approach, the robot aims to calculate the following measurement block as the maximum value of mean equivalent variance and travels through to that block. With different combinations of the simulations, the

results are compared for different synthetic water quality maps, and different kernel functions for Gaussian process regression, according to traveled distance, number of measurements, and the accuracy of predicted value for each water quality value.

Keywords: Water Quality Monitoring, Informative Path Planning, Artificial Potential Function, Gaussian Process Regression, Unmanned Surface Vehicle.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . **iii**

GENİŞ ÖZET . . . . **iv**

ABSTRACT . . . . **ix**

LIST OF FIGURES . . . . **xiii**

LIST OF TABLES . . . . **xv**

LIST OF ALGORITHMS . . . . **xvi**

LIST OF ABBREVIATIONS . . . . **xvii**

1 INTRODUCTION **1**

1.1 Problem Statement 1

1.2 Related Literature 2

1.2.1 Gaussian Process for Mapping Water Quality Parameters 2

1.2.2 Path Planning 3

1.3 Contributions 4

1.4 Approach 5

1.5 Outline of the Thesis 6

2 WATER QUALITY MAPPING **7**

2.1 Gaussian Process 7

2.2 Kernel Functions 9

2.2.1 Constant Kernel 9

2.2.2 Radial Basis Function Kernel 9

2.2.3 Rational Quadratic Kernel 9

2.2.4 White Kernel 10

2.3 Gaussian Regression for Mapping Water Quality Parameters 10

3 MOTION PLANNING FOR WATER QUALITY MONITORING **12**

3.1 Informative Path Planning using Artificial Potential Function 12
3.1.1 Environment Discretization 13
3.1.2 Overall Approach 13
3.1.3 Global Approach 17
3.1.4 Local Approach 20
3.2 Energy Awareness 23

4 SIMULATIONS **24**
4.1 Simulation Settings 24
4.2 Baseline Approach 31
4.3 Simulation Results and Comparison 31
4.3.1 Test Lake 31
4.3.2 Bafa Lake 34

5 CONCLUSION **39**

REFERENCES **40**

LIST OF FIGURES

1.1 A USV following the path to take measurements in a lake. 2

3.1 Represents a new approach of dividing a area with local-global way . . . 13

4.1 The test lake. 24

4.2 Bafa Lake, Didim, Aydin, Turkey. 25

4.3 Ray Casting Algorithm. 26

4.4 Lake water receding. 27

4.5 DO and pH synthetic water quality map WQM-T1 for the test lake. . . 28

4.6 DO and pH synthetic water quality map WQM-T2 for the test lake. . . 28

4.7 DO and pH synthetic water quality map WQM-B for Bafa Lake. 29

4.8 A sample result from the simulations showing the USV trajectory and measurement locations at the test lake when our approach is used (the starting point is set to SP-1, the water quality map is set to WQM-T1). The yellow dot represents the starting point, magenta lines represent the robot's trajectory, and white stars represent measurement locations. . . 33

4.9 Our Approach at Bafa Lake: A sample result from the simulations showing the USV trajectory and measurement locations at Bafa Lake when our approach is used (the starting point is set to SP-1, the water quality map is set to WQM-B). The yellow dot represents the starting point, magenta lines represent the robot's trajectory, and white stars represent measurement locations. 35

4.10 Baseline Approach at Bafa Lake: A sample result from the simulations showing the USV trajectory and measurement locations at Bafa Lake when the baseline approach is used (the starting point is set to SP-1, the water quality map is set to WQM-B). 36

4.11 Averaged absolute difference between synthetic and predicted DO (upper plot) and pH (lower plot) values during block transition time for our proposed artificial potential function approach. 37

4.12 Averaged absolute difference between synthetic and predicted DO (upper plot) and pH (lower plot) values during block transition time for baseline approach. 38

LIST OF TABLES

4.1 Simulation results in which our approach is used with the settings WQM-T1, C*RQ+White and SP-1 32

4.2 The effects of the various kernels using our approach with the settings WQM-T1 and SP-1 33

4.3 The effects of the various kernels using our approach with the settings WQM-T2 and SP-1 34

4.4 Simulation results for **the proposed approach** for Bafa Lake with the settings WQM-B and SP-1 34

4.5 Simulation results for **the baseline approach** for Bafa Lake with the settings WQM-B and SP-1 35

LIST OF ALGORITHMS

2.1 `GPR-Calculation` 8

2.2 `GPR-Mapping` 11

3.1 `Overall approach` 15

3.2 `Mission-Check` 16

3.3 `Global-Mission` 20

3.4 `Multiple-WQP` 21

3.5 `Local-Mission` 23

3.6 `Energy-Awareness` 23

LIST OF ABBREVIATIONS

APF	Artificial potential function
C	Constant kernel
DO	Dissolved oxygen
GP	Gaussian process
GPR	Gaussian process regression
loc_dis_thr	Local distance threshold
loc_var_thr	Local variance threshold
Meas.	Measurement
pH	Potential of Hydrogen
RBF	Radial basis function
RQ	Rational quadratic
USV	Unmanned surface vehicle
WQP	Water quality parameter

CHAPTER 1: INTRODUCTION

The effects of global warming on water resources have made water an increasingly crucial component in human life, impacting areas such as food production, drinking, cleanliness, and overall health [Şen, 2009, Fazai et al., 2019]. Assessing seawater pollution, determining coastal areas recommended for swimming, measuring the quality of drinking water in lakes and reservoirs, and predicting damage to water resources following chemical warfare are some noteworthy examples of its significance in governmental, municipal, and military domains.

Thanks to advancements in science, successful studies have been conducted in the field of environmental sensing and monitoring, encompassing physical, chemical, and biological observations. However, a significant challenge in the environmental monitoring domain lies in the substantial amount of work and effort required for observations across vast landscapes. In contrast to the traditional approach of using fixed sensors for environmental sensing, the emergence of autonomous robotic systems has begun to outpace traditional methods, particularly in water, air, and land environments [Ma et al., 2017]. In light of this, our study introduces a novel approach to informative path planning for the monitoring of water quality across large-scale water bodies using autonomous unmanned surface vehicles.

1.1 Problem Statement

The challenge at hand involves the autonomous monitoring of multiple water quality parameters and the generation of corresponding water quality map for each parameter by a robotic system. To accomplish this task, the robot must determine optimal measurement locations and plan its waypoints while avoiding obstacle. Figure 1.1 provides an illustrative example of the problem.

Figure 1.1: A USV following the path to take measurements in a lake.

1.2 Related Literature

Various water quality monitoring studies have been conducted over time. This study relates to the existing literature in the following areas: water quality mapping using Gaussian process regression, informative path planning, and energy awareness

1.2.1 Gaussian Process for Mapping Water Quality Parameters

In the context of water quality measurement and monitoring, machine learning and artificial neural networks can be used to obtain information from unmeasured areas. However, artificial neural networks require a large amount of pre-information for training and parameter optimization. Since there may be a small amount of training data available, machine learning algorithms are more suitable for this work [Duvenaud, 2014]. Machine learning algorithms can handle regression and classification problems. Thanks to regression, the prediction of unknown quantities can be made like interpolation [Williams and Rasmussen, 2006]. With this preliminaries Gaussian Process (GP) is selected for this thesis, which is a regression model with a probabilistic approach. Gaussian process regression can use train and test inputs as a 2D coordinate system, which makes it useful in fields such as water, atmosphere monitoring, and solar analysis etc. [Stachniss et al., 2009, Marchant and Ramos, 2014, Plonski et al., 2013].

In one of the studies on water quality monitoring, a new algorithm is developed for information-sensitive sampling locations using the region of interest method in the water quality measurement domain [Tan et al., 2018]. In this study, the upper confidence bound of the Gaussian process is used to balance the exploration of unmeasured regions with the exploitation of the mean function of the Gaussian process. This work focuses on one quality parameter, while we investigate more parameters.

Also, for Gaussian process regression, the kernel function is an important variable that affects the performance of water quality mapping. [Zare Farjoudi and Alizadeh, 2021] analyze different kernel functions (radial basis function, rational quadratic, exponential, Matérn) for river water quality coverage. According to this investigation, we use different kernel functions to analyze their effects on the water quality mapping. On the other hand, we multiply the tested kernels with the constant kernel for influence over the kernel, as defined in the work described in [Pedregosa et al., 2011] and investigated in this thesis.

In [Lal and Datta, 2018], genetic programming and Gaussian process regression are compared for monitoring groundwater salinity prediction. As a result of this work, Gaussian process regression gives more accurate results in terms of root mean square error, correlation, and Nash-Sutcliffe efficiency coefficient.

In [Peralta et al., 2021], they propose a new approach for multi-function estimation of water quality parameters using Bayesian optimization and acquisition function fusion. They compare their approach to the predictive entropy search for multi-objective optimization with constraints (PESMOC) algorithm and the genetic algorithm (GA). In this thesis, we propose a new algorithm that handles multiple water quality parameters using discrete random variables.

1.2.2 Path Planning

Water quality mapping can be accomplished by informative path planning. The robot's movement and optimal measurement-taking location from the terrain are part of the informative path planning problem in online learning. The informative path planning problem is handled by various studies.

[Mishra et al., 2018] offered a new effective path planning algorithm with Gaussian process regression for one quality parameter, and it was compared with a standard offline predefined path like a lawn mower algorithm. The new Adaptive Path Planning (AdaPP) algorithm solves the path planning in constrained duration. In a specific duration, AdaPP gives a more accurate solution over lawn mover in terms of root mean square error.

In [Li, 2018], the grid-based survey planning problem is investigated. The terrain is discretized into hexagonal areas, and a new path planner method is developed for area coverage. This method is useful, but for huge terrain areas, it can become difficult and consume more energy.

In [Pamosoaji and Hong, 2013], the entire terrain is divided by triangular shaped areas. Among these areas, a collision-free, goal-point-oriented motion algorithm is proposed using a potential function. The approach mentioned in [Pamosoaji and Hong, 2013] and the approach proposed in this thesis study are similar in that the entire terrain is divided into blocks of a certain shape. However, in our approach, the potential function does not need a goal for traveling among the blocks. Small changes in the attractive potential function value cause the system to stop. In our approach and their study, repulsive potentials are used for the same purpose: to escape collisions.

In [Kathen et al., 2021], a new particle swarm optimization (PSO) approach is developed. This enhanced GP-based PSO uses a fitness function, which is related to the location and speed of the USV. This fitness function targets maximum error locations in local and global areas.

[Ajanović et al., 2018] propose a novel model-based heuristic for energy-optimal motion planning, calculating the cost of transitioning operations to specific locations. To prevent unwanted situations such as running out of battery in the middle of the lake, we also propose considering the energy cost during the operation.

1.3 Contributions

We propose a new approach to informative path planning strategies using the integration of artificial potential function and Gaussian process regression. The main contributions in this thesis are as follows:

- Developing an artificial potential function-based approach for informative path planning,
- Developing a local/global-based approach for wide water bodies,
- Monitoring multiple water quality parameters using Gaussian process regression,
- Developing a simulation environment in Python for lake/pond/sea/gulf using only water borders and sensitivity parameters.

1.4 Approach

In this thesis, we develop a new approach to water quality monitoring using an unmanned surface vehicle (USV). Our approach is designed to accommodate extensive areas and address multiple water quality parameters simultaneously.

Water has physical, chemical, biological, and other characteristics that are used to assess its quality. Among the most common monitored water quality characteristics are temperature, pH, color, turbidity, total solids, acidity, dissolved oxygen (DO), electrical conductivity, alkalinity, chloride, and coliform organisms [Tebbutt, 1997]. In this thesis, a subset of these characteristics is selected for water quality monitoring and mapping.

