Spectral Unmixing and Mapping of Coral Reef Benthic Cover

Rohan Zeng

CMU-RI-TR-23-06

May 9, 2023

The Robotics Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA

Thesis Committee:
David Wettergreen, *chair*
Maxim Likhachev
Dhruv Saxena

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Robotics.

To my family, friends, and mentors that have advised me, supported me, and helped me grow all throughout this journey.

Abstract

Coral reefs are important to the global ecosystem and the local communities and wildlife that rely on the habitat they create. However, coral reefs are also in critical and rapid decline: reefs have degraded over recent decades and what remains is at increasing risk of loss. Only a small fraction of the world's reefs have been studied quantitatively. Mapping these reefs, in particular the benthic cover components of coral, algae, and sand, would allow us to monitor changes and understand their health. We seek methods to better model reefs at global scales. Throughout this work, we train models on different levels of spatial-resolution satellite data and evaluate performance between them. We also develop on a novel method using a Deep Conditional Dirichlet Model to perform unmixing of spectral signatures in different satellite data sources. Accounting for mixing of classes yields improvements in accuracy of the predicted class probabilities. Regions of high uncertainty are also determined based on the model output prediction and can be used to determine informative *in situ* sampling. An ergodic planner is implemented to generate a path through these regions to acquire samples that best improve the coral map. The result of this study is an efficient learning based pipeline that augments existing satellite data and maps coral reefs globally in order to improve our understanding of their condition.

Acknowledgments

There are a plethora of people that I have to thank for helping to make this thesis possible.

First and foremost, I want to express my gratitude to my advisor, David Wettergreen. Thank you for taking a chance and giving me the opportunity to pursue research as an undergraduate senior, and even more for continuing to advise me over these years. I'm truly grateful for your steady guidance, kindness, and support through all of the highs and lows that research brings. You are a mentor and pillar that I look up to, and look forward to working with for years to come.

Thank you to Maxim Likhachev and Dhruv Saxena for taking the time to be on my committee. I really appreciate all of your advice informing and influencing the direction of my work over this past year. Your questions, thoughts, and insights gave me pause, helped me more deeply consider the foundations of my research, and overall strengthened the final output as a result.

I'm fortunate to have had the opportunity to work with both Alberto Candela Garza and Eric Hochberg over these past two years as well. Their own work and goals are what brought this research to life and made it possible for me to pursue this work in the first place. Thank you for sharing your experiences and knowledge in order to help me better understand and appreciate the depth of this field.

To Erin Wong and Ananya Rao, two very close friends that came into this program with me from undergrad at CMU. I'm so grateful to have had the opportunity to share this journey with you both. From late night discussions to projects together, sharing struggles and everything else in between, truly, thank you. We did it!

I'd also to thank my lab mates David Russell, Maggie Hansen, Abby Breitfeld, and Srini Vijayarangan. You all made the office so much fun to be in and always inspire me with your own work and passions. Thank you for helping make the lab an incredible community to be a part of, for sharing your passions with me, and apologies for always hopping around and usurping your desks to work.

From the friends I've made during my time at CMU, thank you especially to my housemates Hardik (Dikky) Singh, Arian Raje, Suraj Bhosale,

Tejpal Virdi, Hans Kumar, and Haran Balakrishnan Selvakumaran. I cannot express enough how much each of you means to me. As friends, as mentors, as family, you all have done so much for me and have been my backbone of support throughout this entire journey. There's no other group I would have rather had to make these memories with. I love you all.

The CMU Bhangra Team has been an environment that has fostered my growth ever since I was a freshman. Through the team, I've made countless friends, connections, and memories that will last a lifetime. You all made life outside of school so much fun, and I'm so fortunate to have continued to share so many of these experiences with everyone even in grad school. Thank you all for being people that I look up to, friends that I know I can always reach out to, laugh, and have fun alongside, for being so supportive of me through it all and helping me grow to my best self.

My RI cohort, my friends from undergrad, and so many more are all deserving of thanks as well. I'm grateful for all of the memories we've made together, whether it was a passing by work session, sharing food together, catching up through Zoom, or more. You all made my time at CMU so enjoyable, and I hope to keep in touch.

Last, but absolutely not least, to my family, and especially my parents. Thank you. For raising me, for taking care of me, for supporting my dreams, for loving me. I could not be where I am today without you and your endless support. I love you.

Funding

This research was supported by funding from the National Aeronautics and Space Administration (NASA) through awards NNX16AB05G and 80NSSC21K1159. Digital- Globe/Maxar data were provided by NASA's Commercial Archive Data for NASA investigators (cad4nasa.gsfc.nasa.gov) under the National Geospatial-Intelligence Agency's NextView license agreement.

Contents

When this thesis is viewed as a PDF, the page header is a link to this Table of Contents.

List of Figures

List of Tables

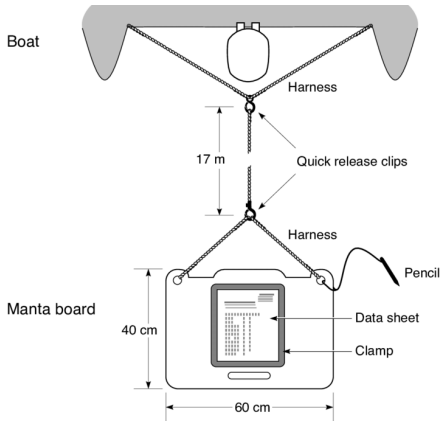
Chapter 1

Introduction

1.1 Problem Motivation

Coral reefs are a necessary ecosystem for the wildlife that rely on them while also being an essential component to the cultural and economic lives of people living near them [Moberg and Folke 1999; Reaka 1997]. Reefs worldwide are in critical and rapid decline, with 33-55% of reefs degrading over recent decades and about 35% of the remaining reefs at risk of loss over the next few decades [ISRS December 2015; Wilkinson 2008]. This decline has been fueled by a combination of two primary factors. The first is a higher sea surface temperature that stresses and causes coral bleaching which reduces coral reproduction and immunity [Wilkinson 2008]. The second is increasing ocean acidity through higher CO_2 concentrations in the atmosphere that acidifies and thus decreases carbonate ion concentration in the ocean [Hoegh-Guldberg et al. 2017; Wilkinson 2008]. The lack of this ion reduces the capacity of coral to generate calcium carbonate skeletons and continue growing. Despite these rising issues, only 0.01-0.1% of the world's reefs have been studied quantitatively [Hochberg and Gierach 2021]. With current global warming trends affecting the health of the coral reefs, it is important that we are able to map these coral regions in order to monitor their health over time.

(a) Diver conducting a survey



(b) Sample setup of the Manta board and attachments

Figure 1.1: Manta tow technique for crown-of-thorns starfish and coral surveys [Miller et al. 2018]. A diver is pulled along by a boat while recording observations on a Manta board for surveying.

1.2 Problem Approach

Coral reef mapping has typically consisted of *in situ*, or on site, missions. Researchers ventured into the field and took pictures or recorded surveys of the coral scene to compile a set of small scale local observations of $10m^2$ regions at separated sites and intervals (Figure 1.1) [Miller et al. 2018]. The composition of the greater coral scene was then estimated by extrapolating these local scenes to a larger scale. However, this approach is quite a timely and costly expenditure that misses the ecological scale variability as it only studies local regions. It also limits the reef survey area to 100s of km^2, only accounting for approximately 0.1% of the total global coral reef coverage (Figure 1.2) [Edmunds and Bruno 1996]. In addition, it is often difficult to replicate *in situ* missions over time, thus limiting the ability to obtain data that is consistent in scope and temporality [Hedley et al. 2016].

In contrast to these issues presented by *in situ* expeditions, the usage of remote sensing allows for mapping of coral reefs that capture the variability and magnitude of the reef regions when surveying the area. Satellite data is an especially cost effective method of remote sensing for this purpose. Although the detail of the data is diminished in comparison to *in situ* or even remote sensed airborne data, the data

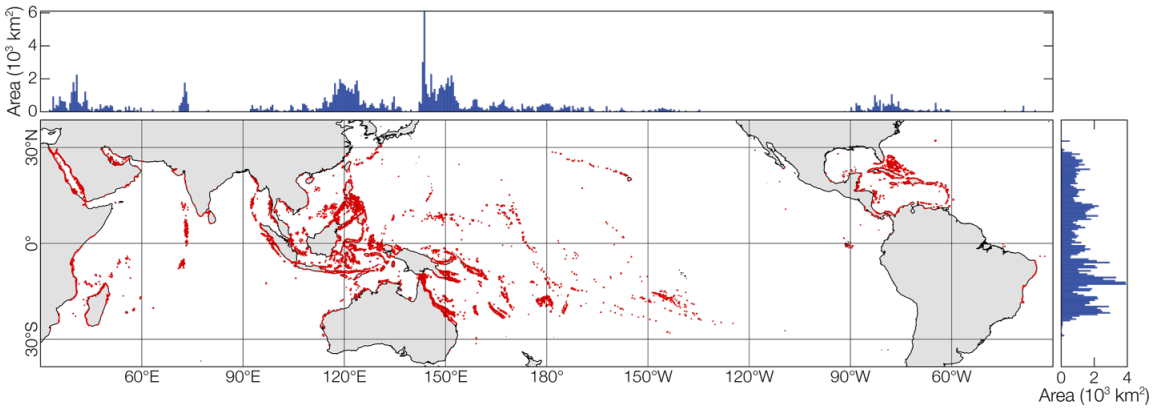


Figure 1.2: Global distribution of coral reefs. The latitudinal and longitudinal extents of coral reefs are shown above and to the right respectively. Reef locations obtained from the United Nations Environment Programme World Conservation Monitoring Centre's Ocean Data Viewer [UNEP-WCMC et al. 2021].

itself costs little to nothing to obtain and thus allows for resources to be diverted towards processing the data instead [Hedley et al. 2016].

With the introduction of remote sensing for coral mapping, maps of greater scale were able to be created of the coral reefs. In particular, in 2001, the Millennium Coral Reef Mapping Project was launched to take advantage of newly available Landsat-7 data to characterize and map the most extensive compendium of coral scenes at the time [Andréfouët et al. 2004]. A standardized scheme was then determined by which the data would be classified for further dissemination towards research purposes across the globe. However, as this classification was done primarily through visual interpretation, it lacks the atmospheric and reflectance processing needed to produce benthic classifications or bathymetric maps [Hedley et al. 2018]. This is significant because the benthic cover of reef ecosystems has proven to be an effective metric for monitoring coral reef health. By characterizing the spectral reflectance of reef benthic communities using remote sensing, different benthic components can be classified and mapped [Hochberg et al. 2003; Hochberg and Atkinson 2000].

We choose to map the three benthic cover classes of coral, algae, and sand (Figure 1.3). However, these three classes are also ambiguous with respect to one another in spectral data. Bleached corals have similar spectral values as sand. Features from their spectral signature are also similar to those covered by algae or live corals due to

(a) Coral (b) Algae (c) Sand/Rock

Figure 1.3: Samples of the benthic cover components to be classified and unmixed [Beeden et al. 2015]

the effect of mixed pixels [Goodman and Ustin 2007; Xu et al. 2021]. This is indicative of the difficulty in being able to properly classify and unmix spectral coral reef scenes with full accuracy. Due to this difficulty in processing their data alongside the fact that the most effective and accurate spectral sensors have been limited to airborne platforms, mapping coral reefs with remote sensing data still remains technically difficult.

We approach this problem through the training of deep neural networks. This not only further decreases the difficulty associated with mapping coral reefs with remote sensing, it also increases the efficiency of the process as well. In machine learning, neural networks are used to classify and cluster data. One way this is achieved is through supervised learning, where input data is provided alongside ground truth labels for training. Models learn to accurately predict the desired output by adjusting their parameters to minimize the loss (difference) between their training predictions and the correct labels. These models can be used to interpret satellite spectral data which, once trained and tuned, will yield a lower cost alternative to traditional processing methods and enable global coral classification.

Recent work employing machine learning approaches have been used with remote sensing for coral reef mapping, resulting in more efficient methods of mapping reefs from spectral data compared to earlier spectral processing algorithms [Dahal et al. 2021; Roelfsema et al. 2020; Li et al. 2020]. One of the most notable compendiums created with these methods is the Allen Coral Atlas, where a Random Forest classifier was used for classification (Figure 1.4) [Lyons et al. 2020]. Further work with neural networks have been trained on the Atlas in order to enable efficient global classification

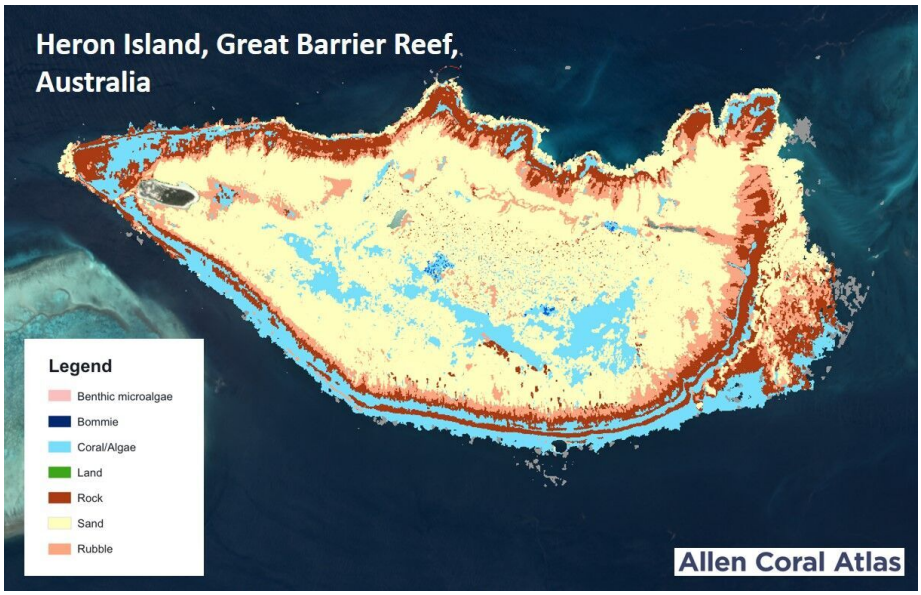
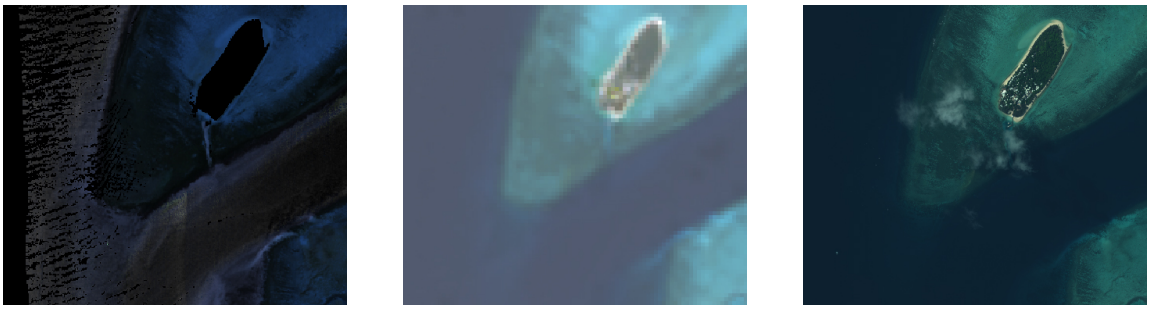


Figure 1.4: Sample classifications by the Allen Coral Atlas on Heron Island [Edkins 2022 (Accessed April 6, 2023)]. Benthic cover classes in the image can be grouped as Coral (Bommie, Coral/Algae), Algae (Benthic Microalgae, Coral/Algae), and Sand (Land, Rock, Sand, Rubble).

and mapping of the coral benthic cover [Li et al. 2020]. However, these efforts have focused on mapping the locations and extents of reefs, not the actual distribution of living coral.

One machine learning study by Candela et al. did focus on mapping the individual distributions of coral, algae, and sand. They utilized Landsat-8 and PRISM data and yielded accurate reconstructions of the benthic scene that closely matched the known ground truth [Candela et al. 2021; Garza 2021]. This study demonstrated the effectiveness of training supervised learning models using labeled hyperspectral, high-spatial-resolution (8 m) data from the Portable Remote Imaging SpectroMeter (PRISM) applied to medium-resolution (30 m), multispectral Landsat-8 data, attaining good predictions despite the difference between label and data sources [Candela et al. 2021]. Here, we follow a similar process to develop models applicable to high-resolution (2 m), multispectral WorldView-3 data (Figure 1.5).

Our models take labeled PRISM data as input, align the labels with the WorldView data, then train Support Vector Machines (SVM), Regression models, and Deep Conditional Dirichlet Models (DCDM) through supervised learning. Regression



(a) PRISM data with spatial resolution 7.8 m

(b) Landsat-8 data with spatial resolution 30 m

(c) WorldView-3 data with spatial resolution 2 m

Figure 1.5: Samples of the Heron Island data products (WorldView-3 data © 2016 Maxar)

models are a type of neural network that regresses model weights in order to better fit the data. The DCDM is a variation of the Regression model that fits a Dirichlet distribution to the Regression model output. Dirichlet distributions model likelihoods over mixtures of components, meaning that they are useful for learning to model mixture distributions and thus unmix these distributions as well. With the higher resolution data, there is an expectation that the models will better learn to identify different classes, distinguish class mixtures, and result in increased spatial accuracy in comparison to the prior lower resolution results.

After the spectral data has been processed, it is important to consider how to best improve the classification and unmixing confidences. One method to approach this problem is to supplement the model predictions with *in situ* expeditions. However, these expeditions still remain to be costly expenditures. By leveraging an approximate knowledge of the map through the model predictions, we can determine a more efficient trajectory for the *in situ* expedition to follow in order to make the best use of the provided resources.

Information priors can be developed from the model output that quanitfy the uncertainty, or conversely the amount of information gain, in the predicted map. Informative path planners utilize these priors to provide a path for an agent to follow. These types of planners prioritize traversing towards regions of high information gain. An ergodic planner further ensures that the trajectory spends time in regions proportional to the expected information gain [Edelson 2020]. The result is an efficient, ergodic trajectory through the coral reef regions. This trajectory optimizes where *in*

situ measurements can be taken to verify or otherwise update the regions where the model is most uncertain.

1.3 Project Goals

There currently does not exist an efficient pipeline for providing data on the global coral reef benthic cover, which is an important component to understanding coral reef condition. We seek to apply previously developed deep learning methods to map and classify these scenes [Candela et al. 2021; Garza 2021]. Though we train the same model frameworks as in prior work, we utilize an updated set of labels alongside higher resolution data with greater global coverage.

In particular, our goal is to evaluate the value of using higher resolution data for training as well as the efficacy of applying novel architectures to unmix spectral data. After classification, we explore how to leverage the uncertainty of the model outputs to inform future *in situ* expeditions for verification and validation of the predictions. Our overall problem statement is that global quantification, and thus understanding, of coral reefs can be efficiently achieved by applying learning methods to mapping coral and further improved with intelligent sampling.

The overall contribution of this thesis is a complete system to classify, unmix, and map global benthic cover as well as determine efficient trajectories for *in situ* expeditions to improve predictions. In particular, the application of the system to the WorldView-3 image archive would yield global estimates of coral reef benthic cover over the past decade. This would allow for significant advancement in our understanding of the trajectory of global coral reef condition.

Chapter 2

Spectral Data and Machine Learning Background

We seek to classify coral reefs using remote sensed data due to the global scale coverage that this data format provides. Remote sensed data can be obtained in different forms depending on the specifications of the device gathering the data. While these data types are distinct, prior methods have demonstrated the effectiveness of using these spectral data formats to train machine learning models for classification tasks as well as unmixing tasks. By processing this data using these models, we can develop an efficient pipeline for classifying global scenes. As such, we choose to train similar machine learning based models on our new data to evaluate performance.

2.1 Spectral Data and Processing

Reflectance is a measure of the electromagetic energy in the light reflected or scattered from the surface. Spectral data consists of measured reflectance values at different wavelengths from the same source. Distinctions in the reflection bands capture a significant amount of information from which the chemistry and composition of the material of which the light is reflected can be derived.

Remote sensed spectral data can be roughly categorized into two categories, hyperspectral and multispectral, and is obtained by measuring the different wavelengths of light from the sun as they are reflected off of the surface of the Earth. Each unique

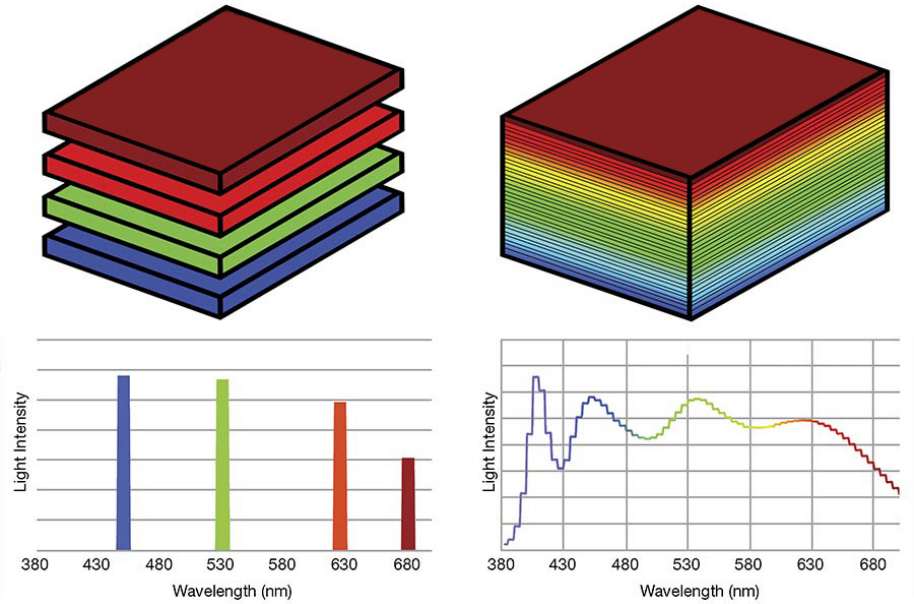


Figure 2.1: Visualization of Multispectral Data (left) vs Hyperspectral Data (right) [Broderson 2020]. Note that band wavelengths are not only limited to the visible spectrum as depicted.

wavelength constitutes a single band of data, where each band can be considered the measurement of the reflectance at the corresponding wavelength over the measured area. Similar to an RGB image, spectral data is thus composed of multiple layers of bands covering a designated area. The spectral resolution refers to the number of bands in the data. The spatial resolution refers to the size of the region measured by each pixel.

The distinction between hyperspectral data and multispectral data lies in the number of bands and the interval between band wavelengths taken. Hyperspectral data has a higher spectral resolution and typically consists of over a hundred bands measured at small intervals. Multispectral data has a lower spectral resolution that consists of much fewer bands, typically less than 20, taken at much larger intervals. Hyperspectral data benefits from being able to capture finer distinctions within its measurements due to its high spectral resolution, but in exchange is more complex than multispectral data due to potential redundancies between its bands [GISGeography 2022 (Accessed March 20, 2023)]. However, both formats of spectral data are suitable for the purpose of coral monitoring and mapping.

2.2 Machine Learning for Mapping

Regression models have been used in previous efforts to classify coral scenes [Dahal et al. 2021; Garza 2021]. Given input and ground truth data alongside a provided loss function to optimize, these models fit functions to the data that minimize the loss. Statistical approaches, e.g. maximum likelihood classification, have been applied extensively for scene-specific mapping. More recently, research has included defining a neural network architecture and regressing model weights with respect to the loss and ground truth over training iterations.

One type of loss function that is applied to train these Regression models is the Kullback-Leibler Divergence (KLD), which is a metric used to quantify the difference between two probability distributions $P(x)$ and $Q(x)$ [Kullback and Leibler 1951]. The loss is defined as follows:

$$D_{KL}(P(x)||Q(x)) = \sum_{x \in X} P(x) log \frac{P(x)}{Q(x)} \tag{2.1}$$

For the purposes of training the Regression model, the KLD loss compares the direct probability values between the output of the model and the labeled values, and the model regresses its weights to minimize the function value.

The KLD also serves as a useful metric for evaluating the model performance post training. A model with perfect performance would have a low KLD value, signifying that the probabilistic difference between the ground truth and the model output is minimized. A similar metric used for model evaluation is the Mean Absolute Error (MAE). Given n pairs of ground truth values y_i and model output values x_i, the MAE is defined as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=0}^{n} |y_i - x_i| \tag{2.2}$$

This value provides an alternative quantification of the model performance through a direct absolute difference between the two values. Similar to the KLD, a lower MAE values signifies better model performance with regards to matching the ground truth labels. A final evaluation metric would be accuracy, which quantifies how well the

model performs standard hard classification. This is defined as follows:

$$\text{Acc} = \frac{1}{n}\sum_{i=0}^{n}[\max(y_i) = \max(x_i)] \tag{2.3}$$

where there are n pairs of ground truth values y_i and model output values x_i.

While accurate classifications are yielded through these approaches, it is also important to note that, as pixels in spectral data encompass regions of varied sizes, these methods do not account for the mixture of classes within the pixels. An addition to the Regression model neural network was thus employed to unmix the benthic cover components from spectral data pixels as a Deep Conditional Dirichlet Model (DCDM) in prior work by Candela [Garza 2021]. This model makes use of a Dirichlet distribution as a probability density function to model likelihoods over mixtures of components together.

The domain of the Dirichlet distribution can be considered as the probabilities associated to K categories $c_1, ..., c_K$. The support of the distribution is the mixing ratios r, where $r \in R \subset \mathbb{R}^K$ with $\Sigma_{k=1}^{K} r_k = 1, r_k \in (0, 1), \forall r \in R$; where R is in the form of an open standard (K-1) simplex. The Dirichlet distribution has a probability density function given as follows:

$$p(r_1, ..., r_K|\alpha_1, ..., \alpha_K) = \frac{\Gamma(\Sigma_{k=1}^{K}\alpha_k)}{\Pi_{k=1}^{K}\Gamma(\alpha_k)}\Pi_{k=1}^{K} r_k^{\alpha_k - 1} \tag{2.4}$$

where Γ is the gamma function. The Dirichlet distribution is defined by the concentration parameters $\alpha_1, ..., \alpha_K > 0$, where larger α_k values indicate that category c_k is more likely and vice versa [Garza 2021].

Given this Dirichlet distribution, the relationship between the mixing ratios r and the input spectral data can be modeled using a probabilistic conditional distribution $p_\theta(r|y)$, where the weights of the neural network are given as θ and $y \in Y \subset \mathbb{R}^n$ are the input spectra. The conditional component of the prior distribution can be thought of as:

$$(r|y) \sim Dir(\alpha_\theta(y)) \tag{2.5}$$

Rather than solely minimizing a KLD loss during training, the DCDM also attempts to maximize a cumulative log-likelihood function of the training set $\mathcal{D} =$

$\{(y^i, r^i)\}_{i=1}^N$ for input spectra y and output ground-truth mixing ratios r. The log-likelihood function is defined as:

$$\mathcal{L} = (\theta|\mathcal{D}) = \sum_{(y,r\in\mathcal{D})} logp_\theta(r|y) \tag{2.6}$$

$$= \sum_{(y,r\in\mathcal{D})} log(\frac{\Gamma(\Sigma_{k=1}^K \alpha_k(y))}{\Pi_{k=1}^K \Gamma(\alpha_k(y))} \prod_{k=1} K r_k^{\alpha_k(y)-1}) \tag{2.7}$$

$$= \sum_{(y,r\in\mathcal{D})} [log\Gamma(\sum_{k=1}^K \alpha_k(y)) - \sum_{k=1}^K log\Gamma(\alpha_k(y)) + \sum_{k=1}^K (\alpha_k(y)-1) log r_k] \tag{2.8}$$

The first two terms in the log-likelihood summation serve as regularizers for α, while the last term measures the difference between the predicted and true mixing ratios. A weighted sum of the negative log-likelihood (NLL) loss can be used in conjunction with the KLD loss to train the DCDM. This allows the DCDM to simultaneously optimize the weights in the neural network for both the mixing ratios through the NLL loss and the ground truth labels through the KLD loss.