To tackle the challenge of covering large areas, we introduce a local/global approach in which we discretize the entire region into local and global sections. The global area is further subdivided into grid-shaped blocks, each containing a predefined set of sampling points. Within these sections, the system employs distinct algorithms for water quality mapping and motion planning to optimize efficiency.

To calculate the optimal measurement locations, we use Gaussian Process Regression (GPR), a machine learning regression method. GPR provides both the mean and variance of water quality parameters at each sampling point. Our objective is to minimize variance while ensuring full coverage of the entire terrain.

The USV autonomously navigates through the water body and takes measurements within its local block, unless the variance is lower than the given threshold value. After that, the USV systematically traverses from one block to another. These movements, occurring both within and between the blocks, are provided via the potential function, which serves as our path-planning method. This approach allows us to comprehensively monitor water quality across the entire area.

Due to the limited battery capacity of the USV, it also employs a distance-based strategy for waypoint calculations. The USV determines whether or not the distance to the nearest shore of the lake falls below a predefined threshold. If so, it autonomously redirects itself to a recharging station to ensure continuous operation.

1.5 Outline of the Thesis

The thesis consists of five main chapters. Chapter 1 serves as the introduction, encompassing a problem statement, a literature survey, the thesis's contributions, the proposes approach, and an outline of the subsequent sections. In Chapter 2, water quality mapping is described in detail. In Chapter 3, a novel method for informative motion planning is explained. In Chapter 4, the simulation platform, parameters, baseline approach, and results are provided. In Chapter 5, the results are discussed.

CHAPTER 2: WATER QUALITY MAPPING

Since our aim is to monitor multiple water quality parameters in a lake/pond using an autonomous surface vehicle, Gaussian process regression (GPR) is used for mapping these parameters. The output of the GPR is then used as an input to the informative path planner explained in Chapter 3.

2.1 Gaussian Process

Gaussian Process (GP) is a collection of random variables, which is any finite number of which have joint Gaussian distribution. GP is defined with a mean function $\mu(x)$, a real process function $f(x)$ and covariance (kernel) function $k(x, x')$ [Williams and Rasmussen, 2006].

$$\mu(x) = E[f(x)] \tag{2.1}$$

$$k(x, x') = E[(f(x) - \mu(x))(f(x') - \mu(x'))] \tag{2.2}$$

According to this, GP is defined by Equation 2.3:

$$f(x) \sim \mathcal{GP}(\mu(x), k(x, x')) \tag{2.3}$$

The joint Gauss distribution is given in the following equation:

$$\begin{bmatrix} f \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix} \right) \tag{2.4}$$

where f_* is a function output based on training points, $K = k(X, X)$ is $N \times N$ dimensional covariance of training pairs, $K_* = k(X, X_*)$ is $N \times N_*$ dimensional sub-covariance matrix of training and test points, $K_{**} = k(X_*, X_*)$ is $N_* \times N_*$ dimensional sub-covariance matrix of test points.

For noisy observations, we assume that the Gaussian noise ϵ is distributed with variance σ^2. The prior on noisy observations is described in Equation 2.5 [Williams and

Rasmussen, 2006]:

$$cov(y) = K + \sigma_n^2 I \tag{2.5}$$

When a noise term is added to Equation 2.4, the joint distribution of the observed target and function values at the test location is explained in Equation 2.6 [Williams and Rasmussen, 2006].

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K + \sigma_n^2 I & K_* \\ K_*^T & K_{**} \end{bmatrix}\right) \tag{2.6}$$

Using Equation 2.6, the predictive distribution equations for Gaussian process regression can be written as Equation 2.7-2.9 [Williams and Rasmussen, 2006].

$$f_*|X, y, X_* \sim N(\overline{f}_*, cov(f_*)) \tag{2.7}$$

$$\overline{f}_* = E[f_*|X, y, X_*] = K_*^T [K + \sigma_n^2 I]^{-1} y \tag{2.8}$$

$$cov(f_*) = K_{**} - K_*^T [K + \sigma_n^2 I]^{-1} K_* \tag{2.9}$$

According to these equations, the general approach can be summarized as Algorithm 2.1 [Williams and Rasmussen, 2006]. In the GPR calculations, the Cholesky decomposition is used to solve the matrix inversion.

Algorithm 2.1: `GPR-Calculation`

Input: X, y, k, σ_n^2, x_*

/* X: inputs, y: targets, k: covariance function, σ_n^2: noise level, x_*: test input */

Output: $\overline{f}_*$, $Var[f_*]$

/* $\overline{f}_*$: mean, $Var[f_*]$: variance */

$L = cholesky(K + \sigma_n^2 I)$;

$\alpha = L^T \backslash (L \backslash y)$;

$\overline{f}_* = k_*^T \alpha$;

$v = L \backslash k_*$;

$Var[f_*] = k_{**} - v^T v$;

return $\overline{f}_*$, $Var[f_*]$

2.2 Kernel Functions

Gaussian process regression (GPR) is a type of regression model in machine learning that presents the most probable range of values in a function whose measurement values are given as input at some points [Williams and Rasmussen, 2006]. As defined in Equation 2.4, the kernel function K must be defined in GPR. The kernel function determines how the model covers the input and predicts the output based on the input [Duvenaud, 2014]. This kernel function has various types such as the constant function, radial basis function (RBF), periodic function, and rational quadratic function, white kernel. The appropriate kernel function is selected as a result of operations such as addition, subtraction, or multiplication.

2.2.1 Constant Kernel

The constant kernel (C) is used in product-kernels to scale the amplitude of the kernel, and in sum-kernels to rearrange the mean of a Gaussian process. It is defined as Equation 2.10 [Pedregosa et al., 2011].

$$k(x, x') = \alpha \forall x, x' \tag{2.10}$$

where α is a constant value.

2.2.2 Radial Basis Function Kernel

The radial basis function (RBF), which is also known as the squared exponential function, is a stationary kernel. It has a hyperparameter l as an input, which defines the characteristic length scale of the kernel. The kernel is defined as Equation 2.11 [Duvenaud, 2014].

$$k(x, x') = exp\left(-\frac{d(x, x')^2}{2 * l^2}\right) \tag{2.11}$$

where l is characteristic length scale and d is Euclidean distance.

2.2.3 Rational Quadratic Kernel

The rational quadratic (RQ) kernel is derived from the RBF kernel as an infinite sum with different characteristic length scales. The kernel is defined as Equation 2.12. In this equation [Pedregosa et al., 2011], l is the characteristic length scale, which controls the smoothness of the function that is being modeled, and α is the scale mixture parameter, which controls the variance of the noise in the data.

$$k(x, x') = \left(1 + \frac{d(x, x')^2}{2 * \alpha * l^2}\right)^{-\alpha} \tag{2.12}$$

2.2.4 White Kernel

White kernel, which is used for modeling the noise in the data as normally distributed and independent, is named from the white noise signal. It is generally used to combine with other kernels [Duvenaud, 2014, Pedregosa et al., 2011]. Equation 2.13 describes the white kernel:

$$k(x, x') = \sigma_n^2 * I \tag{2.13}$$

where σ_n^2 is variance and I is identity matrix.

2.3 Gaussian Regression for Mapping Water Quality Parameters

Gaussian process regression is used for generating a map for each water quality parameter in the whole terrain with obtained data from measurement points.

The GPR can take an input for X_{train} and X_{test}, not only a vector but also an mxn matrix. In Equations 2.14 and 2.15, X_{train}, y_{train} are measured water quality inputs, and their measured locations. X_{test} is the input location matrix of the model for prediction, y_{test} is the output matrix for the predicted value of the model, σ_{test} is the standard deviation matrix for the predicted output of the model [Pedregosa et al., 2011].

$$X_{train} = \begin{bmatrix} x_{meas,1} & y_{meas,1} \\ x_{meas,2} & y_{meas,2} \\ \vdots & \vdots \\ x_{meas,m} & y_{meas,m} \end{bmatrix}, \ y_{train} = \begin{bmatrix} y_{meas,1} \\ y_{meas,2} \\ \vdots \\ y_{meas,m} \end{bmatrix} \tag{2.14}$$

$$X_{test} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_m & y_m \end{bmatrix} \quad y_{test} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \quad \sigma_{test} = \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_m \end{bmatrix} \tag{2.15}$$

Algorithm 2.2: `GPR-Mapping`

Input: X_{train}, y_{train}, X_{test}
Output: y_{test}, σ_{test}, σ_{avg}, and Var_{test}
1 define the kernel function;
2 define the GPR model;
3 fit the GPR for input X_{train}, y_{train};
4 predict y_{test}, σ_{test} and Var_{test} from the model;
5 Calculate average of σ_{test} as σ_{avg}

Gaussian process regression (GPR) algorithm is provided in Algorithm 2.2. During the simulations, Algorithm 2.2 is used with the Scikit-learn Python library to perform water quality mapping. To define the kernel, the RBF and RQ kernels, which are commonly used for spatial domain applications, are used and compared. Additionally, these kernels are multiplied with a constant kernel. Moreover all kernels are added with a white kernel to account for noisy sensor measurements. Algorithm 2.2 can be used for each water quality parameter separately. The output σ_{avg} is used as an input for calculating the global attractive potential and its gradient. The maximum value of Var_{test} is used to determine the next measurement location in the local approach. The total Var_{test} is used to switch the operation mode between local and global.

CHAPTER 3: MOTION PLANNING FOR WATER QUALITY MONITORING

In a monitoring task, the measurement value and the location of the measurement are two of the inputs. To address this need, a new method of informative path planning is detailed in Chapter 3. The uncertainty output from Gaussian process regression serves as an input for informative path planning to determine movement to a specific location. On the other hand, the measurement location outputs from informative path planning, along with measurement values, serve as inputs for Gaussian process regression in the next step of monitoring. In this way, both topics support each other and the simultaneous movement and monitoring operations are executed.

Additionally, the developed approach has an energy awareness feature to prevent problems like running out of battery and interruptions during operation. The approach developed in this case study is described in Section 3.2.

3.1 Informative Path Planning using Artificial Potential Function

Environmental monitoring tasks can be accomplished using two distinct measurement methods: employing a stationary measurement station or employing a mobile measurement system. For example, air quality monitoring can be accomplished from fixed stations or with balloons sent into the air with sensor devices [Sankar and Norman, 2009]. Similarly, for water quality monitoring, measurements can be gathered from fixed stations positioned in lakes, rivers, seas, and gulfs, as well as from boats on the water's surface. In cases where measurements from a limited number of shoreline locations are not enough to cover the entire area, online measurements can be conducted using a mobile measurement system. In this context, selecting the optimal measurement location problem addresses to the informative path planning approach.

The monitored terrain can be a small lake or a wide sea. The method recommended for covering wide regions is explained in Section 3.1.1, while the general algorithmic approach is outlined in Section 3.1.2. The application of the land discretization suggested method in Section 3.1.1 is detailed as a global solution in Section 3.1.3 and as a local solution in Section 3.1.4.

3.1.1 Environment Discretization

For water quality monitoring, the application terrain can be a large area. Randomly zigzag movements in the terrain require a lot of effort and energy for monitoring operations. In this study, we propose a new approach to address this problem.In the global approach, the area is discretized into blocks. In the local approach, the blocks are further divided into cells. The size of each cell is related to the sensor's sensitivity. Figure 3.1 illustrates this explanation.