The Dirichlet distribution can be considered a probability distribution that models multinomial distributions. Rather than directly predicting the probability of each class based on the model inputs, the distribution instead characterizes the probability of different mixing ratios of the classes occurring. This probability is then used to determine a mixing ratio and predict the composition of each pixel during testing. By applying the Dirichlet distribution to model the likelihoods over the mixture classes, the DCDM is thus effectively able to account for different mixtures in order to unmix the components in its output prediction.

Both the Regression model and the DCDM have demonstrated prior success in their application towards classifying and unmixing coral reef regions from hyperspectral data. As such, these two models are both well suited towards processing multispectral data as well, even with the distinctions between the two data formats. The primary considerations to be made when training the models on hyperspectral versus multispectral data are adjusting their hyperparameters as well as modifying the architecture to better represent the smaller amount of bands, which will be covered in latter sections.

In addition, we have identified the metrics by which we will be evaluating the

models, namely accuracy, KLD, and MAE. Accuracy will yield a fair metric for understanding how accurate the models are at determining the hard classifications of dataset. KLD and MAE will serve as an evaluation of how well the model is able to match the ground truth compositions, or otherwise how well the model is able to unmix the dataset.

Chapter 3

Path Planning Background

One of the core problems in robotics is path planning, or determining a trajectory for an agent to follow relative to an objective. These objectives can range from finding a path to a target, covering a map in its entirety for surveillance, or even just gathering information about the environment. Our application of planners uses the latter objective as a post processing step on the model output. We seek to use the uncertainty of the predictions to inform what trajectory through the map would most efficiently improve our knowledge of the region's benthic cover composition.

3.1 Informative Path Planners

Path planning is a core problem in robotics that seeks to determine a trajectory for agents to follow relative to an objective. These objectives can range from finding the shortest path between a starting and target point to determining the most efficient method of covering the entire planning region. Informative Path Planning (IPP) is used to determine where an agent should intelligently sample to gain information. The objective function of IPP prioritizes a trajectory through the map that allows the agent to gain the greatest amount of information. Prior research has demonstrated the efficiency and effectiveness of IPP in generating sampling routes [Edelson 2020].

Agents sent into the field often have resource or time limitations that prevent them from exploring the entire map. Given a probabilistic model, or prior, of the quantity being studied, IPP determines trajectories that maximize information gain

in the map within these constraints [Binney and Sukhatme 2012; Park 2008]. This allows the agent to efficiently traverse to and observe areas that will yield the most information while minimizing resource expenditure, balancing both exploration and exploitation in the path.

Monte Carlo Tree Search methods are simple but effective methods of applying IPP. These methods find optimal decisions by randomly sampling in the discretized search domain and building a decision tree with a reward based on the results. The final tree yields a set of sampling locations for the planner to take [Browne et al. 2012; Kodgule 2019]. Though these methods are optimal, they are computationally intensive as they constantly need to maintain and build up the entire tree while sampling new locations.

Regression methods are a more efficient method of IPP. One approach used regression to learn the prediction update for Monte-Carlo methods based on past simulations. This allows for the evaluation of potential information gain from candidate paths much earlier in the tree, thus expediting the IPP problem [Park 2008]. Another approach used Gaussian Process regression with autonomous planetary rovers for active spectroscopic mapping. Given previously sampled measurements, Gaussian Process regression is used to learn the spatial distribution of spectra in a spectroscopic map. This allows the rover to estimate unvisited locations and determine where to traverse to next in order to gain the most information [Candela et al. 2020].

Another efficient application of IPP that takes into account prior information was towards data collection over a region by autonomous underwater vehicles through a branch and bound method. In this study, information about the area of interest from previous data collection efforts is initially provided to the robots. This prior is represented as a scalar field of the desired measured quantity. As they explore, the robots periodically collect sensor measurements to update their known information map in order to inform and adapt their path. The resulting trajectories yielded near optimal paths for collecting information across the field, demonstrating the effectiveness of utilizing known and gained information in the map to plan trajectories [Binney et al. 2010; Binney and Sukhatme 2012].

Asymptotically optimal rapidly exploring random trees (RRT*), rapidly exploring random graphs (RRG), and probabilistic road maps (PRM*) have been combined with the above branch and bound optimization in order to generate informative trajectories

as well [Hollinger and Sukhatme 2014; Binney and Sukhatme 2012]. This method yields efficient information gathering in the space while retaining the asymptotic optimality of the constituent planning algorithms. This simply means that as the system runs, the planner will eventually sample enough or all of the environment to determine the optimal path.

All of these methods yield trajectories that are able to effectively gather information from the environment. Ergodic theory can then be applied to ensure that the sampling and traversal of the environment is efficient under the constraints of the IPP problem.

3.2 Ergodic Path Planning

Ergodic theory studies the time-averaged behavior of dynamical systems. These concepts can be used for IPP due to the inherent submodularity of the stochastic information provided to the planner [Edelson 2020; Orrieri et al. 2019]. This simply means that regions of high information density can be set as objectives for coverage. An ergodic trajectory will traverse to and spend time in these regions proportional to their expected information gain. Additionally, given enough time, an ergodic trajectory will cover and measure all states in the space, resulting in all available information about the space being gained [Edelson 2020; Mathew and Mezić 2011].

Some useful quantification of information for informative path planners to optimize over are entropy and variance. These two metrics are different measures of uncertainty. Entropy quantifies the disparity of the distribution from a uniform distribution while variance measures the average distance of outcomes from the mean in the distribution. Both have been successfully implemented in the objective function of IPP for information gathering tasks in prior work [Gautam et al. 2017; Mathew and Mezić 2011; Dressel and Kochenderfer 2018]. In the context of ergodic path planning, this means that the trajectory will spend more time in regions of uncertainty when developing the information map.

For a single trajectory to attain ergodic behavior, an optimization problem is posed that minimizes the difference between the time-averaged statistics of the agent and the information map. The agent's time-averaged statistics are computed as

follows:

$$C^t(x, \gamma_t) = \frac{1}{t} \sum_{\tau=0}^{t-1} \delta(x - \gamma_i(\tau)) \tag{3.1}$$

γ is the trajectory of the agent and is defined as $\gamma : (0, t] \rightarrow \mathcal{X}$, t is the discrete time horizon of the trajectory, δ is the Dirac delta function, and $\mathcal{X} \in \mathcal{R}^d$ in the d-dimensional search domain. The fraction of time spent at a position $x \in \mathcal{X}$ is quantified by this property [Mathew and Mezić 2011; Rao et al. 2022].

The uncertainty based information map defines the likelihood of generating informative measurements at any point within the map. Within ergodic search, the planner minimizes the difference between the Fourier spectral decomposition of both the agent's time-averaged statistics and this expected information distribution. In other words, the objective of the planner is to reduce the uncertainty in the map to 0. This is determined through the minimization of an ergodic metric Φ, defined as a weighted sum of the difference between the spectral coefficients of the two distributions [Mathew and Mezić 2011].

$$\Phi(\gamma(t)) = \sum_{k=1}^{m} \alpha_k (c_k(\gamma_t) - \xi_k)^2 \tag{3.2}$$

In this case, c_k and ξ_k are the Fourier coefficients for the agent's time-averaged statistics and the desired spatial distribution respectively, and α_k serves as the weights for each coefficient difference. These weights are typically defined as:

$$\alpha_k = \sqrt{(1 + ||k||^2)^{-(d+1)}} \tag{3.3}$$

This allows the weights to be inversely proportional to the frequency components, as lower frequency components have a greater impact on the overall information gain than higher frequency ones.

The optimization that ergodic coverage seeks to solve is the determination of optimal controls $u^*(t)$ for an agent, where the dynamics of the agent are described by the function $f : \mathcal{Q} \times \mathcal{U} \rightarrow \mathcal{TQ}$, such that:

$$\begin{aligned} u^*(t) &= \text{argmin}_u \Phi(\gamma(t)), \\ \text{subject to } \dot{q} &= f(q(t), u(t)), ||u(t)|| < u_{max} \end{aligned} \tag{3.4}$$

where $q \in \mathcal{Q}$ is the state and $u \in \mathcal{U}$ denotes the set of controls. Trajectory optimization for planning feed-forward trajectories over a defined time horizon can be used to find the optimal control input at every time step for this problem [Miller and Murphey 2013].

3.3 Sparse Ergodic Path Planning

The prior optimization acts under the assumption that the agent is sampling at every single point in the trajectory. However, oftentimes the ergodic trajectory may pass through regions of low information density which do not yield as much information per sensor measurement. As such, measuring at every step is an inefficient use of resources. Modifications to the prior ergodic functions were then developed for sparse ergodic optimization, which seeks to optimize the points where sensor measurements are taken over the ergodic trajectory while still yielding high information gain. This demonstrated that there does exist an optimal set of sensing measurements for a given coverage scenario following ergodicity [Rao et al. 2022].

The new ergodic optimization for sparse sensing is expanded upon from Equation 3.4 as:

$$\begin{aligned} u^*(t), \lambda^*(t) &= \text{argmin}_{u,\lambda} \Phi'(\gamma(t)), \\ \text{subject to } \dot{q} &= f(q(t), u(t)), ||u(t)|| < u_{max} \end{aligned} \tag{3.5}$$

where the new variable $\lambda(t) \in \{0, 1\}$ serves as a decision variable. Sparsity is then promoted in the sample measurements by regularizing λ with an L_1 metric. This yields the augmented ergodic metric from Equation 3.2:

$$\Phi'(\gamma(t)) = \sum_{k=1}^{m} \alpha_k (c_k(\gamma_t, \lambda_t) - \xi_k)^2 + \sum |\lambda_k| \tag{3.6}$$

The time-averaged statistics of the agent and the information map from Equation 3.1 is then updated as:

$$C'^t(x, \gamma_t) = \frac{1}{\sum_t \lambda_t} \sum_{\tau=0}^{t} \lambda(t) \delta(x - \gamma_i(\tau)) \tag{3.7}$$

where $\lambda(t) \in \{0, 1\}$ [Rao et al. 2022].

As it is difficult to solve mixed integer programming problems due to a lack of a gradient, $\lambda(t)$ is generalized from discrete values of $\lambda(t) \in \{0, 1\}$ to a bounded continuous range of $\lambda(t) \in [0, 1]$ during optimization. After λ is optimized over this range, it is then projected back into the discrete values of $\{0, 1\}$ [Rao et al. 2022].

In situ expeditions are a timely and costly expenditure, which was why we approach the problem of coral mapping through machine learning with remote sensed data. However, these approaches still have uncertainties associated with their prediction as they are not completely accurate. The output of the models is able to not only leverage the efficiency of remote sensing for mapping, but also inform routes for the expeditions by serving as an information prior for IPP. In addition, with regards to the resource costs of *in situ* expeditions, sparse ergodic planners are a fully appropriate approach to this problem in order to ensure that the sensing costs are as efficient as possible.

Chapter 4

Approach

The raw data products used in this research were preprocessed with atmospheric and surface reflectance correction. In addition, due to an imbalance of classes in the dataset, the data is further balanced through clustering in order to prepare for training the model. Two primary model architectures are trained on the dataset, the Regression Model and the Deep Conditional Dirichlet Model (DCDM). Variances are calculated from the model prediction outputs and weighted in order to obtain information priors. These priors are then passed through the sparse ergodic planner in order to optimize a simulated *in situ* trajectory through the map.

4.1 Data

The data used in this research were in the form of remote sensed hyperspectral and multispectral data from different satellites and sensing platforms, particularly PRISM, Landsat-8, and WorldView-3. All of the input data were preprocessed for atmospheric correction and surface reflectance from the water. Corresponding labels were obtained by taking the preprocessed PRISM data and passing them through ENVI spectral image processing and analysis software, verifying the resulting output through personal experience and measurements from NASA's *in situ* Earth Venture Suborbital-2 COral Reef Airborne Laboratory (CORAL) mission.

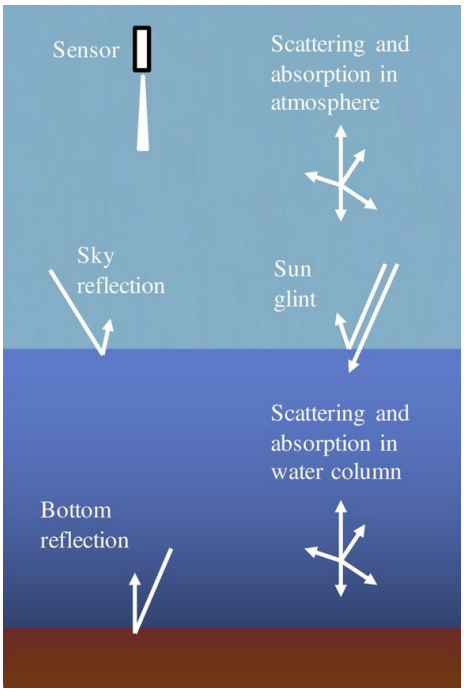


Figure 4.1: Environmental effects on spectral data measurements [Thompson et al. 2017]. These effects from the atmosphere, at the water surface, and in the water column are estimated, inverted, and corrected in order to obtain the seafloor benthic reflectance.