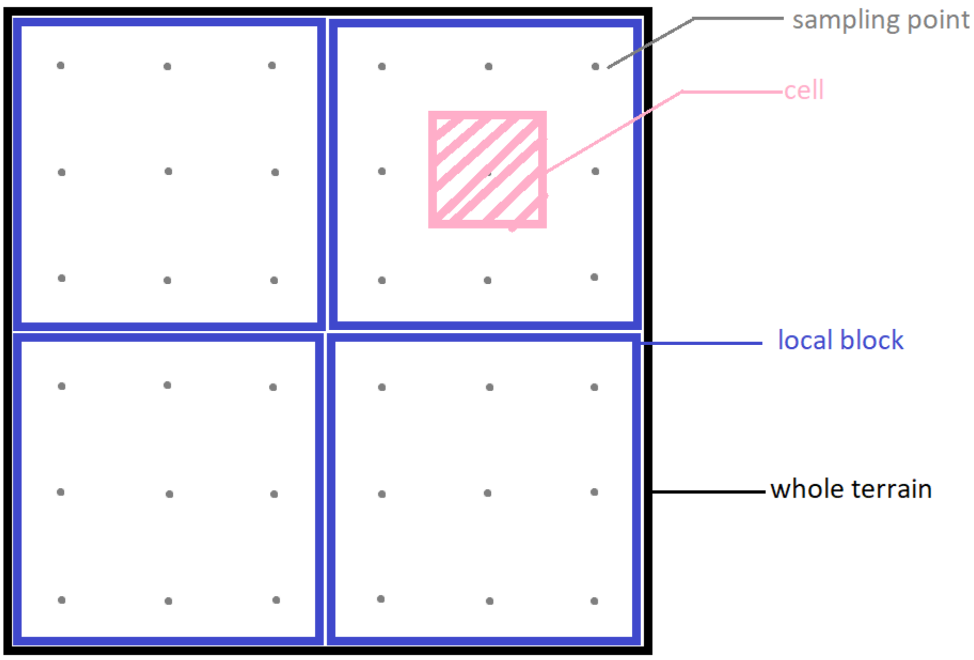

Figure 3.1: Represents a new approach of dividing a area with local-global way

3.1.2 Overall Approach

To autonomously monitor water quality in various locations like lakes, ponds, seas, etc., a robot must locate the most suitable waypoints. This challenge is approached using informative path planning. After completing water measurements, the robot must choose a new and optimal measurement location. Through online learning, this new site should prioritize energy efficiency and dependable water quality data.

Considering this, the general algorithm is introduced in this section, taking energy efficiency into account. As previously mentioned, the entire area is discretized into blocks, and these blocks are further discretized into cells. Within this landscape, the robot carries out both global and local missions. During the global mission, the robot moves between blocks until a global stop criterion is met. Meanwhile, during the local mission, the robot moves between cells within a single block until a local stop criterion is met.

The local stop criteria triggers a switch in the algorithm's operation mode between local and global. The global stop criteria brings the algorithm to an end. Both stop criteria are set to ensure that the uncertainty value, for both local and global missions, falls below the desired threshold in terms of variance.

Throughout this calculation period of each while loop process, the duration Δt is measured. Moreover, the outputs of the local and global missions are provided to the robot's angle and velocity. With these outputs, the following location is controlled using the Energy-Awareness Algorithm 3.6. With these outputs, the motion command is sent to the robot. The complete process is outlined in Algorithm 3.1.

By the way, the *loc* array is the array where location information is kept, the *mloc* array is the array where measured location information is kept. For each water quality parameter (WQP), measured values are kept in the *mv* array.

The entire process is controlled by the Mission Check Algorithm, which monitors the mission status. When the mission is local, the algorithm calculates the distance ($d_{goal,r}$) between the robot's location and the highest uncertainty in the block. The algorithm keeps the mission local until this distance is less than the desired distance value. If not, it sets the *localswitch* flag to 1, indicating that a new measurement is far enough away. In this case, the algorithm also calculates the average variance of each cell within the block. If this average is lower than the desired value, the mission switches to global

Algorithm 3.1: `Overall approach`

/* generate 3-dimensional X_{test} list */

for blockid in range # of blocks **do**
  for cellid in range # of cells **do**
    $X_{test,blockid}$.append(x_{cellid},y_{cellid})
  X_{test}.append($X_{test,blockid}$)
$mission$ = LOCAL;
$globalswitch$=0;
$localswitch$=1;
r_x, r_y=startpoint;
loc.append([r_x, r_y])// add location to array
mloc.append([r_x, r_y])// add location to array
/* take measurement and add to mv array for each wqp */
for wqid in range(# of WQP) **do**
  mv_{wqid}.append(value);
while True **do**
  if $mission$==LOCAL **then**
    V,θ= `Local-Mission`(mloc, mv$_{wqid}$,X_{test})
  else if $mission$==GLOBAL **then**
    V,θ= `Global-Mission`(mloc,mv$_{wqid}$,X_{test});
  loc=`Energy-Awareness`(V, θ,loc) // Check the following iteration is suitable for travel
  Send Motion Command (V, θ);
  $mission$, $exitflag$ = `Mission-Check`($mission$, $globalswitch$, $localswitch$, loc, $mloc$, mv_{wqid}, $maxeqvar$);
  if $exitflag$==True **then**
    break;

and *globalswitch* is set to 1. The *localswitch* flag is used to control whether Gaussian process regression is running or not in the local mission. The *globalswitch* flag is used to run Gaussian process regression for each block when starting the global mission.

In the global mission, the algorithm checks the maximum attractive potential of unvisited blocks in the temporal domain. If the potential change is repeated, indicating a stationary situation, then the mission mode is switched to local, *localswitch* is set to 1, the new measured value is added to the *mv* list, and the final location is added to the *mloc* list. The algorithm also checks the variance normal of each block. If this variance normal is below the desired value, it sets the *exitflag* to true, which terminates the while loop in the overall algorithm and terminates the entire process.

Algorithm 3.2: `Mission-Check`

Input: $mission$, $globalswitch$, $localswitch$, loc, $mloc$, mv_{wqid}, $maxeqvar$
Output: $mission$, $exitflag$

r=loc[-1]`// set r to last element of loc array`
if mission==LOCAL **then**
    $distance$=$d_{goal,r}$;
    if $distance$<loc_dis_thr **then**
        localswitch=1;
        for wqid in range(# of WQP) **do**
            mv_{wqid}.append(value)`// take measurement and add it to mv list for each WQP`
        $mloc$.append($[r_x, r_y]$);
        if $maxieqvar$<$loc - var - thr$ **then**
            $mission$=GLOBAL;
            $globalswitch$=1
else
    $maxiUatt$=find maximum attractive potential in the Equation 3.6;
    if value change between $maxiUatt_{att,t-3}$ and $maxiUatt_{att,t}$ **then**
        $localswitch$=1;
        $globalswitch$=0;
        for wqid in range(# of WQP) **do**
            mv_{wqid}.append(value)`// take measurement and add it to mv list for each WQP`
        $mloc$.append($[r_x, r_y]$);
        $mission$=LOCAL
    $Var_{eq}list$=`Multiple-WQP`(Var_{avg}list for each wqp);
    $sumvar$=sum($Var_{eq}list$);
    if $sumvar$<glb_var_thr **then**
        $exitflag$=True

3.1.3 Global Approach

In the global approach, a potential function is used for path planning between blocks. A potential function is a differentiable function that can be used to generate a robot's movements in a vector flow. This makes it suitable for path planning algorithm design.

The robot moves in a gradient vector field, which can be visualized as the flow between a positively charged robot and a negatively charged goal. The potential function is an energy, and its gradient represents a force. This gradient is a vector that shows the direction of maximum energy difference [Choset et al., 2005].

The potential function is denoted by U and contains attractive and repulsive potentials. The attractive potential represents the energy between the robot and its goal location, and the repulsive potential represents the energy between the robot and an obstacle. The obstacle has the same charge as the robot, which forces them to avoid each other.

$$U(q) = U_{att}(q) + U_{rep}(q) \tag{3.1}$$

For easy setup, let the attractive function be related to the distance between the robot and the goal. The equations for the attractive function and its gradient are shown in Equations 3.2 and 3.3 [Choset et al., 2005].

$$U_{att}(q) = \frac{1}{2}\zeta d^2(q, q_{goal}) \tag{3.2}$$

$$\nabla U_{att}(q) = \zeta(q - q_{goal}) \tag{3.3}$$

For easy calculation, let the repulsive function be inversely proportional to the distance between the robot and the closest obstacle $D(q)$. The equations for the repulsive function and its gradient are shown in Equations 3.4 and 3.5 [Choset et al., 2005].

$$U_{rep}(q) = \frac{1}{2}\eta \left(\frac{1}{D(q)} - \frac{1}{Q*} \right)^2 \tag{3.4}$$

$$\nabla U_{rep}(q) = \eta \left(\frac{1}{Q*} - \frac{1}{D(q)} \right) \nabla D(q) \tag{3.5}$$

In the global approach, when the robot aims to move between blocks, it uses an adaptive potential function algorithm. Distances between visited and unvisited blocks and the robot, standard deviations of the GPR for each water quality parameter of each block, and the distance between the robot and border locations of water terrain are parameters of potential function.

In the new approach, the adaptive potential function is the addition of attractive and repulsive potentials. The attractive potential(U_{att}) is related to the distance between the robot and the middle point of unvisited blocks, and the average standard deviation of the GPR results for each water quality parameter in the block. The repulsive potential has not only an effect on visited blocks ($U_{rep,block}$), but also an effect on the border of the terrain ($U_{rep,border}$).

The borders are simulated by the collection of straight lines. The border lines are divided into a list of points at specific distances. The distance between each point and the robot's location is part of the repulsive potential. The potential functions are defined in Equations 3.6, 3.7, and 3.8.

$$U_{att} = \sum_{n=1}^{N_u} \left(\frac{\alpha_{att} * \prod_{n=1}^{N_{wqp}} \sigma_{avg,c}}{d^2_{n^{th} block,r}} \right) \tag{3.6}$$

$$U_{rep,block} = \sum_{n=1}^{N_v} \frac{\alpha_{rep,block}}{d^2_{n^{th} block,r}} \tag{3.7}$$

$$U_{rep,border} = \begin{cases} \sum_{n=1}^{N_b} \left(\alpha_{rep,border} \left(\frac{1}{d_{n^{th} border,r}} - \frac{1}{\alpha_{dist}} \right)^2 \right) & d_{n^{th} border,r} < \alpha_{dist} \\ 0 & else \end{cases} \tag{3.8}$$

where $N_u$is the number of unvisited blocks. N_v is the number of visited blocks. N_{wqp} is the number of water parameters. $\sigma_{avg,c}$ is the average standard deviation of the water parameters. r is the robot's location. N_b is the number of border points. α_{att}, $\alpha_{rep,block}$, and $\alpha_{rep,border}$ are constants for the attractive, repulsive block, and repulsive border potentials, respectively. α_{dist} is the distance constant, which is used to determine how far away the border points affect the robot. d is the distance function.