4.1.1 Data Products

The CORAL mission was used to collect data and map portions of the Great Barrier Reef, Main Hawaiian Islands, Mariana Islands, and Palau in 2016-2017 [Hochberg and Gierach 2018]. Conventionally, remotely sensed mapping of coral reefs has relied either on in-scene statistics (so-called "supervised classification") or on algorithmic correction for atmosphere, sea surface, and water column radiative transfer effects to derive seafloor reflectance. The derived bottom reflectances are used to discriminate between benthic classes. The algorithmic approach is not scene-specific and theoretically can be applied globally; hyperspectral data affords the best retrievals [Lee and Carder 2002]. The PRISM data of Heron Island used in this study were obtained on 17 September, 2016, from an airborne platform at 8.5 km altitude. PRISM provided water-leaving reflectance from 380-1050 nm at approximately 3 nm spectral resolution and 8 m spatial resolution. For this mission, using shallow-water reflectance, PRISM data

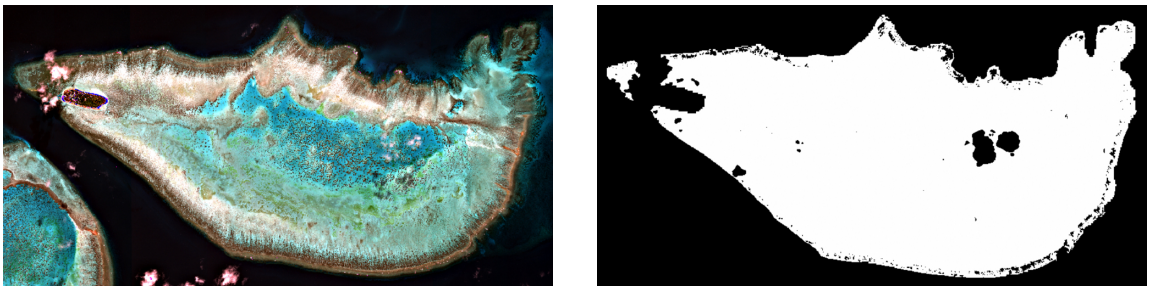
were processed using Bayesian optimization to provide seafloor benthic reflectance and benthic functional type (Figure 4.1) [Thompson et al. 2017; Maritorena et al. 1994]. This in in turn was used as input in logistic regression to predict probabilities of coral, algae, and sand for each pixel. CORAL's map products were validated using diver-based, in-water observations to yield final PRISM data and labels.

The Landsat-8 data of Heron Island were acquired on 26 October 2016; the surface reflectance products were then retrieved from both the USGS EarthExplorer (https://earthexplorer.usgs.gov/) and USGS Global Visualization Viewer (https://glovis.usgs.gov/). The data itself is composed of 7 bands ranging from 440-2200 nm with unequal spectral resolution and 32 m spatial resolution. The WorldView-3 data were collected on 4 July, 2016 through the Maxar WorldView-3 satellite and was provided through NASA/GSFC and the NextView license agreement. It consisted of 8 bands of 420-950 nm wavelengths with approximately 75 nm spectral resolution and 2 m spatial resolution.

The purpose behind transferring the labels of the PRISM data to WorldView-3 rather than directly labeling WorldView-3 data is that direct labels require in-water observations, which are extremely limited owing to logistical constraints. The CORAL mission, however, obtained hyperspectral imagery for about 43,000 km^2 of the Great Barrier Reef and its surrounding area, affording a very large and sufficiently accurate training set. Our work focuses on a demonstration at Heron Island, where in-water validation data exist. It seeks to classify and unmix the benthic scene by using high resolution spectral data priors from WorldView-3 as inputs to our network in order to yield more accurate per pixel labels. This allows our network to make more distinctive classifications on pixel regions while also being able to more accurately perform unmixing through the DCDM on the benthic coverage.

4.1.2 Raw Data Processing

Since our data was captured in different resolutions and alignments, we use ENVI, an image processing and analysis software, to coregister the scenes as well as scale them to corresponding resolutions. In order to retain the most spatial information within the scene, the Landsat-8 data and PRISM labels were both upscaled to the resolution of the WorldView-3 data using bilinear interpolation as opposed to downscaling the



(a) Base WorldView-3 Data (b) Sample Mask

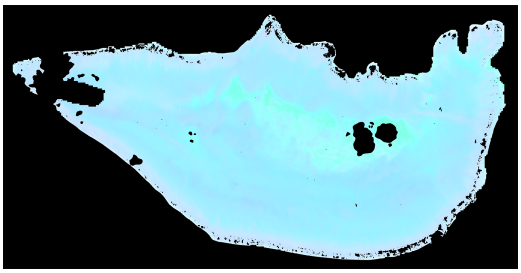
Figure 4.2: Masks were applied to the datasets ensuring that only pixels above 5 m in depth would be considered during training. A sample of the mask applied is shown above, where white indicates valid pixels and black indicates invalid pixels. (WorldView-3 data © 2016 Maxar).

WorldView-3 data. Bilinear interpolation was chosen over nearest neighbors when performing the scaling to retain continuity between neighboring pixels rather than consider upscaled pixels as discrete patches. We also applied a mask to the data during training such that only pixels shallower than 5 m in depth are used (Figure 4.2). This is to minimize inaccuracies arising from radiative transfer effects: shallow data are more reliable. Additionally, as shorter wavelengths are more affected by water refraction and wavelengths outside the visible spectrum are more absorbed by water, we only trained on visible wavelength bands across all of the spectral data. The processed data and labels are presented in Figure 4.3.

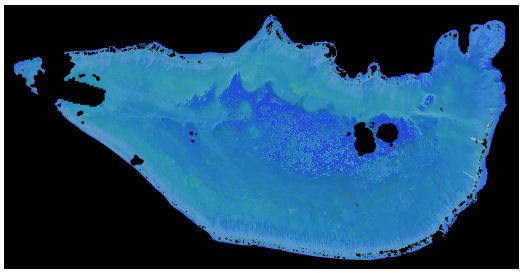
4.1.3 Balancing and Clustering

There is a high abundance of algae labels within the scene in comparison to the sand pixels, and especially with regards to the coral pixels. The dataset we train on in this project consists of 56.93% algae, 42.55% sand, and just 0.5296% coral. As a result, this tends to bias any model output towards algae predictions as it is the most frequent label observed in training. A high enough performance is able to be attained just through solely classifying everything as algae or sand without considering any coral classifications.

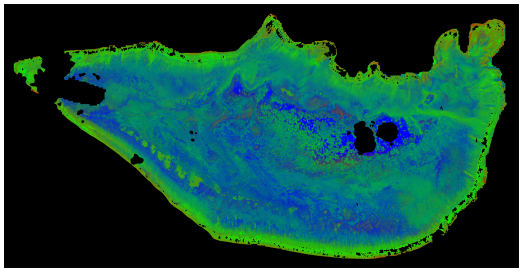
In order to combat this issue, after the data is aligned and masked, it is then balanced through clustering. Clusters of the data are taken to obtain a subset that



(a) Landsat-8 Data



(b) WorldView-3 Data



(c) Labeled PRISM Data. The benthic classes are represented as Coral (Red), Algae (Green), and Sand (Blue).

Figure 4.3: Samples of the final processed and masked data products. (WorldView-3 data © 2016 Maxar).

better represents the three distinct classes when training. This also allows the model to more effectively generalize across all of the classes in testing. Two primary methods of clustering the data were applied within this research.

The first method of balancing applied during training was through the k-means++ algorithm. This algorithm identifies a specified number of cluster centers within the dataset and groups datapoints based on their proximity to those cluster centers [Arthur and Vassilvitskii 2007]. In our case, k-means++ is applied on the labels in order to isolate three clusters of datapoints. Our output dataset is then composed by selecting an equivalent number of points from each cluster. While this method is able to obtain a balance of pixels that are representative of the three clusters, the clusters themselves may not necessarily be fully representative of each class. This is especially the case when considering not only the abundance of algae in the dataset and the likelihood of it being part of any cluster, but also the fact that the pixels themselves are mixtures of classes. This in turn makes it difficult to completely separate the

classes into three distinct clusters.

The second method applied was balancing the classes with a maximum balancing before applying the k-means++ algorithm. The maximum class per pixel is chosen to represent every pixel in the dataset, and the class with the smallest number of pixels determines the number of randomly sampled pixels from the other classes. The primary benefit of this approach is that all of the classes are guaranteed to be represented equally during training with regards to their hard label classifications. However, due to the upper bound on the maximum number of datapoints that may be drawn from each class, the data subsets may not necessarily be able to represent all of the nuances in the class mixtures. The k-means++ application attempts to represent these nuances through clustering the pixels in the subset dataset.

Once the data is properly balanced and clustered, it is ready to be passed into the different models. These models are adaptive to the number of bands in the input data, and their exact architectures are covered in the following section.

4.2 Models

We evaluate several Fully Connected Network (FCN) architectures to train over the course of our research as well as provide comparison to a linear Support Vector Machine (SVM). The linear SVM initially maps inputs into a higher dimensional feature space $\mathcal{Z}$ through a non linear mapping. Within this feature space $\mathcal{Z}$, subsets of support vectors representing the training inputs are used to determine the maximum margin, or distance, between points in different classes. These support vectors are used to construct the optimal hyperplane in the original input space that separates the different input classes [Cortes and Vapnik 1995].

FCNs are neural networks where each of the nodes in a single layer receive inputs from all of the nodes in the prior layer. These types of networks are also known as multilayer perceptrons. The initial FCN models used for this research were developed in work by Candela on scene unmixing [Garza 2021]. In our research, the two types of models employed from that resource are the Regression model and the DCDM.

The Regression model architecture is set up through three dense hidden layers, all with Scaled Exponential Linear Unit (SELU) activation, as well as a final dense layer with a softmax activation. The width of the hidden layers for the base network

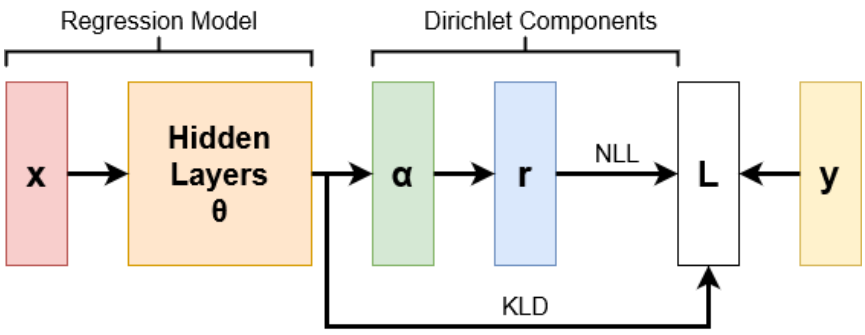


Figure 4.4: The architecture of the FCNs takes a spectrum x as input, predicting a set of probabilities as the output from the hidden layers θ prior to calculating the KLD loss as part of the Regression model. The DCDM takes the output of the hidden layers θ to predict the values of α, the concentration parameters of the Dirichlet distribution. These values are used to infer the mixing ratios r and are then used to calculate the NLL loss. The DCDM employs both KLD and NLL loss in order to estimate the mixing of benthic components while fitting the labeled probabilities.

is determined by the number of bands in the input data and the desired number of output classes. This architecture was chosen to unmix endmembers of the network and serve as a neural network baseline comparison to the DCDM (Table 4.1, Figure 4.4).

The DCDM architecture is set up with the same initial framework as the Regression model architecture. However, as we are attempting to obtain a set of mixture probabilities, we add an intermediary layer α with a modified sigmoid function that outputs between 0-100. This layer is used as an input into the Dirichlet distribution to model the different distributions of mixtures. Afterwards, a softmax is applied to the output of α in order to get a set of concentration parameters for the Dirichlet model mixing estimation. This architecture was chosen to learn parameters for the Dirichlet distribution to unmix benthic classes (Table 4.2, Figure 4.4).

We use two loss functions when training our models. The KLD is a metric used to compare the direct probability values between the output of the model and the labeled values. In the case of the Dirichlet model we are also trying to train the model to account for unmixing, so we incorporate the NLL loss which is used to optimize the unmixing parameters in the α layer of the DCDM.

The primary goal of the machine learning model is to be able to classify and unmix

Layer	Type	Activation	Number of Units
Input	-	-	n
Hidden Layer 1	Linear	SELU	0.75*n+0.25*m
Hidden Layer 2	Linear	SELU	0.5*n+0.5*m
Hidden Layer 3	Linear	SELU	0.25*n+0.75*m
Output	Linear	Softmax	m

Table 4.1: Regression Model Architecture

Layer	Type	Activation	Number of Units
Input	-	-	n
Hidden Layer 1	Linear	SELU	0.75*n+0.25*m
Hidden Layer 2	Linear	SELU	0.5*n+0.5*m
Hidden Layer 3	Linear	SELU	0.25*n+0.75*m
α	Linear	Sigmoid	m
r	-	Softmax	m

Table 4.2: Deep Conditional Dirichlet Model Architecture

the three benthic classes, consisting of coral, sand, and algae, based on distinctions in the spectral data. However, as both coral and sand are quite similar in composition and color to one another (especially as some forms of sand are derived from dead coral) and we are provided with limited data for training, we do not expect to obtain a particularly high classification accuracy from the model. We are instead able to demonstrate improved performance for this problem in comparison to previous approaches and methods.

4.3 Information Map Prior

In order to perform IPP across the coral region and determine an ergodic trajectory, it is necessary for the model to be provided with an information prior. The uncertainty of the neural networks typically originate from either uncertainty in the data (aleatoric uncertainty) or uncertainty in the model (epistemic uncertainty) [Gawlikowski et al. 2022]. To quantify the uncertainty and generate a prior, different subsets of the

training dataset are used to train the same model. The variance of the model outputs through this approach on the test dataset quantifies the uncertainty within the model. In addition, a random prior where all of the probabilities are randomly initialized is provided to the IPP for baseline comparison.

When planning a trajectory through the map for further *in situ* measurements, we are primarily concerned with observing the distribution of coral in the field as opposed to algae or sand. As such, we can also apply a bias towards the coral class when developing the information prior. After calculating the variance of the coral, algae, and sand classes from the different model outputs, we normalize these maps through a max norm in order to ensure that their uncertainty is of the same magnitude. Afterwards, a weighted sum of the normalized maps yields the final uncertainty prior used as the information map for the planner. The average variance across the three classes for the output is calculated as a comparative prior as well.

Due to the computational limits of the device that the experiments are run on, we are only able to determine a sparse ergodic trajectory for the scene at a 16m^2 per pixel resolution. Similar to the preprocessing of the spectral data, the model output predictions were rescaled using bilinear interpolation to retain neighboring pixel continuity. ENVI was used to process the outputs and yield the final priors to input into the planner.

4.4 Planner Setup and Evaluation

The dynamics of the agent for the planner are defined as the constrained discrete time dynamical system:

$$x(t+1) = f(x,u) = x(t) + \tanh(u) \tag{4.1}$$

where $x(t) \in \mathcal{R}^2$ and $u \in \mathcal{R}^2$ is bounded by the tanh function. This bound is applied in order to constrain the results to the range $[0,1]$. The discrete values of $\lambda(t) \in \{0,1\}$ used in the sparse ergodic planner are instead modeled as a bounded continuous variable $\lambda(t) \in [0,1]$, which improves computational efficiency without negatively impacting performance [Rao et al. 2022].

The ergodic planner seeks to minimize the Fourier spectral decomposition of

the two distributions. As such, the Fourier basis modes that satisfy the Neumann boundary conditions must be calculated and applied across the distributions [Mathew and Mezić 2011]. For a rectangular domain of $U = [0, L_1] \times [0, L_2]$, the Fourier basis are found by the following:

$$F_k(x) = \frac{1}{h_k} \cos(\frac{k_1 \pi x_1}{L_1}) \cos(\frac{k_2 \pi x_2}{L_2}) \tag{4.2}$$

where

$$h_k = \sqrt{\int_0^{L_1} \int_0^{L_2} \cos^2(\frac{k_1 \pi x_1}{L_1}) \cos^2(\frac{k_2 \pi x_2}{L_2}) dx_1 dx_2} \tag{4.3}$$

is a normalization factor that ensures that F_k has L2 norm equal to 1, $k = [k_1, k_2]$ are positions in the domain, and x is the base distribution input value.

As we are applying a sparse ergodic planner to determine a trajectory through the region, equations 3.5, 3.6, and 3.7 are applied to determine the optimization function formulation, the ergodic metric to optimize in the optimization loop, and the time-averaged statistics of the agent respectively. The primary difference between a standard and sparse ergodic planner is the inclusion and optimization of $\lambda(t)$, which determines when a sensor measurement should be taken.

An additional barrier cost function and L1 norm penalty are added to the optimization problem in equation 3.5. The barrier cost function is defined as:

$$\mathcal{J}(\gamma(t)) = 100 * (\max(0, \gamma(t) - 1)^2 + \max(0, -\gamma(t))^2) \tag{4.4}$$

where γ is the trajectory of the agent. The L1 norm penalty is defined as:

$$L_1(t) = \int_t |\lambda(t)| dt \tag{4.5}$$

where $\lambda(t)$ is the bounded continuous range of values denoting whether a sensor measurement is taken or not. These two additions to the optimization function seek to stabilize the system. The barrier function ensures that the trajectory remains within the bounds of the map. The L1 norm penalty bounds the values of the sensor measurements while also promoting sparsity by encouraging smaller values.

Afterwards, the initial conditions for the planner are defined. These include

defining the starting position of the agent, determining the number of steps allocated for the agent to take, and initializing the set of controls u to 0 and λ values for the sensor measurements to 1. Finally, the trajectory is optimized through the application of the Adam optimizer with learning rate 1e-3, using the modified optimization function from equation 3.5 with equations 4.4 and 4.5 [Kingma and Ba 2014].

After a trajectory has been determined and a final information map is obtained, we can update the model output using the model labels as an approximation of the results from an *in situ* expedition. We define this model prediction update as follows:

$$M' = M + (\mathcal{L} - M) * (1 - \frac{\mathcal{D} - 0.5}{0.5}) \tag{4.6}$$

where

$$\mathcal{D} = \text{sigmoid}(|c(\gamma, \lambda) - \xi|) \tag{4.7}$$

M is the model output prediction, $\mathcal{L}$ are the labels, and c and ξ are the Fourier coefficients for the agent's final time-averaged statistics and the desired spatial distribution respectively. Equation 4.7 scales the difference between c and ξ to the range $[0, 1]$. Note that this function assumes a linear relationship between the information gain and the model update.

The same metrics of accuracy, KLD, and MAE used to initially evaluate the model output are applied to the updated output as well. This allows us to observe the impact each distinct prior had on improving the coral map. It also provides a basis of comparison between the Regression model and DCDM outputs in terms of how much an *in situ* expedition would be able to improve the respective outputs. While regions that the agent does not pass over are also updated by the metric in Equation 4.6, the relative performance of the updated predictions are still maintained, allowing for fair comparison of how these trajectories would fare in the field.

The entire pipeline in this research is established as taking input data, preprocessing it, passing it through a model for training and evaluation, taking the model output variances as priors for the planner, and using the information map difference from the planned trajectory to update the model predictions. The modifications to the optimization function for the sparse ergodic planner also serve to ensure that the trajectory and sensor measurements are well bounded during training. We now cover the specific parameter tuned experimentation and evaluation using this pipeline.

Chapter 5

Experiments

5.1 Mapping and Unmixing

Our models were trained with a batch size of 32, Adam optimizers, and learning rates of 1e-3 for all loss functions [Kingma and Ba 2014]. For the DCDM, the KLD was weighted by 0.95 and the NLL was weighted by 0.05 to obtain the total loss. We trained on a dataset clustered with k-means++, as well as on a dataset first balanced with maximum balancing then k-means++.

We evaluated our models using accuracy, KLD, and MAE. Accuracy is defined as the number of correct maximum probability classifications per pixel over the total number of pixels, KLD is a metric the measures the statistical distance between two probability distributions, and MAE measures the average absolute difference between actual and predicted values. Accuracy was chosen in order to represent general performance in hard classification of the benthic cover. KLD and MAE quantify how closely our predictions align with the actual probabilities, and thus how well the model is able to unmix the benthic cover classes. In addition to the evaluation of these values across all of the classes (Table 5.1, Table 5.3), model performance on coral as the sole class of interest is also included (Table 5.2, Table 5.4).

Beginning with the k-means++ clustered dataset, all of the methods achieve comparable accuracy, with SVM performing slightly better on the Landsat data while the Regression model performs slightly better on the WorldView data as seen in Table 5.1. The reason behind the SVM's performance is due to classifying all pixels as

Data	Model	Accuracy	KLD	MAE
Landsat	SVM	72.68%	—	—
Landsat	Regression	72.07%	0.0393	0.1867
Landsat	DCDM	72.22%	0.0419	0.1878
WorldView	SVM	69.79%	—	—
WorldView	Regression	70.38%	0.0487	0.2054
WorldView	DCDM	69.91%	0.0512	0.2056

Table 5.1: Overall test set results using Landsat and WorldView data clustered with k-means++ on SVM, Regression Model, and DCDM

Data	Model	Accuracy	KLD	MAE
Landsat	SVM	0%	—	—
Landsat	Regression	25.60%	0.0937	0.3078
Landsat	DCDM	22.34%	0.1015	0.3163
WorldView	SVM	0%	—	—
WorldView	Regression	6.170%	0.1127	0.3535
WorldView	DCDM	22.14%	0.1025	0.3215

Table 5.2: Coral test set results using Landsat and WorldView data clustered with k-means++ on SVM, Regression Model, and DCDM

algae or sand without any coral classifications. The high abundance of the former two classes in the dataset even after clustering allows the SVM to attain a high accuracy without classifying coral, as seen in Table 5.2.

Between a Regression model and DCDM trained on the same dataset, the performance across all metrics appear to be roughly similar with regards to all of the classes as seen in Table 5.1. Landsat achieves an accuracy of approximately 72.15%, a KLD of 0.406, and an MAE of 0.1872 while WorldView performs slightly worse with an accuracy of about 70.15%, a KLD of 0.0500, and an MAE of 0.2055. A slight increase in accuracy is observed with the Landsat data between the two models. In all other metrics, the DCDM appears to perform slightly worse in comparison to the Regression model. These results all seem to contradict our expectations for improvement in performance with high spatial-resolution WorldView data alongside the unmixing capabilities of the DCDM. However, this is considering all of the classes together in the evaluation, where both algae and sand dominate the test set.

Data	Model	Accuracy	KLD	MAE
Landsat	SVM	63.17%	—	—
Landsat	Regression	70.35%	0.0446	0.2011
Landsat	DCDM	71.45%	0.0542	0.214
WorldView	SVM	62.43%	—	—
WorldView	Regression	69.96%	0.0544	0.2162
WorldView	DCDM	69.13%	0.0632	0.2308

Table 5.3: Overall test set results using Landsat and WorldView data clustered with maximum balancing and k-means++ on SVM, Regression Model, and DCDM

Data	Model	Accuracy	KLD	MAE
Landsat	SVM	88.16%	—	—
Landsat	Regression	75.77%	0.0605	0.2012
Landsat	DCDM	0%	0.3383	0.6075
WorldView	SVM	89.27%	—	—
WorldView	Regression	73.21%	0.0584	0.2011
WorldView	DCDM	76.55%	0.0664	0.2028

Table 5.4: Coral test set results using Landsat and WorldView data clustered with maximum balancing and k-means++ on SVM, Regression Model, and DCDM

With regards to the coral test set results in Table 5.2, a couple of distinct improvements can actually be observed with the WorldView data in comparison to the Landsat data. Landsat decreases in performance across all evaluation metrics between the Regression model and the DCDM, while WorldView is able to improve its accuracy by almost 3 times while also decreasing its KLD and MAE, indicative of successful unmixing. While the evaluation of the models on all of the classes do not appear to meet our expectations, there is supportive evidence for improved performance as a result of the higher spatial-resolution data and the DCDM. This is demonstrated by the fact that they are able to successfully isolate and improve classifications on the more minor coral class.

Moving on to the maximum balanced then k-means++ clustered dataset, the coral class now has increased representation in the dataset than the direct k-means++ clustering. As a result, SVM now classifies coral quite well in Table 5.4, attaining the highest accuracy across all methods. However, it is unable to separate pixels with

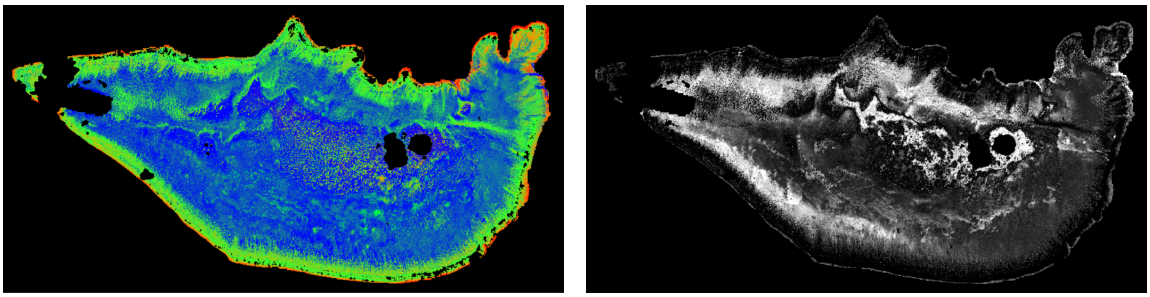
similar compositions to multiple classes. SVM is thus unable to perform as well on all the classes at once and achieves significantly lower accuracy on both datasets than the Regression model and the DCDM, as seen in Table 5.3.

From Table 5.3, similar trends to Table 5.1 can be seen with the Landsat and WorldView datasets between the Regression model and the DCDM performance. Landsat experiences a slight increase in accuracy while WorldView slightly decreases in performance across all metrics, once again in opposition to our expectations. Considering results on just the coral pixels in Table 5.4, we observe that once again WorldView achieves better accuracy with the DCDM but also experiences a decrease in performance with KLD and MAE. A drastic decrease in performance with coral classification is observed on the Landsat data, where it attains 0% accuracy. While the exact cause is undetermined, we conclude that characteristics of the clustered Landsat dataset alongside the softmax activation cause the Dirichlet component to standardize the concentration parameter value of coral to 1. As algae and sand typically attain higher concentration parameter values due to their prioritization, this results in zero coral classifications.

Overall, the SVM as a standard machine learning approach does give fair accuracy, but distinctly fails to model any mixing of classes. The Regression model achieved better KLD and MAE than the DCDM in general as the DCDM tended to increase these metrics on all classes when applied. However, the DCDM does give better representation for coral as demonstrated by increases in accuracy when classifying coral observed in both types of clustering cases. Finally, while Landsat appears slightly more suited for overall classification accuracy, WorldView data tends to outperform it when focusing on classifying coral. Note that these conclusions are based on a dataset with minimal coral representation. While distinctions are observed between models and datasets, performance across the board is quite similar. More data is needed to establish statistical significance for these conclusions.

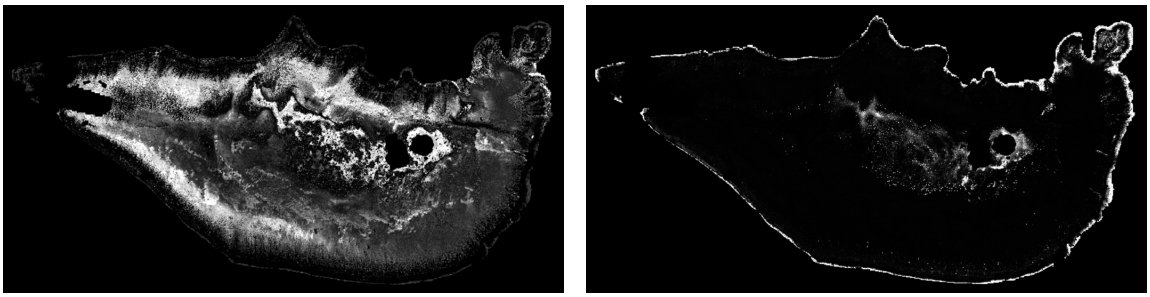
5.2 Planning for Uncertainty

For the planning component, we proceed with the Regression model and the DCDM both trained with the hyperparameters defined as in Section 5.1 on WorldView data with maximum balancing and k-means++ clustering as a standard. The training



(a) The average prediction across all four models.

(b) The base weighted prior.



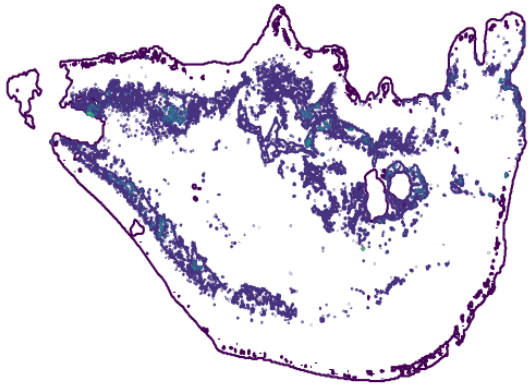
(c) The equally weighted prior.

(d) The coral weighted prior.