The gradient of the attractive potential is calculated by taking a partial derivative with respect to x and y, and it is shown in Equations 3.9 and 3.10:

$$\nabla U_{att,x} = \sum_{n=1}^{N_u} \left(\frac{\alpha_{att} * \prod_{n=1}^{N_{wqp}} \sigma_{avg,c} * (n^{th}block_x - r_x)}{d^3_{n^{th}block,r}} \right) \tag{3.9}$$

$$\nabla U_{att,y} = \sum_{n=1}^{N_u} \left(\frac{\alpha_{att} * \prod_{n=1}^{N_{wqp}} \sigma_{avg,c} * (n^{th}block_y - r_y)}{d^3_{n^{th}block,r}} \right) \tag{3.10}$$

The gradients of the repulsive block and repulsive border potentials are also calculated by taking partial derivatives with respect to x and y, and they are described between Equations 3.9 and 3.14.

$$\nabla U_{rep,block,x} = \sum_{n=1}^{N_v} \left(\frac{\alpha_{rep,block} * (r_x - n^{th}block_x)}{d^3_{n^{th}block,r}} \right) \tag{3.11}$$

$$\nabla U_{rep,block,y} = \sum_{n=1}^{N_v} \left(\frac{\alpha_{rep,block} * (r_y - n^{th}block_y)}{d^3_{n^{th}block,r}} \right) \tag{3.12}$$

$$\nabla U_{rep,border,x} = \begin{cases} \sum_{n=1}^{N_b} \left(\frac{2 * \alpha_{rep,border} * (n^{th}border_x - r_x) * (d_{n^{th}border,r} - \alpha_{dist})}{\alpha_{dist} * d^4_{n^{th}border,r}} \right) & d_{n^{th}border,r} < \alpha_{dist} \\ 0 & else \end{cases} \tag{3.13}$$

$$\nabla U_{rep,border,y} = \begin{cases} \sum_{n=1}^{N_b} \left(\frac{2 * \alpha_{rep,border} * (n^{th}border_y - r_y) * (d_{n^{th}border,r} - \alpha_{dist})}{\alpha_{dist} * d^4_{n^{th}border,r}} \right) & d_{n^{th}border,r} < \alpha_{dist} \\ 0 & else \end{cases} \tag{3.14}$$

The sum of the gradients for x and y provides the direction towards which the robot should move. The gradient of the total potential, ∇U_{total}, is used to calculate the direction angle using the arc-tangent function.

$$U_{total} = U_{att} + U_{rep,block} + U_{rep,border} \tag{3.15}$$

$$\nabla U_{total,x} = U_{att,x} + U_{rep,block,x} + U_{rep,border,x} \tag{3.16}$$

$$\nabla U_{total,y} = U_{att,y} + U_{rep,block,y} + U_{rep,border,y} \tag{3.17}$$

According to these potential equations, Algorithm 3.3 can be used to move the robot block by block.

Algorithm 3.3: `Global-Mission`

Input: $mloc$, mv_{wqp},X_{test}
Output: V,θ

define α_{att}, $\alpha_{rep,block}$, $\alpha_{rep,border}$ and $\alpha_{dist} constant$;
if globalswitch==1 **then**
    for wqid in range(N_{wqp}) **do**
        for blockid in range(# of blocks) **do**
            σ_{avg}, and Var_{avg}=`GPR-Mapping`($mloc$,mv,X_{test});
            $\sigma_{avg,wqid}$ list.append(σ_{avg});
            $Var_{avg,avg}$ list.append(Var_{avg});
    globalswitch=0
calculate Equations 3.6, 3.7 and 3.8`// calculate attractive and repulsive potentials`
calculate Equations 3.9-3.14`// calculate partial derivatives`
calculate Equations 3.15,3.16 and 3.17`// calculate sum of potentials and its derivatives`
θ =arctan($\nabla U_{total,y}/\nabla U_{total,x}$);
V=Max speed of boat;
return V,θ

In Algorithm 3.3, when the mission is changed from local to global, the GPR algorithm is first calculated for each block and for each water quality parameter only once. The GPR output of the standard deviations of each block and water quality parameter are used to calculate the attractive potential and its gradient. The attractive and repulsive potentials and their gradients are then calculated separately. The direction angle of the robot is calculated using the sum of the gradients (partial derivatives for x and y) of the potentials. For a global walk, the velocity of the boat can be selected to its maximum speed.

3.1.4 Local Approach

In the local approach, the robot must finish exploring a block before moving on to the next block, until the maximum variance of the cells in the current block is less than a threshold.

This study addresses the monitoring of multiple water quality parameters. It is assumed that these water quality parameters are not correlated with each other. This means that their standard deviations and variances are not related to each other, and they can be summed together according to Equation 3.18 [Yates and Goodman, 2014].

$$Var[A+B] = Var[A] + Var[B] \tag{3.18}$$

When the GPR algorithm is run for each water quality parameter in a local block, each parameter has a different mean and variance output. The goal of the local approach is to minimize the uncertainty for each water quality parameter by taking samples from locations with the maximum variance. To handle multiple water quality parameter variance values, Equation 3.18 is used to calculate a new equivalent standard deviation.

According to this information, Algorithm 3.4 explains how it works. A list of variances of water quality parameters in the block is weighted summed and the algorithm returns the equivalent variance. The location of the maximum equivalent variance is used to select the target location for the next measurement location.

Algorithm 3.4: `Multiple-WQP`

Input: # of Var_{wqp}
Output: σ_{eq}
set a weight for each wqp α_{wqp} s.t. the sum of the weights is equal to 1;
calculate equivalent variance $Var_{eq} = \sum_{i=1}^{N_b} {}_{wqp,i} * \alpha_{wqp,i}$;
return Var_{eq}

When the mission is local, the robot aims to move to the location with the maximum uncertainty. During this operation, the potential function algorithm is also needed to prevent the robot from crashing into obstacles (borders). In this case, the standard potential function is used. The attractive potential occurs between the robot's location and the location with the maximum uncertainty. Additionally, the repulsive potential occurs between the robot's location and the border locations at a certain distance α_{dist}. Using this information, the attractive and repulsive functions along with their partial derivatives are given in Equations 3.19-3.24.

$$U_{att,local} = \alpha_{att,local} * d_{goal,r}^2 \tag{3.19}$$

where $\alpha_{att,local}$ is a constant for the attractive potential in the local approach and $d_{goal,r}$ is the distance between the goal (maximum uncertainty location) and the robot's location.

$$\nabla U_{att,local,x} = \alpha_{att,local} * 2 * (goal_x - r_x) \tag{3.20}$$

$$\nabla U_{att,local,y} = \alpha_{att,local} * 2 * (goal_y - r_y) \tag{3.21}$$

$$U_{rep,local} = \begin{cases} \sum_{n=1}^{k} \left(\alpha_{rep,local} \left(\frac{1}{d_{n^{th}border,r}} - \frac{1}{\alpha_{dist}} \right)^2 \right) & d_{n^{th}border,r} < \alpha_{dist} \\ 0 & else \end{cases} \tag{3.22}$$

where $\alpha_{rep,local}$ is a constant for the repulsive potential in the local approach.

$$\nabla U_{rep,local,x} = \begin{cases} \sum_{n=1}^{k} \left(\frac{2*\alpha_{rep,local}*(n^{th}border_x - r_x)*(d_{n^{th}border,r} - \alpha_{dist})}{\alpha_{dist}*d^4_{n^{th}border,r}} \right) & d_{n^{th}border,r} < \alpha_{dist} \\ 0 & else \end{cases} \tag{3.23}$$

$$\nabla U_{rep,local,y} = \begin{cases} \sum_{n=1}^{k} \left(\frac{2*\alpha_{rep,local}*(n^{th}border_y - r_y)*(d_{n^{th}border,r} - \alpha_{dist})}{\alpha_{dist}*d^4_{n^{th}border,r}} \right) & d_{n^{th}border,r} < \alpha_{dist} \\ 0 & else \end{cases} \tag{3.24}$$

With this knowledge, the total potential and its gradients (partial derivatives for x and y) can be written in Equations 3.25-3.27

$$U_{total,local} = U_{att,local} + U_{rep,local} \tag{3.25}$$

$$\nabla U_{total,local,x} = \nabla U_{att,local,x} + \nabla U_{rep,local,x} \tag{3.26}$$

$$\nabla U_{total,local,y} = \nabla U_{att,local,y} + \nabla U_{rep,local,y} \tag{3.27}$$

According to these potential equations, Algorithm 3.5 can be written for movement in the local block. When the mission is changed from global to local, the GPR algorithm is first calculated for each water quality parameter in the block only once. Using the algorithm for multiple water quality parameters, the average variance is calculated for each cell in the block. The maximum variance value is noted for the mission check algorithm. The goal is also selected for the location of maximum variance. After that, the equations for the potential function and its gradients are calculated, and the direction angle is found. The velocity changes with the ratio between the last calculated attractive potential and the first calculated one. When the robot gets closer to the goal, the velocity decreases because the attractive potential decreases.

Algorithm 3.5: `Local-Mission`

Input: $mloc$, mv_{wqp},X_{test}
Output: V,θ
define $Vmax$;
if $localswitch{=}{=}1$ **then**
    for wqid in range(N_{wqp}) **do**
        `GPR-Mapping`($mloc$,mv,X_{test});
    Var_{eq}=`Multiple-WQP`(Var_{avg}list for each wqp);
    $maxieqvar$=max(Var_{eq});
    $goal$=find $maxieqvar$ location in block;
    $localswitch$=0
calculate Equations 3.19-3.24;
calculate Equations 3.25-3.27;
θ=arctan($\nabla U_{total,local,y}/\nabla U_{total,local,x}$);
V=$Vmax$*$\frac{U_{att,local}}{U^{1st}_{att,local,}}$;
return V,θ

3.2 Energy Awareness

The robot has a limited travel range in lakes, ponds, and seas. For example, if the robot runs out of battery in the middle of the terrain, the researcher will need to rescue it. To address this issue, we propose a new approach in this paper. The robot calculates and keeps track of its travel distance. Using this information, the robot can determine whether a new target movement point is feasible or not. This approach is given in Algorithm 3.6.

Algorithm 3.6: `Energy-Awareness`

Input: V,θ,loc
Output: loc
Energy cost value $cost$ is zero at initial;
define $energydistancelimit$;
$r_x = r_x + V * cos(\theta) * \delta t$;
$r_y = r_y + V * sin(\theta) * \delta t$;
$cost = cost + d_{loc[-1],r} + d_{r,nearstborderpoint}$;
if $cost< energydistancelimit$ **then**
    $cost = cost - d_{r,nearstboarderpoint}$;
    return loc
else
    loc.append(nearstboarderpoint);
    $cost$= 0;
    return loc

CHAPTER 4: SIMULATIONS

This chapter presents the simulations and their results using the approaches described in Chapters 2 and 3. Section 4.1 provides the details on the simulation environment, synthetic data generation, and selection of parameters. In Section 4.2, we apply our developed approach and the baseline approach using various settings. In Section 4.3, the simulation results are demonstrated and compared.

4.1 Simulation Settings

We first generate a test world to have a realistic environment. To achieve this, we employ Google Earth Pro for terrain selection. Using the waypoint feature in Google Earth Pro, we draw the borders of the lake terrain. The simulation program also requires two bookmarks, each covering the minimum and maximum latitude and longitude of the waypoint-defined borders. As a result, two water bodies with different shapes were generated for the simulations, which are shown in Figures 4.1 and 4.2. The dimensions of Bafa Lake and the test lake are approximately $15km \times 6km$ and $5.5km \times 5.5km$, respectively.

Figure 4.1: The test lake.

Figure 4.2: Bafa Lake, Didim, Aydin, Turkey.

These waypoints and bookmarks can be downloaded from Google Earth Pro in KML format and are used as input to the simulation program. These latitude and longitude coordinates are transformed into the east-north coordinate system, enabling calculations in meters.

As previously mentioned in Section 3.1.1, the water body is composed of cells, which are further categorized into water and soil cells. The detection of these categories is accomplished using the Ray Casting Algorithm [Ye et al., 2013]. In this algorithm, the border's waypoints define a polygon shape. Each horizontal grouping of cells is subjected to a ray, and the algorithm counts the number of intersections between each cell and the polygon's edges. If the count number is even, then the cell is considered outside the polygon; conversely, if the count number is odd, then the cell is considered inside the polygon. The algorithm is illustrated in Figure 4.3.

Each water body has a specific shape, such as rectangular or square. As proposed in this thesis, the terrain is discretized into blocks, and each block is discretized into cells. Two neighboring cells must have a distance less than a specific distance, which must cover the sampling location according to the sensor measurement range. With this distance, one block is generated with a size of $N \times N$ cells. According to the shape of the terrain, the horizontal and vertical blocks must provide enough coverage, and the number of horizontal and vertical blocks is related to the shape of the area. In this thesis, two shapes of area are investigated for simulations. One of them is selected

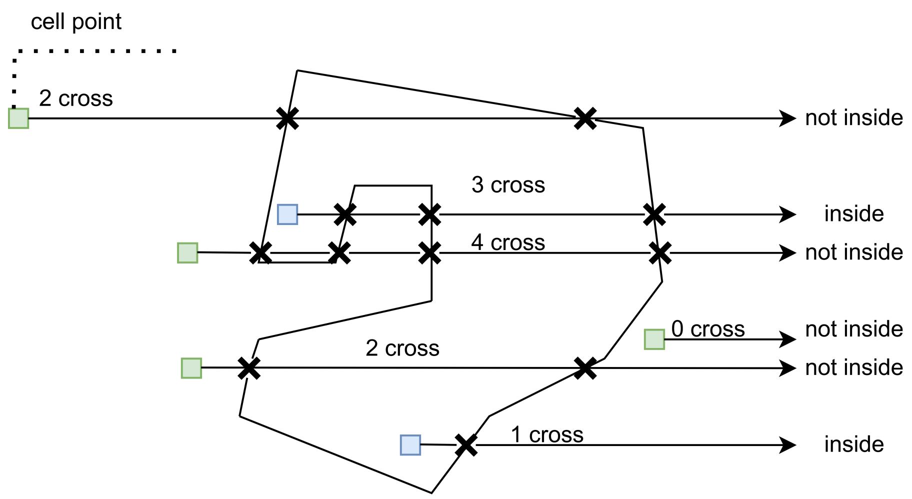

Figure 4.3: Ray Casting Algorithm.

to be a nearly square-shaped test lake, while the other one is Bafa Lake, which has a rectangular shape.

However, the area of the lake may change depending on the season. Additionally, the depth of the shoreline may not be sufficient for the USV navigation, or there may be reeds present, blocking the boat's movement. To prevent such situations, a parameter is defined for the simulation that represents recession on the water in meters, as illustrated in Figure 4.4. With this information, the water receding distance parameter for the test lake and Bafa Lake is selected as 100 and 150 meters, respectively. Moreover, according to their shapes, the horizontal and vertical block numbers for the test lake are selected as 5×5, and the horizontal and vertical block numbers for Bafa Lake are selected as 9×4. Each block has 20×20 cells, and both lakes are illustrated for the simulations in Figures 4.1 and 4.2. In both simulation setups, the lake is surrounded by a group of boundary lines. These lines are represented as a set of boundary points at a specific distance of 70 meters, generating repulsive force to the USV.

Since our approach is developed for monitoring multiple water quality parameters, two water quality parameters are selected for the simulations: Dissolved oxygen (DO) and potential of hydrogen (pH), which are commonly used for water quality monitoring. For Algorithm 3.4, the weights of DO and pH are set to 0.6 and 0.4, respectively.

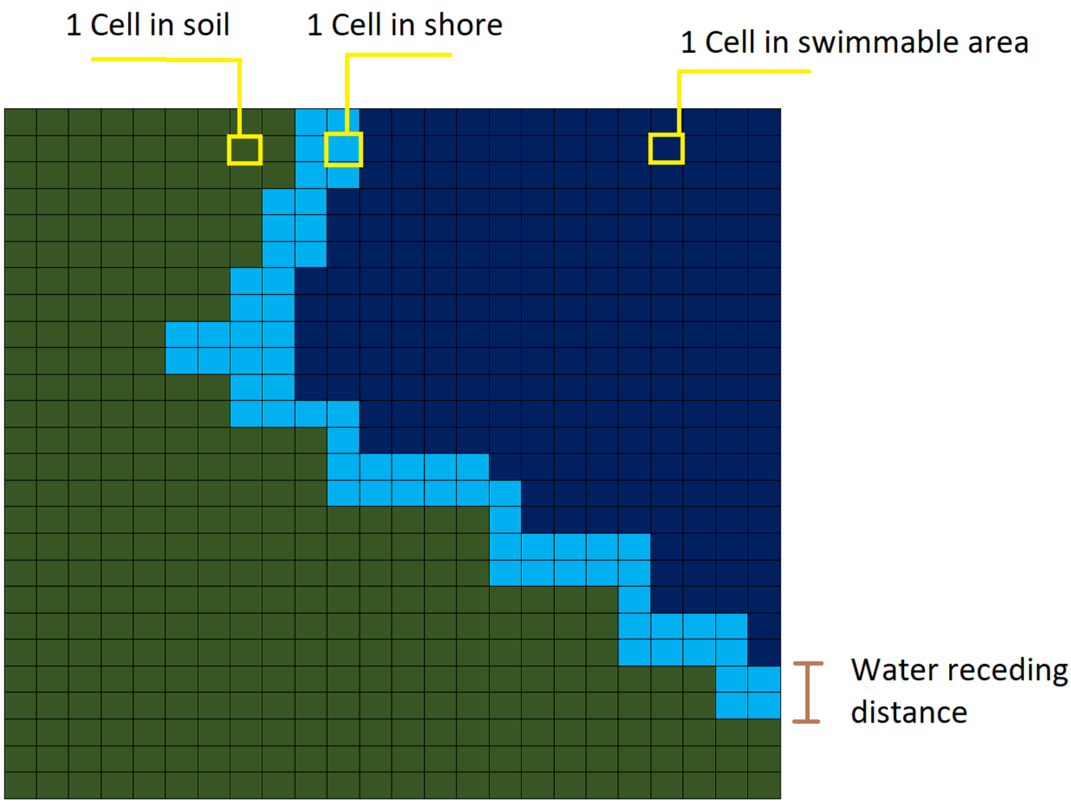

Figure 4.4: Lake water receding.

For the simulations, we need synthetic water quality values. To handle this demand, the inverse distance weighting interpolation method is used. In the moveable locations in the terrain, point locations of specific distances are selected, and values in the range of pH (0-14) and DO (0-16 mg/L) [Uddin et al., 2021] are selected. With these values, the inverse distance interpolation method is used for each block and cells in these blocks and it is described in Equation 4.1.

$$w(x) = \frac{\sum_{i=1}^{N_s} u(x_i) * \phi(d_i)}{\sum_{i=1}^{N_s} \phi(d_i)} \tag{4.1}$$

where N_s is the number of point values, $u(x_i)$ is the value of point x_i, and $\phi(d_i)$ is the weight as minus the square of the Euclidean distance between point x_i and the cell. In this way, for the test lake, two different synthetic water quality maps (WQM-T1 and WQM-T2) are generated for DO and pH, as shown in Figures 4.5 and 4.6.

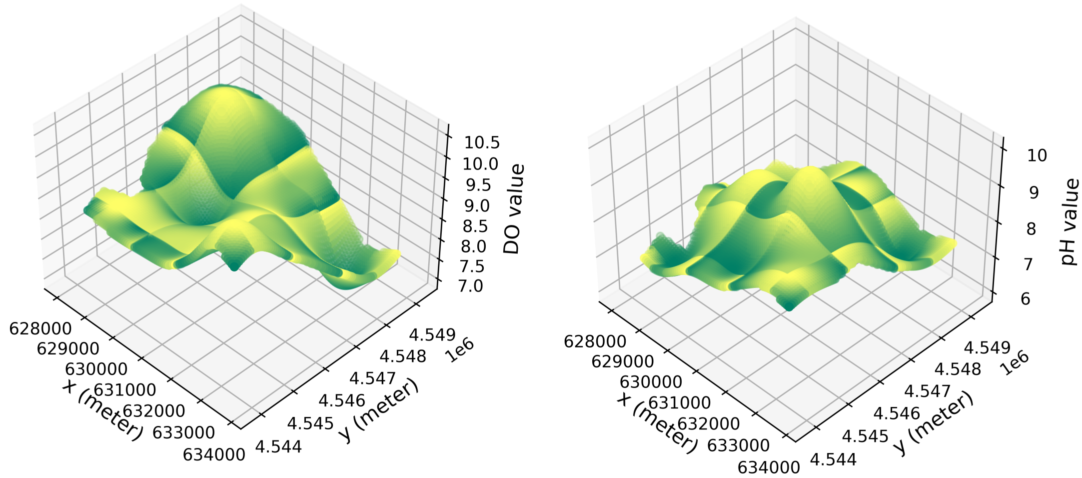

Figure 4.5: DO and pH synthetic water quality map WQM-T1 for the test lake.

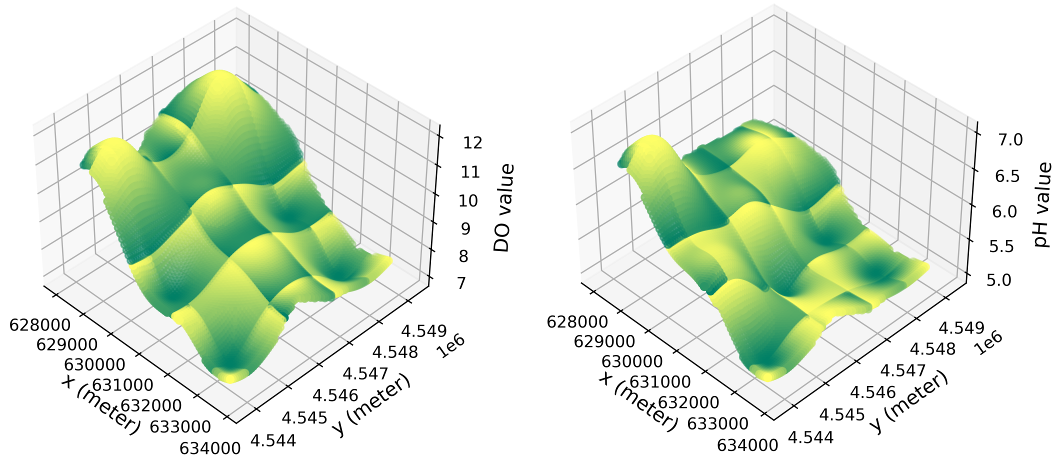

Figure 4.6: DO and pH synthetic water quality map WQM-T2 for the test lake.

Moreover, one synthetic water quality map (WQM-B) for DO and pH is generated using this method for Bafa Lake and it is shown in Figure 4.7.

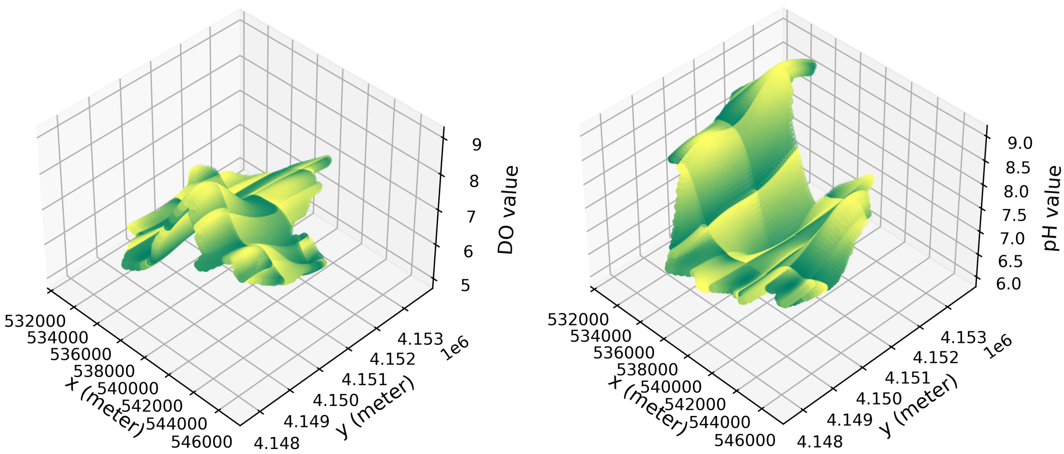

Figure 4.7: DO and pH synthetic water quality map WQM-B for Bafa Lake.