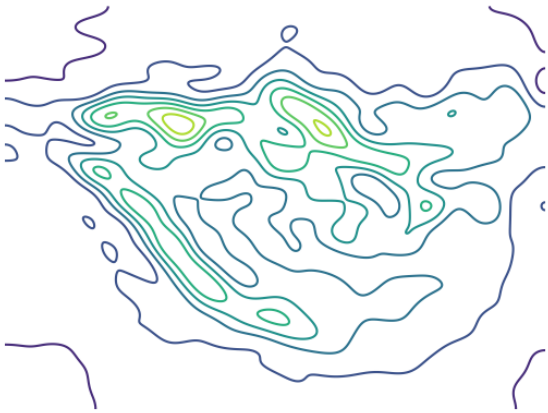
Figure 5.1: Average Prediction and Weighted Variance Priors. Four subsets of the maximum balanced and k-means++ clustered dataset were used to train DCDMs. The average model predictions and weighted variances were calculated from the outputs.

dataset was equally divided into four parts, and four models were trained on each of these data subsets. The average of the output predictions is then taken to be the overall base model prediction. The variance of each of the classes across the four predictions is calculated and normalized, before a weighted sum is calculated. During testing, the weights of each class variance $[\omega_C, \omega_A, \omega_S]$ were set for the coral, algae, and sand variances respectively. Overall, we calculated a weighted prior biased towards coral with $[\omega_C, \omega_A, \omega_S] = [0.6, 0.2, 0.2]$, an equal prior where the variances were averaged with $[\omega_C, \omega_A, \omega_S] = [0.33, 0.33, 0.33]$, and a coral prior considering only the coral variance with $[\omega_C, \omega_A, \omega_S] = [1, 0, 0]$ (Figure 5.1).

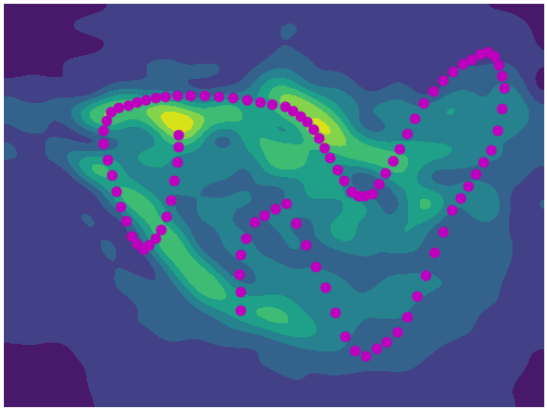
The starting location for the trajectory that the ergodic planner optimizes was set to the point of maximum variance in each prior. An Adam optimizer with learning rate 1e-3 was applied for the planner to optimize the ergodic loss function [Kingma and Ba 2014]. A total of 1000 iterations were allotted for the planner to adjust the



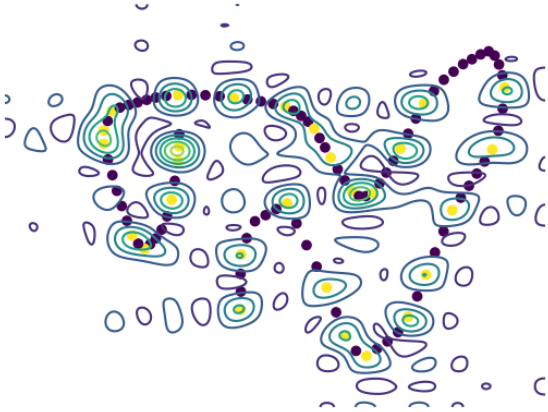
(a) The weighted prior found by taking a weighted sum of the class variances from the four model outputs.



(b) The weighted prior after a Fourier basis is applied for smoothing. Warmer colors have higher weights.



(c) The output trajectory from the planner.

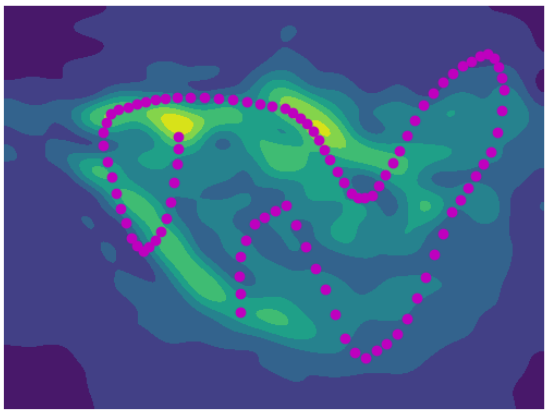


(d) The final information map of the agent from the trajectory.

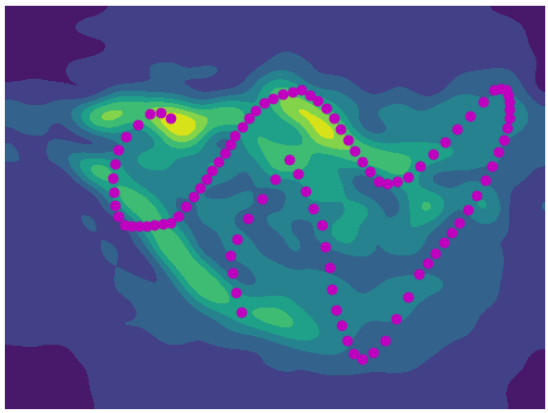
Figure 5.2: Outputs for the planner on the weighted prior from the DCDM

trajectory, and the length of the trajectory was set to 100 steps. The set of controls u were all initialized to 0 while the sensing values λ were all initialized to 1 for sensing at every step.

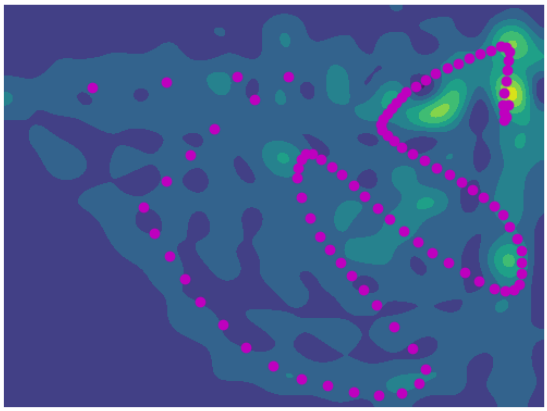
A fourier basis function is applied across the input prior to yield the target information map. This smooths out the probability distribution for the planner to optimize across. In our application, $L_1 = L_2 = 1$ and a meshgrid is used to evaluate across the entire scope of the map. In addition, we simplify the Fourier basis calculations by setting the normalization factor h_k to the first value of h_0 calculated



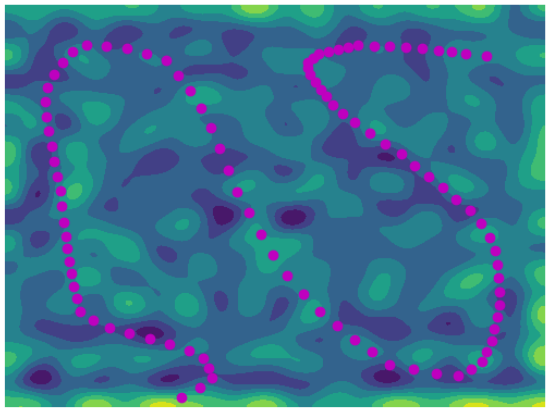
(a) The weighted prior trajectory.



(b) The equal prior trajectory.



(c) The coral prior trajectory.



(d) The random prior trajectory.

Figure 5.3: Final trajectories across all the different priors for the DCDM

using Equation 4.3, as this yielded smoother generalized contours in the information map [Rao et al. 2022]. The final simplified Fourier basis equation in our application is thus defined as:

$$F_k(x) = \frac{1}{h_0} \cos(\pi k_1 x_1) \cos(\pi k_2 x_2) \tag{5.1}$$

Over the course of the training, the planner tries to match the information map of the trajectory to this target map by minimizing the ergodic loss, which is a weighted difference between these two maps as defined by Equation 3.6. The agent's final trajectory yields the final information map used to update the average model prediction, thus simulating an *in situ* expedition (Figure 5.2, Figure 5.3). A linear

Model	Acc	KLD	MAE
Regression	70.36%	0.1331	0.4556
DCDM	69.59%	0.1337	0.4557

Table 5.5: Average model prediction performance for the Regression model and the DCDM on all classes. These were calculated using the average model prediction from the four model runs against the ground truth labels for all classes.

Model	Prior	Accuracy	KLD	MAE
Regression	Weighted	92.97%	0.0399	0.1847
Regression	Equal	92.15%	0.0396	0.1837
Regression	Coral	94.46%	0.0376	0.1725
Regression	Random	96.31%	0.0256	0.1388
DCDM	Weighted	92.62%	0.0385	0.1790
DCDM	Equal	92.11%	0.0390	0.1809
DCDM	Coral	94.15%	0.0353	0.1643
DCDM	Random	95.85%	0.0257	0.1387

Table 5.6: Model prediction performance after planner updates on all classes

relationship is assumed between the amount of information gain and the update to the model prediction.

Similarly to the evaluation of the machine learning models, accuracy, KLD, and MAE are used to evaluate the final output maps. The performance of the average predictions of the four models are obtained by comparing them against the ground truth labels for all classes (Table 5.5) as well as for just the pixels of the ground truth where coral is the maximum class (Table 5.7). The performance of the updated predictions using the information maps of the planned trajectories are also included for evaluation. They are separated into performance across all the classes (Table 5.6) and on just the pixels in the ground truth with coral as a maximum (Table 5.8). As Equation 4.6 even updates the average model predictions for points that the trajectory does not come close to, we are interested in observing the relative performance of the updated predictions as opposed to their magnitude.

Regarding the initial performance of the models on all classes, both models perform roughly equivalently, though the Regression model slightly outperforms the DCDM (Table 5.5). The model outputs experience an increase in performance across all

Model	Accuracy	KLD	MAE
Regression	71.49%	0.1397	0.4840
DCDM	71.70%	0.1380	0.4793

Table 5.7: Average model prediction performance for the Regression model and the DCDM on coral class. These were calculated using the average model prediction from the four model runs against the ground truth label pixels that had coral as the maximum class.

Model	Prior	Accuracy	KLD	MAE
Regression	Weighted	90.50%	0.0180	0.0984
Regression	Equal	91.45%	0.0147	0.0859
Regression	Coral	88.70%	0.0417	0.1897
Regression	Random	87.54%	0.0395	0.1837
DCDM	Weighted	93.03%	0.0144	0.0876
DCDM	Equal	91.97%	0.0136	0.0836
DCDM	Coral	88.70%	0.0426	0.1930
DCDM	Random	89.23%	0.0389	0.1817

Table 5.8: Model prediction performance after planner updates on coral class

metrics when updates from the information map are applied, regardless of the prior type. In the final outputs as seen in Table 5.6, though the Regression model still retains higher accuracy than the DCDM between equivalent priors, the DCDM is able to achieve lower KLD and MAE. This is indicative of results that more closely match the distribution of ground truth benthic class mixtures and suggest that the DCDM is able to more intelligently sample locations that improve its performance than the Regression model.

The updated predictions on each model achieve similar performance across all metrics on all classes regardless of the prior used (Table 5.6). Out of them, the random prior and the coral prior achieve the best performance, while the weighted and equal priors follow close behind and are roughly equivalent. These results can be attributed to the coverage of the final trajectories as seen in Figure 5.3. As we are updating the entire map based on the update function 4.6, prioritizing exploration over exploitation yields better results on overall classes. Even though the coral prior seeks to exploit high uncertainty regions of coral pixels, due to the dispersion of

regions of high information in the prior, the ergodic planner instead yields a more exploratory trajectory.

When considering the coral pixel classifications, the DCDM initially outperforms the Regression model across all metrics in Table 5.7. All of the metrics are improved through updates from the information map as seen in Table 5.8, where the DCDM continues to outperform the Regression model overall. Notably, while their accuracy is still similar, the weighted and equal priors now perform the best in comparison to the coral and random priors. Even though all of the priors had similar values over the metrics with all classes, the KLD and MAE of the weighted and equal priors is much lower than those of the coral and random priors with respect to just the coral pixels. This demonstrates how their trajectories and samples are able to better exploit coral uncertainties.

Overall, the general increase in performance with information map updates suggest that *in situ* expeditions would definitely improve benthic cover mapping. The DCDM especially benefited from the update, generally achieving better performance than the Regression model across all metrics after the update. The weighted and equal priors demonstrated how exploitation of a class when sampling can improve the prediction of that class. In contrast, having regions of high information too spread out as in the case of the coral and random priors can cause the trajectory to prioritize exploration, leading to better general prediction performance as opposed to individual class performance.

Chapter 6

Conclusions

Coral reefs are still in rapid decline to this day, thus necessitating a global scale method of monitoring their benthic cover in order to understand the trajectory of their condition. Remote sensed smallsat data has been demonstrated to be a cost effective method of obtaining global scale data. In addition, developments in deep learning approaches for spectral reconstruction using this data for training have proven successful as well [Candela et al. 2021; Garza 2021]. By employing these approaches to train and map global scale spectral data, we are able to yield a system that satisfies our need for mapping and monitoring benthic cover distributions over time.

In this project, we are able to successfully train Regression models and Deep Conditional Dirichlet Models to classify and unmix multispectral data with roughly 70% accuracy. While this value may appear fairly low with regards to other classification or regression tasks, all three classes in the benthic cover have similar spectral values and signatures to one another. The ambiguity between the three classes gives reason for this accuracy. We demonstrate that an increase in spectral training data resolution improves the performance of the model in classifying lesser known classes such as coral in this case. In addition, we also find that the DCDM is able to successfully unmix coral pixels to improve their classification on the high spatial-resolution data, though more tuning may be necessary for lower spatial-resolution applications.

One concern from the experimental results is that even though an improvement in accuracy is consistently observed with the application of the DCDM towards

coral pixels, the KLD and MAE tend to decrease in performance. As these metrics quantify unmixing, this contradicts the expected performance and the success found in prior work. One explanation for this behavior is simply the limitations of the data. Our data has a minimal amount of coral in comparison to algae and sand. Though we attempt to mitigate this issue through clustering, the clustered datasets may still not be fully representative of all classes and mixtures, thus leading to a decrease in performance. Another explanation for the observed behavior is the network architecture itself, as the architecture is adaptive to the number of bands in the input data. The prior work trained on hyperspectral PRISM data which has 92 bands, while we train on multispectral WorldView-3 and Landsat-8 data which have 8 and 7 bands respectively. As such, our network is drastically smaller than what was previously used. This in turn limits the variability with which our model can represent mixtures and predict on the dataset, simply due to the reduction in size.