From this point of view, in this thesis, GPR is used as a regression model. To implement this model, the Scikit-learn library [Pedregosa et al., 2011] in Python is used. The regressor requires the ‘kernel‘ and ‘n restarts optimizer‘ parameters. ‘n restarts optimizer‘ is used to restart the optimizer when the regressor gets stuck in a local optimum. Different variations of the ‘kernel‘ parameter are explored. Radial basis function (RBF) and rational quadratic (RQ) kernels are used, as well as a constant kernel and a white kernel. The four kernels are summed to account for white noise in the measurements.For the test lake simulations, the characteristic length scale l for the RBF and RQ kernels is set to 65. This corresponds to keeping the values of the kernel nearly the same within a distance of 65 units, which is also the distance between neighboring cells. For the RQ kernel, the scale mixture parameter α is set to 3. α determines the relative importance of large-scale and small-scale variations [Duvenaud, 2014]. Finally, multiplying a kernel with a constant kernel scales the amplitude of the kernel. This effect can be observed by using the constant kernel with a constant parameter of 10. Additionally, the kernels are summed with a white kernel with a variance of 0.003.

Since we use four different kernels in Gaussian process regression, their parameters are described as follows:

- RBF($l = 65$) + White($\sigma^2 = 0.003$)
- C(10)*RBF($l = 65$) + White($\sigma^2 = 0.003$)

- RQ($l = 65$, $\alpha = 3$) + White($\sigma^2 = 0.003$)
- C(10)*RQ($l = 65$, $\alpha = 3$) + White($\sigma^2 = 0.003$)

Moreover, the parameters of the four different kernels are described for Bafa Lake as follows:

- RBF($l = 120$) + White($\sigma^2 = 0.003$)
- C(10)*RBF($l = 120$) + White($\sigma^2 = 0.003$)
- RQ($l = 120$, $\alpha = 3$) + White($\sigma^2 = 0.003$)
- C(10)*RQ($l = 120$, $\alpha = 3$) + White($\sigma^2 = 0.003$)

In the simulations, the maximum speed of the USV is set to four meters per second. The maximum traveling distance is assumed to be 25,000 meters due to the limited energy source. Additionally, each iteration in the simulation δt takes 25 seconds due to the long computation time. Gaussian noise with a standard deviation of 0.05 is added when the robot takes a measurement from the synthetic water quality map.

As previously defined in Sections 3.1.3 and 3.1.4, our approach has two different missions: local and global. In the global mission, a new potential function method is used for movement among blocks. α_{att}, $\alpha_{rep,block}$, and $\alpha_{rep,border}$ are constants for the attractive, repulsive block, and repulsive border potentials, respectively. For the test lake, the constant values are selected as 3×10^9, 6×10^5, and 4×10^6, respectively. For Bafa Lake, the constant values are selected as 6×10^9, 1×10^6, and 8×10^6, respectively. Also, α_{dist} is the distance constant, which is used for determining how far away the border points affect the robot, and this constant is selected as 300 meters for both lakes.

On the other hand, $\alpha_{att,local}$ and $\alpha_{rep,local}$ are the attractive and repulsive potential constants during the local mission, and they are chosen as 2×10^{-2} and 5×10^7 for the test lake. Also, for Bafa Lake, they are chosen as 2×10^{-2} and 5×10^8. In the local mission, the robot aims to move to the maximum uncertainty location. During this travel, the threshold closeness parameter, loc_dist_thr, is applied at a value of 150 and 250 meters for the test lake and Bafa Lake, respectively.

When the USV changes its operation mode from local to global, the maximum equivalent variance value in the block is compared with the threshold variance value, loc_var_thr. For the test lake simulation, this value is chosen to be 0.05. Also, to stop the algorithm, the sum of the variance of each block is calculated and compared with the global threshold value, glb_var_thr, which is selected to be 0.5 for the test lake and Bafa Lake.

4.2 Baseline Approach

As to the baseline approach, the same method of local mission is used for the global mission. When the USV changes its mission from local to global, it calculates the average variance of each block. The block with the highest average variance is the goal block. During the robot's travel to the center of gravity of the goal block, it uses the local potential function approach. The center of the goal block makes an attractive potential to the USV, and the border points of the goal block make a repulsive potential to the USV.

To evaluate the performance of our proposed approach, we compare it with the baseline approach in Section 4.3.

4.3 Simulation Results and Comparison

In this section, our approach and the baseline approach, kernel types used for Gaussian process regression (GPR), and characteristic features of GPR are compared and analyzed in two different simulation water bodies (lakes).

4.3.1 Test Lake

In the test lake, we perform simulations using two distinct synthetic water quality maps (WQM-T1 and WQM-T2). We use these maps, along with the same starting point (SP-1) and four Gaussian process regression kernels (RBF+White, C*RBF+White, RQ+White, and C*RQ+White), to assess the performance of our approach. We run each setting five times to obtain statistically significant results.

To analyze the effect of the number of repetitions for the same parameters, the simulation for our proposed artificial potential function approach is run five times in the same synthetic water quality map (WQM-T1), with the C*RQ+White kernel (with the parameters C=10, l=65, α=3, σ^2=0.003) and the starting point SP-1. The results are shown in Table 4.1. Table 4.1 has the following columns: The number of repetitions, the traveled distance, the number of measurements, the mean, and standard deviation of the absolute difference between the synthetic and predicted DO and pH water quality values (mean-DO, std. dev.-DO, mean-pH, std. dev.-pH). Because the Gaussian process is a stochastic process, the predicted DO and pH, and these variance values may change for each cell. This situation affects the determination of the next measurement locations, therefore the number of measurements, traveled distance, and the absolute difference between synthetic and predicted water quality parameters.

Table 4.1: Simulation results in which our approach is used with the settings WQM-T1, C*RQ+White and SP-1

Repetition ID	# of Meas.	Traveled Dist.	DO		pH	
			Mean	Std.	Mean	Std.
1	59	51725	0.0444	0.0375	0.0980	0.1171
2	59	55214	0.0620	0.0595	0.0767	0.1200
3	50	55876	0.0803	0.0906	0.0800	0.0842
4	58	67781	0.0518	0.0688	0.0474	0.0353
5	55	58761	0.0507	0.0435	0.0592	0.0557

The traveled path is shown in Figure 4.8 for synthetic water quality map (WQM-1), using the C*RQ+White kernel (with the parameters $C = 10$, $l = 65$, $\alpha = 3$, $\sigma^2 = 0.003$), starting point SP-1, and repetition ID-5 of the simulation.

To analyze the effect of kernel choice, the simulations are conducted with two different water quality maps (WQM-T1 and WQM-T2). The results of four different kernel simulations, including the traveled distance, the number of measurements, the mean, and the standard deviation of the absolute difference between synthetic and predicted DO and pH water quality values, are averaged over five repetitions and detailed in Tables 4.2 and 4.3. For each setting, the simulation is repeated five times, and the results are average values for each kernel. The RBF+White kernel could not finish

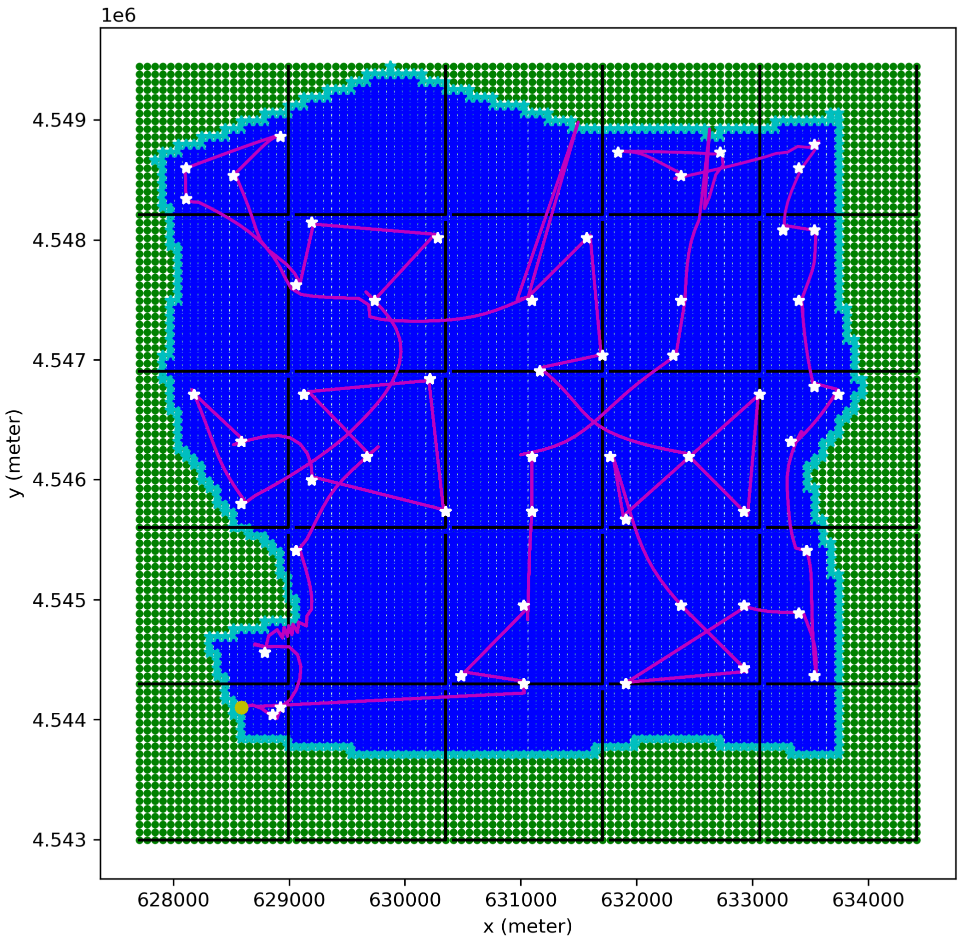

Figure 4.8: A sample result from the simulations showing the USV trajectory and measurement locations at the test lake when our approach is used (the starting point is set to SP-1, the water quality map is set to WQM-T1). The yellow dot represents the starting point, magenta lines represent the robot's trajectory, and white stars represent measurement locations.

the process during the simulations. After it scans all of the blocks, the total average variance value of each block is greater than glb_var_thr. Also it overfits (fitting the error). Therefore, the results are not written in Table 4.2.

Table 4.2: The effects of the various kernels using our approach with the settings WQM-T1 and SP-1

Kernel	# of Meas.	Traveled Dist.	DO		pH	
			Mean	Std.	Mean	Std.
RBF+White	failed	failed	failed	failed	failed	failed
C*RBF+White	71	67139	0.0773	0.0756	0.1120	0.1091
RQ+White	62	54162	0.08148	0.0960	0.1154	0.1659
C*RQ+White	56	57871	0.0578	0.0600	0.07231	0.0825

When we examine Table 4.2, we observe that the RBF+White kernel cannot map

WQP-T1 with the selected hyperparameters. This kernel fits the error and is therefore overfitted. This situation also occurs when the water quality map is set to WQP-T2. Additionally, when comparing kernels, the RQ+White kernel performs better than the RBF+White kernel in terms of traveled distance and water quality mapping accuracy. Multiplying a kernel with a constant kernel increases the number of traveled distance, but it also gives more accurate values.

To analyze the effects of the kernels using a different water quality map for DO and pH, the simulations are conducted with the starting point SP-1 and the kernel functions of the synthetic water quality map WQM-T2. The results of our proposed approach is given in Table 4.3.

Table 4.3: The effects of the various kernels using our approach with the settings WQM-T2 and SP-1

Kernel	# of Meas.	Traveled Dist.	DO		pH	
			Mean	Std.	Mean	Std.
RBF+White	failed	failed	failed	failed	failed	failed
C*RBF+White	69	69955	0.0932	0.1020	0.08798	0.07074
RQ+White	63	55877	0.07641	0.0778	0.04418	0.04129
C*RQ+White	57	61276	0.0844	0.0980	0.04939	0.04805

4.3.2 Bafa Lake

The simulation is also conducted for a different water body, namely Bafa Lake, to verify the chosen kernel (C*RQ+White) and compare our proposed global mission approach with the baseline approach. One starting point and one synthetic water quality map for DO and pH are generated and selected for the comparison. The simulations are repeated five times, and the results are shown in Tables 4.4 and 4.5.

Table 4.4: Simulation results for **the proposed approach** for Bafa Lake with the settings WQM-B and SP-1

Repetition ID	# of Meas.	Traveled Dist.	DO		pH	
			Mean	Std.	Mean	Std.
1	52	81189	0.0686	0.0870	0.05781	0.05719
2	52	72427	0.0638	0.0778	0.0610	0.0780
3	48	77608	0.0713	0.1056	0.0496	0.0459
4	57	73423	0.04831	0.0409	0.0641	0.0597
5	43	71606	0.0827	0.0917	0.0673	0.0688
Average	50	75251	0.06699	0.0806	0.0599	0.06196

Table 4.5: Simulation results for **the baseline approach** for Bafa Lake with the settings WQM-B and SP-1

Repetition ID	# of Meas.	Traveled Dist.	DO		pH	
			Mean	Std.	Mean	Std.
1	31	69610	0.07081	0.0920	0.0721	0.0714
2	26	63909	0.0954	0.1095	0.1146	0.0896
3	36	82490	0.0732	0.0981	0.0660	0.0660
4	40	84317	0.0771	0.0894	0.0643	0.0531
5	41	85405	0.0717	0.0855	0.0569	0.0495
Average	35	77146	0.07915	0.0949	0.0748	0.0659

When we compare the results in Tables 4.4 and 4.5, we can observe that the traveled distance of our proposed algorithm is less than that of the baseline approach. On the other hand, the number of measurements for our proposed approach is higher than that of the baseline approach relative to the traveled distance. Additionally, our approach has a lower average difference between synthetic and predicted DO and pH values.

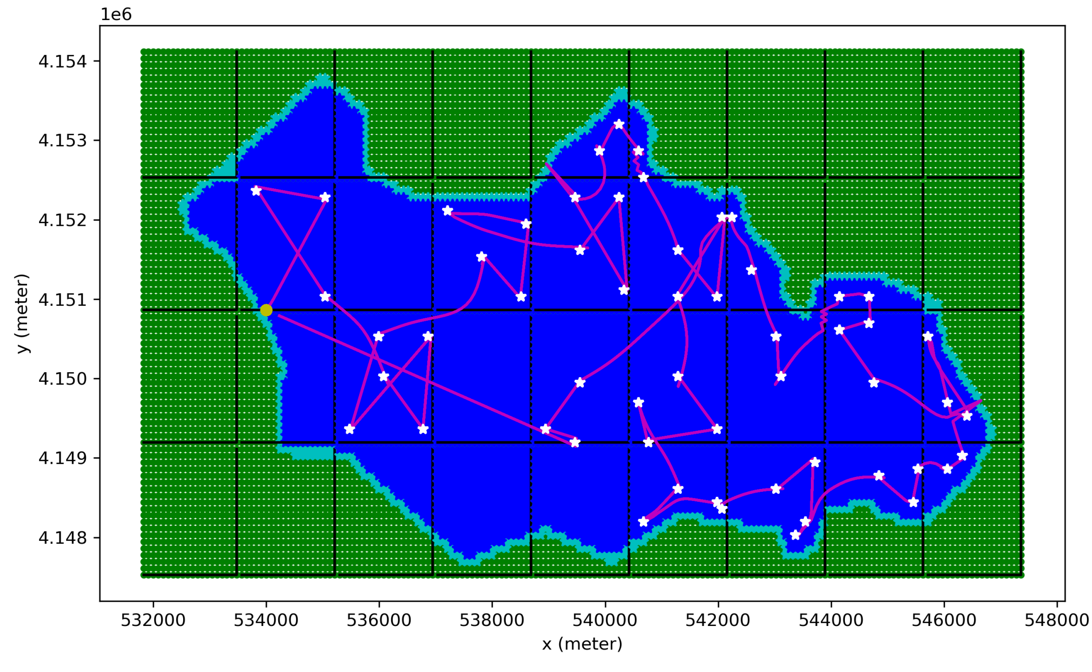

Figure 4.9: Our Approach at Bafa Lake: A sample result from the simulations showing the USV trajectory and measurement locations at Bafa Lake when our approach is used (the starting point is set to SP-1, the water quality map is set to WQM-B). The yellow dot represents the starting point, magenta lines represent the robot's trajectory, and white stars represent measurement locations.

Figure 4.9 describes the path of the robot traveled for the fourth repetition of our approach. Figure 4.11 illustrates the decrease in the mean and standard deviation of the absolute difference between the synthetic and predicted DO and pH water quality values for each cell during the block transition operation as an error plot.

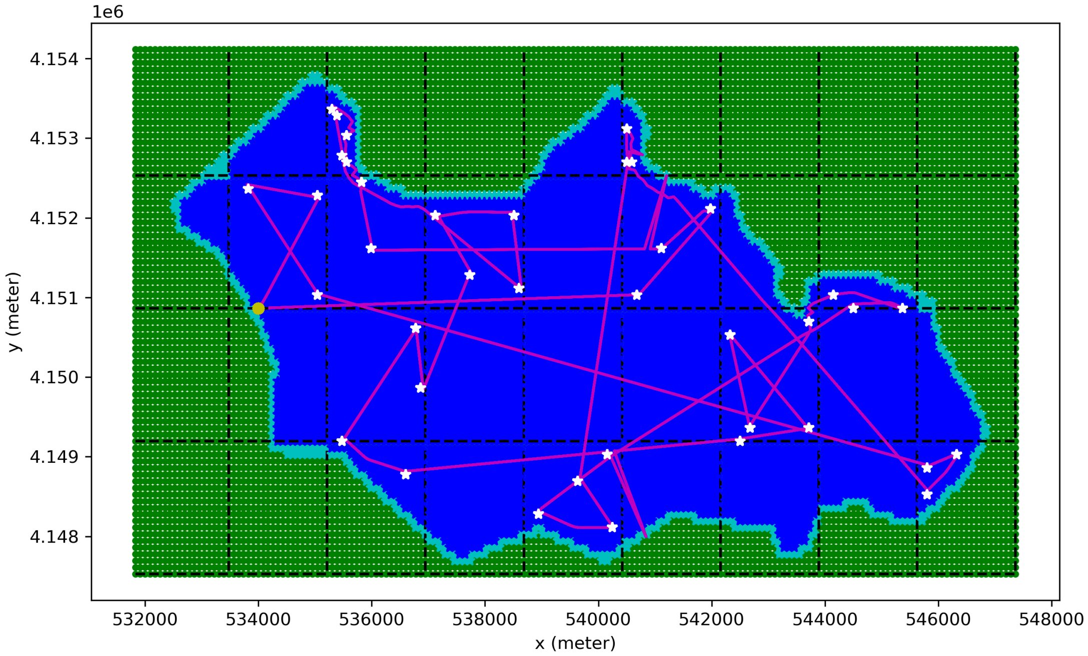

Figure 4.10: Baseline Approach at Bafa Lake: A sample result from the simulations showing the USV trajectory and measurement locations at Bafa Lake when the baseline approach is used (the starting point is set to SP-1, the water quality map is set to WQM-B).

On the other hand, for the baseline approach with the same simulation settings, describes the path of the robot traveled for the first repetition of our approach. Figure 4.12 illustrates the decrease in the mean and standard deviation of the absolute difference between synthetic and predicted DO and pH water quality values for each cell during the block transition operation as an error plot. According to completion time of the monitoring task, our approach is faster than the baseline approach.

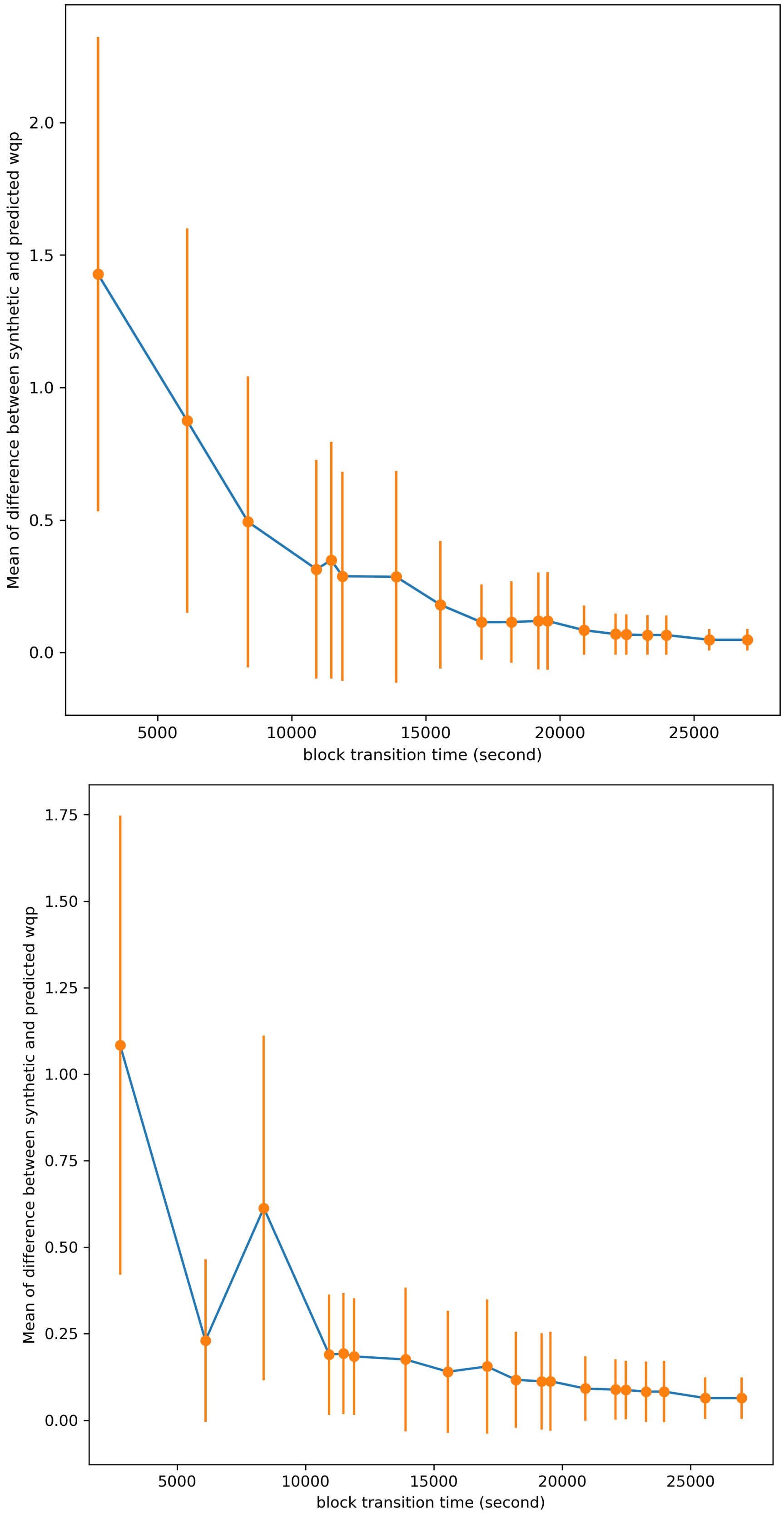

Figure 4.11: Averaged absolute difference between synthetic and predicted DO (upper plot) and pH (lower plot) values during block transition time for our proposed artificial potential function approach.