From the ergodic planner application, we demonstrate that intelligently sampling locations for *in situ* expeditions is an effective way to improve predictions with efficient resource usage. We find that having a prior slightly biased towards a class allows the trajectory to exploit regions of high uncertainty for that class and attain higher prediction performance on the class as a result. In contrast, care must be taken that the weighted prior does not cause regions of high uncertainty to be too dispersed, otherwise the planner will prioritize exploration without consideration of the biased class. We determine that the DCDM benefited the most from being updated with intelligent sampling, generally outperforming the Regression model across all metrics.

6.1 Project Contributions

The contributions of this thesis are summarized as follows:

- Trained a pipeline to successfully map coral reef benthic cover on global scale data with nearly 70% accuracy
- Demonstrated an increase in spatial resolution yields positive impacts in classifying spectral data
- Demonstrated that the Deep Conditional Dirichlet Model is able to successfully unmix coral pixels on higher spatial resolution data

- Implemented a sparse ergodic planner to help inform future *in situ* expeditions for mapping coral reefs based on the model output
- Demonstrated that weighted priors are generally able to improve performance on the class of interest

6.2 Future Work

Currently, the progress and conclusions we are able to make in this thesis are limited by our dataset. The labeled data we are provided and train on consists of 56.93% algae, 42.55% sand, and just 0.5296% of the most significant class, coral. In addition, we only work with the single scene of Heron Island in this thesis. Due to these limitations and the inherent bias in the labels, we are only able to make conclusions that suggest the performance of different datasets and models without statistical significance guarantees. This is especially apparent with regards to the DCDM when classifying coral. Improvements in performance were observed when trained with higher resolution WorldView data, while decreases in performance were observed with lower resolution Landsat data across the same set. As such, future work should look to collect and label more data for which the machine learning approaches explored in this thesis can be applied and evaluated in order to achieve statistically significant conclusions.

The Regression model and the DCDM both optimize a standard KLD loss function alongside an NLL loss function for the DCDM. Due to the imbalance of the training set, one possible modification to the model training would be to develop a weighted loss function to optimize. This would allow the model to penalize incorrect classifications of significant classes more heavily in comparison to others in order to better classify and unmix those classes after training.

The Regression model and the DCDM architectures are currently determined by the number of bands in the input data. As we provide multispectral inputs to the models, the architectures themselves are quite small, thus limiting the range of functions the network is able to model. The size also makes it difficult to apply approaches such as dropout layers, which can help improve the generalizability of the model while preventing overfitting. Further work could seek to determine how

wider and/or deeper architectures are able to better model the ambiguities in the spectral signature of the benthic classes. An optimal architecture size could then be determined for the purposes of benthic classification. Care would need to be taken in order to ensure that these larger networks do not overfit to the dataset.

Additionally, as the models trained in this thesis take in pixel inputs, spatial information about the input data is lost. Convolutional Neural Networks (CNN) train over regions of input pixels and thus are able to learn not only spectral information, but take into account spatial relations as well. This is especially significant for the purposes of coral reef mapping, as neighboring pixels are likely to contain continuous mixture distributions. As significant portions of the current dataset are masked out due to radiative transfer effects in deep water pixels, more continuously labeled pixel regions should be obtained in order to train the model. Otherwise, the masked pixels should be appropriately handled during data preprocessing and training.

Once learned models are obtained, whether from our work or from future developments, applying them to the entirety of the WorldView-3 archive of reef imagery would result in global quantification of coral reefs. From these predicted maps, the scenes with the highest model uncertainties that could benefit the most from *in situ* sampling are determined. The sparse ergodic planner then yields trajectories for *in situ* expeditions to follow and sample from. This would not only improve the coral maps, but would also allow for practical verification of our simulated method of informative path planning for model updating.

Finally, while the approaches discussed in this thesis were primarily developed and trained for application towards coral reef monitoring and mapping, they are generally applicable to any spectral dataset. Future work could seek to apply these models and techniques to different habitat scenes for classification and unmixing of different environments. Comparisons could then be made that explore the effectiveness of these approaches to different habitat scenes or based on the desired output classes.

Appendix A

Appendix

A.1 Hyperparameters

Several different hyperparameters were tuned over the course of training and testing the Regression model and the DCDM. A description of each hyperparameter as well as their possible values (if the hyperparameter values were discrete) is summarized in Table A.1.

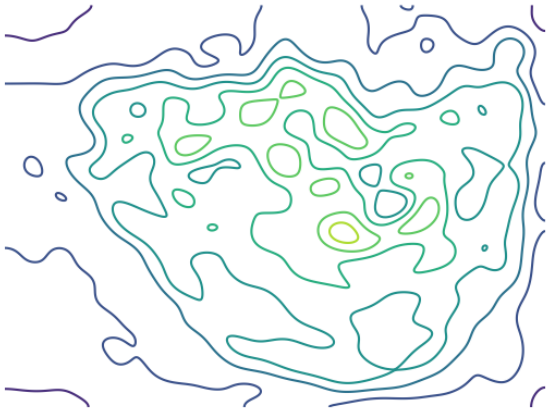
A.2 Ergodic Planning Trajectories

The visualization of the priors, Fourier bases, output trajectories, and final information maps of the agents across all priors for the Regression model and the DCDM are included below.

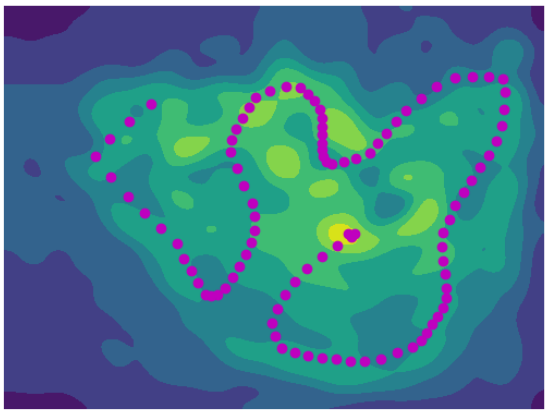
Hyperparameter	Description
Normalization	The normalization applied to the input data such that all inputs are of the same magnitude. [Max, L2]
Clustering Method	The clustering method used on the input training data. This sought to mitigate the effect of bias in the data by obtaining a data subset that would be more representative of all of the input classes. [Kmeans++, Maximum Balancing and Kmeans++]
Activation Function	The activation function applied to the hidden layers of the neural networks. These are important as they introduce non linearity into the system, allowing the network to model more complex representations and functions. [ReLU, SELU]
Learning Rate	The rate at which the model weights are adjusted during training. A higher learning rate will adjust the model weights faster but may overshoot local and global optima, while a lower learning rate will reach these optima but may take a long time to adapt the model weights. This value is modified by the Adam Optimizer over the course of the model training.
Lambda	The scalar value applied to the concentration parameters α of the Dirichlet distribution as part of the modified sigmoid function in the DCDM. A greater lambda value increases the overall magnitude of the α vector, meaning that the distribution has smaller variance.
Loss Weights	The weighting of the KLD and NLL loss functions in the DCDM. The sum of the two loss weights equals 1, and the sum of the weighted loss functions serves as the loss to minimize while training the model.

Table A.1: Hyperparameter Descriptions and Values

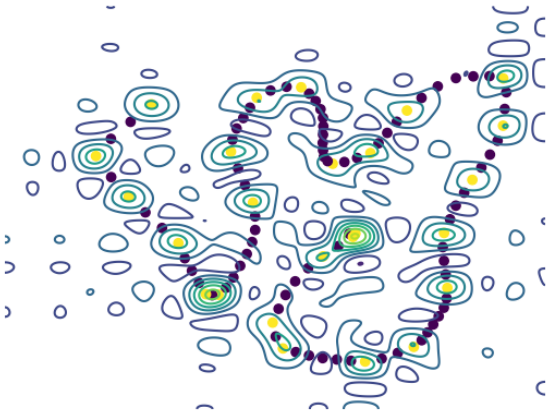
(a) The weighted prior found by taking a weighted sum of the class variances from the four model outputs.



(b) The weighted prior after a Fourier basis is applied for smoothing. Warmer colors have higher weights.



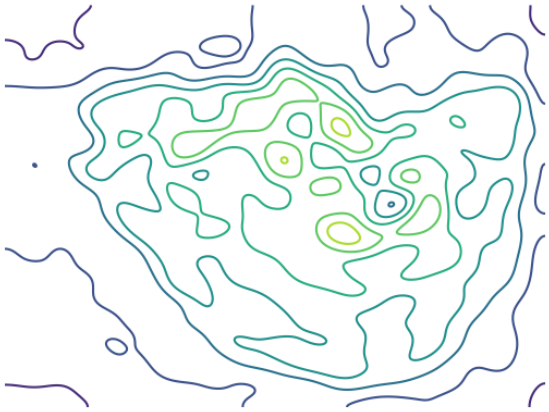
(c) The output trajectory from the planner.



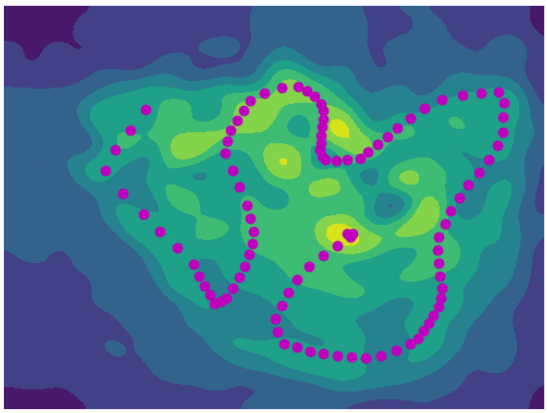
(d) The final information map of the agent from the trajectory. Locations where sparse sensed measurements were taken are indicated in yellow.

Figure A.1: Outputs for the planner on the weighted prior from the Regression model

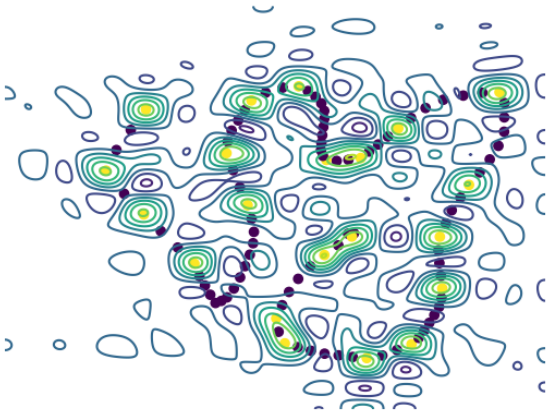
(a) The equally weighted prior from the four model outputs.



(b) The equally weighted prior after a Fourier basis is applied for smoothing. Warmer colors have higher weights.



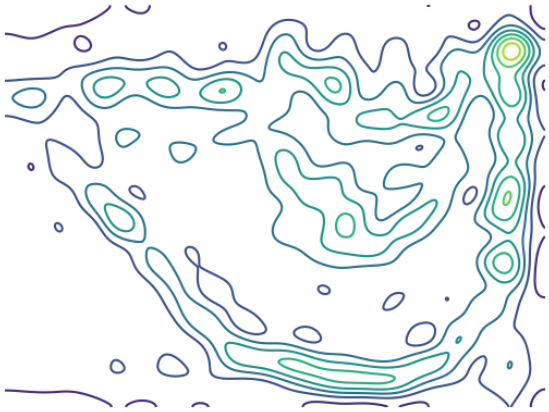
(c) The output trajectory from the planner.



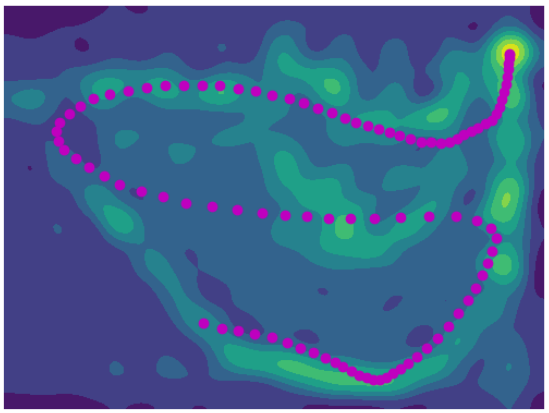
(d) The final information map of the agent from the trajectory. Locations where sparse sensed measurements were taken are indicated in yellow.

Figure A.2: Outputs for the planner on the equal prior from the Regression model

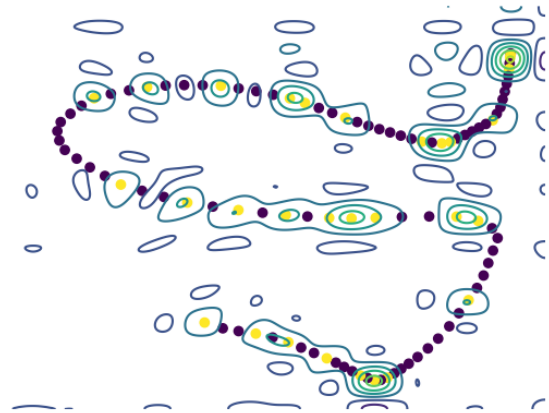
(a) The coral variance prior from the four model outputs.



(b) The coral variance prior after a Fourier basis is applied for smoothing. Warmer colors have higher weights.

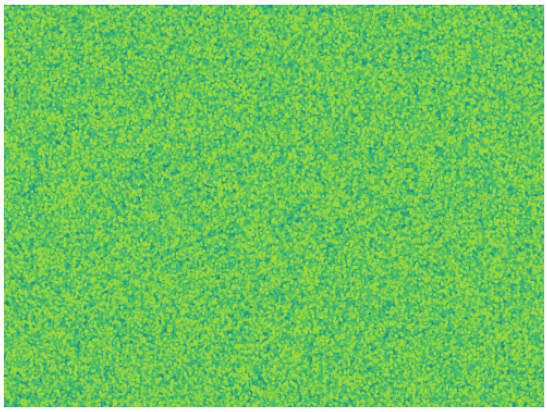


(c) The output trajectory from the planner.