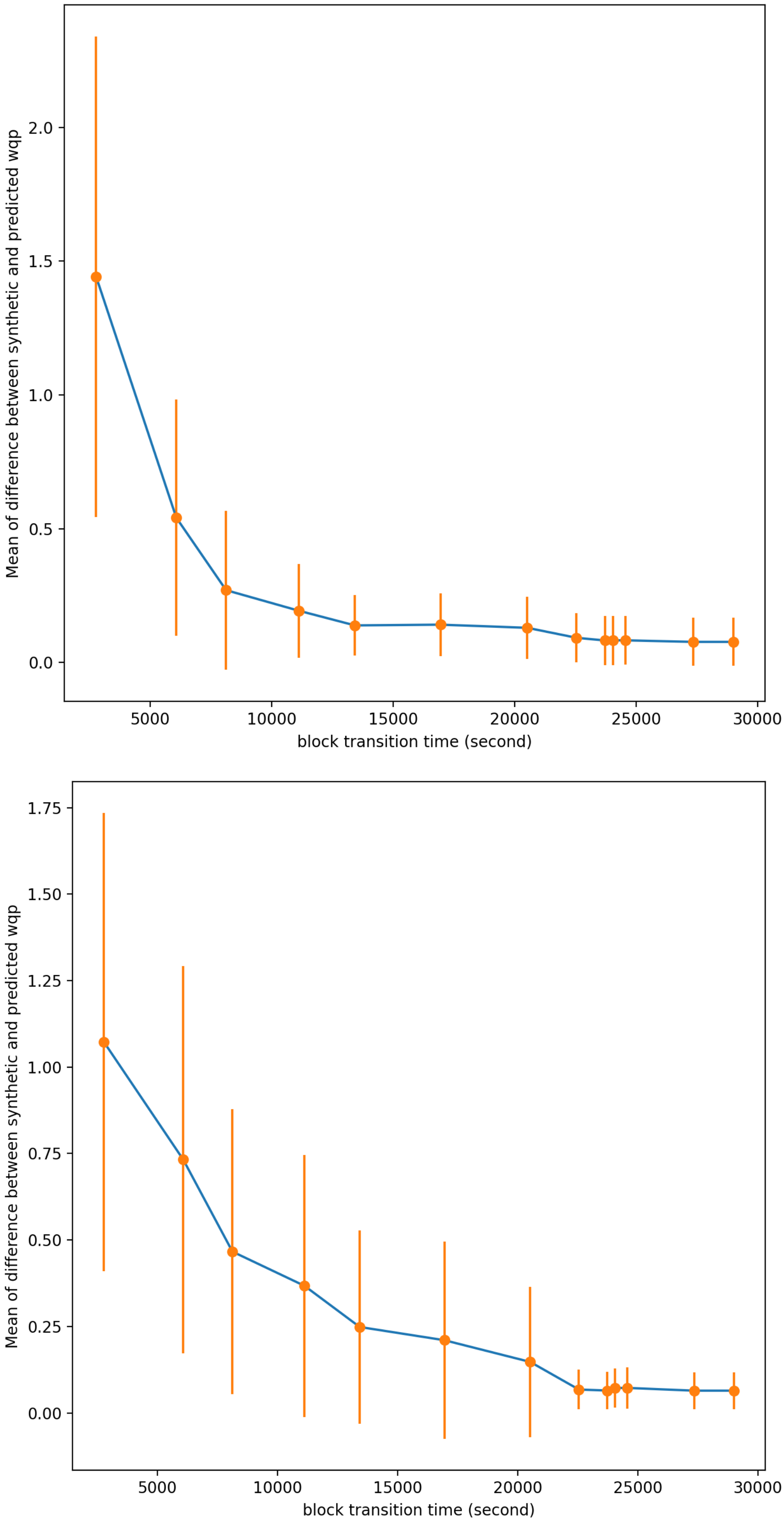

Figure 4.12: Averaged absolute difference between synthetic and predicted DO (upper plot) and pH (lower plot) values during block transition time for baseline approach.

CHAPTER 5: CONCLUSION

In this thesis, we proposed a novel approach to artificial potential function (APF)-based water quality monitoring. To account for wide areas, the whole terrain is discretized into blocks, and each block is discretized into cells. Our proposed APF approach focuses on block transition operations as the global approach. This approach categorizes blocks as measured and unmeasured. Visited blocks and the borders of the terrain affect the robot with a repulsive potential, while unvisited blocks affect the robot with an attractive potential.

On the other hand, our proposed approach is also compared with a baseline approach. The baseline approach aims to move the robot to the target location block with the maximum variance from the averaged variance of each block. After the simulations, our proposed APF approach traveled less distance than the baseline approach. In our approach, the robot travels to the nearest unmeasured block and applies the local approach to learn water quality parameters. On the other hand, the baseline approach targets the robot to the farthest part of the unmeasured blocks.

Moreover, our proposed approach gives more accurate results than the baseline approach, if we compare the mean of each cell's absolute difference between synthetic and predicted water quality values for DO and pH.

We also compared the kernels used for Gaussian process regression. The RQ kernel gives more accurate results than the RBF kernel. Additionally, when we multiply a kernel with the constant (C) kernel, the variance output of the regression is increased, so the simulation needs to travel more and take more measurements to stop the algorithm.

For future work, the developed approach can be extended to analyze water quality values in the temporal domain. Additionally, multiple USVs can be used for mapping and patrolling, and heterogeneous systems that combine USVs and UAVs can be used to monitor both on and above water.

REFERENCES

[Ajanović et al., 2018] Ajanović, Z., Stolz, M., and Horn, M. (2018). A novel model-based heuristic for energy-optimal motion planning for automated driving. IFAC-PapersOnLine, 51(9):255–260.

[Choset et al., 2005] Choset, H., Lynch, K. M., Hutchinson, S., Kantor, G. A., and Burgard, W. (2005). Principles of robot motion: theory, algorithms, and implementations. MIT press.

[Duvenaud, 2014] Duvenaud, D. (2014). Automatic model construction with Gaussian processes. PhD thesis, University of Cambridge.

[Fazai et al., 2019] Fazai, R., Mansouri, M., Abodayeh, K., Puig, V., Selmi, M., Nounou, H., and Nounou, M. (2019). Multiscale gaussian process regression-based glrt for water quality monitoring. In 2019 4th Conference on Control and Fault Tolerant Systems (SysTol), pages 44–49. IEEE.

[Kathen et al., 2021] Kathen, M. J. T., Flores, I. J., and Reina, D. G. (2021). An informative path planner for a swarm of asvs based on an enhanced pso with gaussian surrogate model components intended for water monitoring applications. Electronics, 10(13):1605.

[Lal and Datta, 2018] Lal, A. and Datta, B. (2018). Genetic programming and gaussian process regression models for groundwater salinity prediction: machine learning for sustainable water resources management. In 2018 IEEE conference on technologies for sustainability (SusTech), pages 1–7. IEEE.

[Li, 2018] Li, T. (2018). Robotic path planning for environmental field estimation and its application in aquatic monitoring. PhD thesis, University of British Columbia.

[Ma et al., 2017] Ma, K.-C., Liu, L., and Sukhatme, G. S. (2017). Informative planning and online learning with sparse gaussian processes. In 2017 IEEE International Conference on Robotics and Automation (ICRA), pages 4292–4298. IEEE.

[Marchant and Ramos, 2014] Marchant, R. and Ramos, F. (2014). Bayesian optimisation for informative continuous path planning. In 2014 IEEE International Conference on Robotics and Automation (ICRA), pages 6136–6143. IEEE.

[Mishra et al., 2018] Mishra, R., Chitre, M., and Swarup, S. (2018). Online informative path planning using sparse gaussian processes. In 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO), pages 1–5. IEEE.

[Pamosoaji and Hong, 2013] Pamosoaji, A. K. and Hong, K.-S. (2013). A path-planning algorithm using vector potential functions in triangular regions. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 43(4):832–842.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830.

[Peralta et al., 2021] Peralta, F., Reina, D. G., Toral, S., Arzamendia, M., and Gregor, D. (2021). A bayesian optimization approach for multi-function estimation for environmental monitoring using an autonomous surface vehicle: Ypacarai lake case study. Electronics, 10(8):963.

[Plonski et al., 2013] Plonski, P. A., Tokekar, P., and Isler, V. (2013). Energy-efficient path planning for solar-powered mobile robots. Journal of Field Robotics, 30(4):583–601.

[Sankar and Norman, 2009] Sankar, P. and Norman, S. R. (2009). Embedded system for monitoring atmospheric weather conditions using weather balloon. In 2009 International Conference on Control, Automation, Communication and Energy Conservation, pages 1–4. IEEE.

[Şen, 2009] Şen, Z. (2009). Global warming threat on water resources and environment: a review. Environmental geology, 57:321–329.

[Stachniss et al., 2009] Stachniss, C., Plagemann, C., and Lilienthal, A. J. (2009). Learning gas distribution models using sparse gaussian process mixtures. Autonomous Robots, 26(2):187–202.

[Tan et al., 2018] Tan, Y. T., Kunapareddy, A., and Kobilarov, M. (2018). Gaussian process adaptive sampling using the cross-entropy method for environmental sensing and monitoring. In 2018 IEEE International Conference on Robotics and Automation (ICRA), pages 6220–6227. IEEE.

[Tebbutt, 1997] Tebbutt, T. H. Y. (1997). Principles of water quality control. Elsevier.

[Uddin et al., 2021] Uddin, M. G., Nash, S., and Olbert, A. I. (2021). A review of water quality index models and their use for assessing surface water quality. Ecological Indicators, 122:107218.

[Williams and Rasmussen, 2006] Williams, C. K. and Rasmussen, C. E. (2006). Gaussian processes for machine learning, volume 2. MIT press Cambridge, MA.

[Yates and Goodman, 2014] Yates, R. D. and Goodman, D. J. (2014). Probability and stochastic processes: a friendly introduction for electrical and computer engineers. John Wiley & Sons.

[Ye et al., 2013] Ye, Y., Guangrui, F., and Shiqi, O. (2013). An algorithm for judging points inside or outside a polygon. In 2013 Seventh International Conference on Image and Graphics, pages 690–693. IEEE.

[Zare Farjoudi and Alizadeh, 2021] Zare Farjoudi, S. and Alizadeh, Z. (2021). A comparative study of total dissolved solids in water estimation models using gaussian process regression with different kernel functions. Environmental Earth Sciences, 80(17):557.