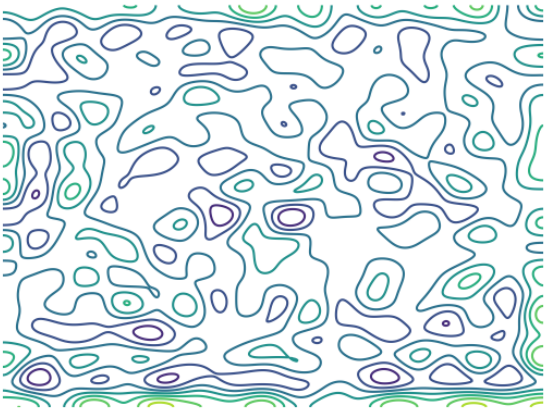


(d) The final information map of the agent from the trajectory. Locations where sparse sensed measurements were taken are indicated in yellow.

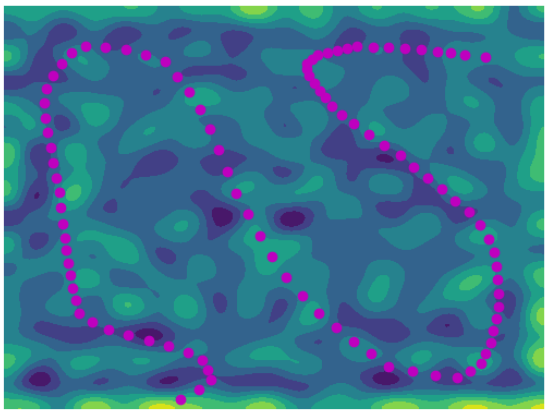
Figure A.3: Outputs for the planner on the coral prior from the Regression model



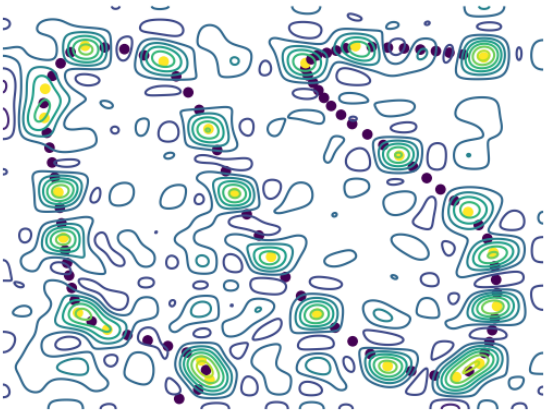
(a) The random prior.



(b) The random prior after a Fourier basis is applied for smoothing. Warmer colors have higher weights.

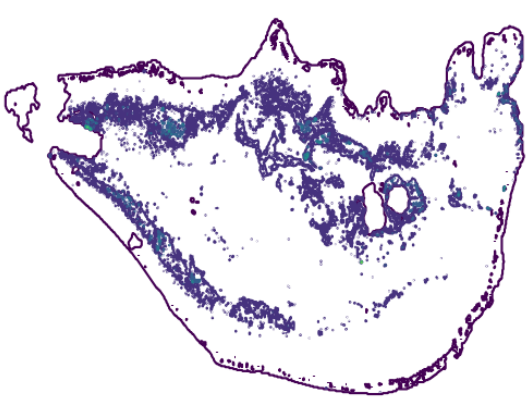


(c) The output trajectory from the planner.

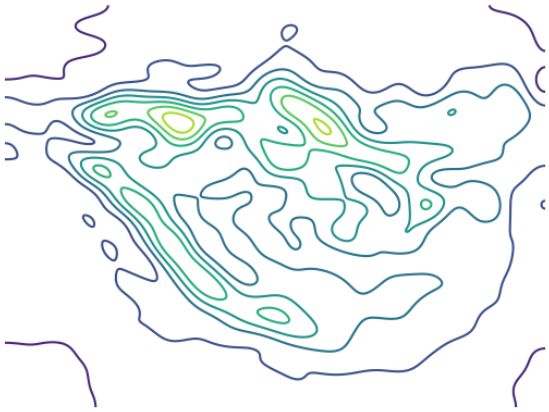


(d) The final information map of the agent from the trajectory. Locations where sparse sensed measurements were taken are indicated in yellow.

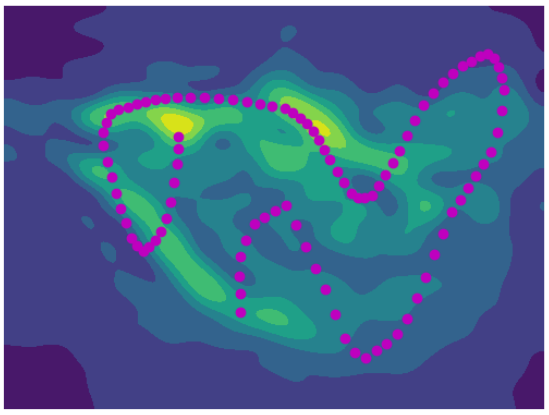
Figure A.4: Outputs for the planner on the random prior from the Regression model



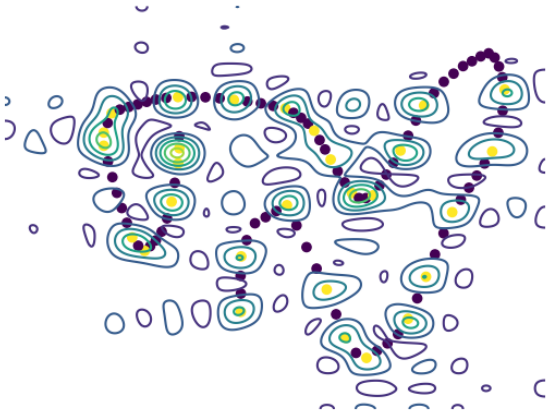
(a) The weighted prior found by taking a weighted sum of the class variances from the four model outputs.



(b) The weighted prior after a Fourier basis is applied for smoothing. Warmer colors have higher weights.

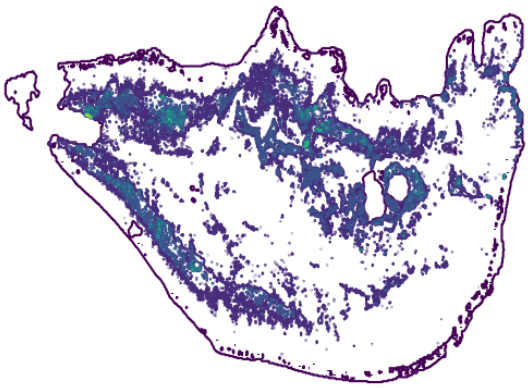


(c) The output trajectory from the planner.

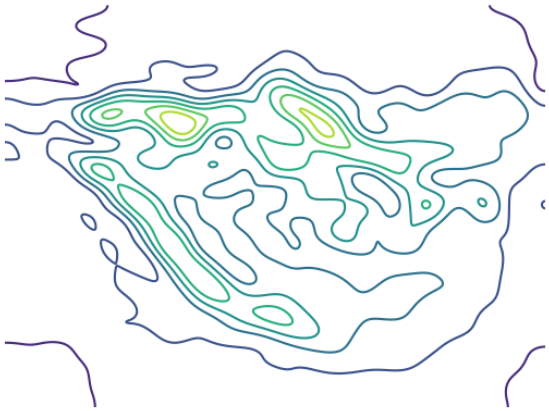


(d) The final information map of the agent from the trajectory. Locations where sparse sensed measurements were taken are indicated in yellow.

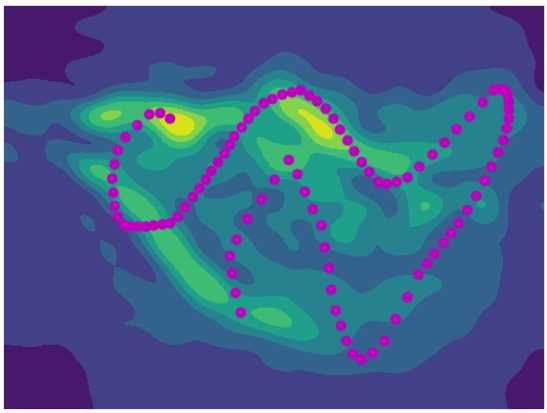
Figure A.5: Outputs for the planner on the weighted prior from the DCDM



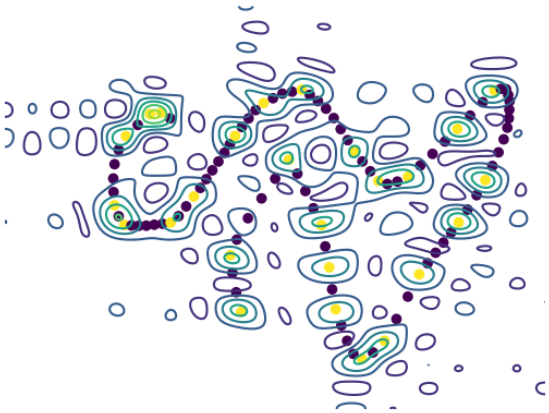
(a) The equally weighted prior from the four model outputs.



(b) The equally weighted prior after a Fourier basis is applied for smoothing. Warmer colors have higher weights.



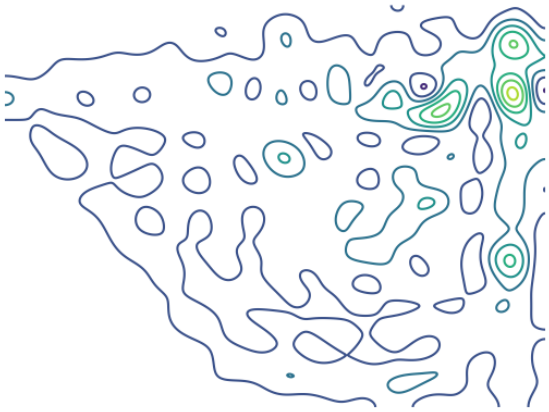
(c) The output trajectory from the planner.



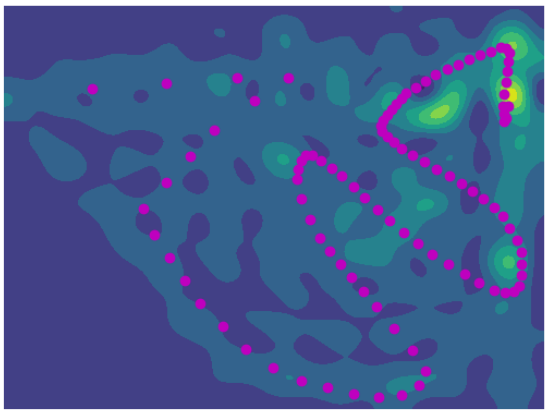
(d) The final information map of the agent from the trajectory. Locations where sparse sensed measurements were taken are indicated in yellow.

Figure A.6: Outputs for the planner on the equal prior from the DCDM

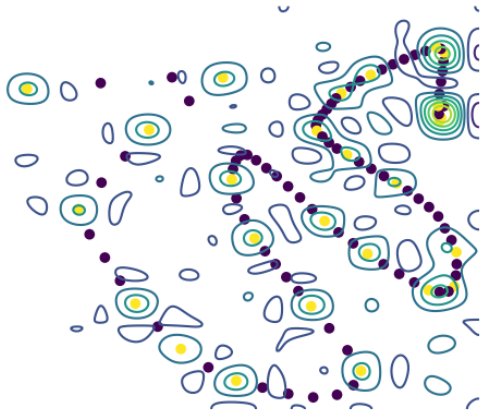
(a) The coral variance prior from the four model outputs.



(b) The coral variance prior after a Fourier basis is applied for smoothing. Warmer colors have higher weights.

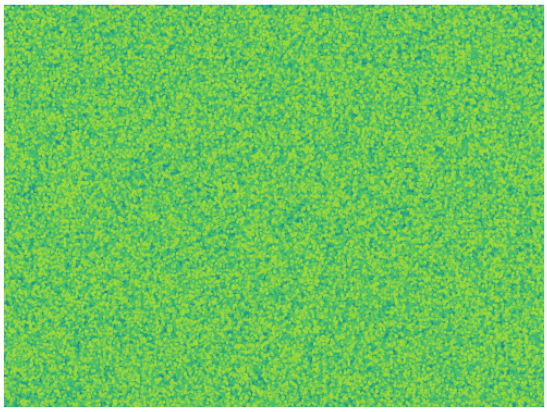


(c) The output trajectory from the planner.

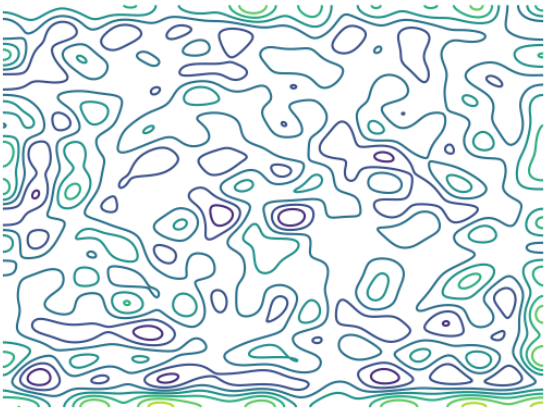


(d) The final information map of the agent from the trajectory. Locations where sparse sensed measurements were taken are indicated in yellow.

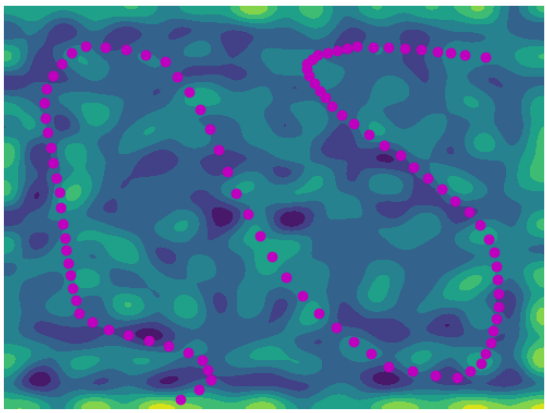
Figure A.7: Outputs for the planner on the coral prior from the DCDM



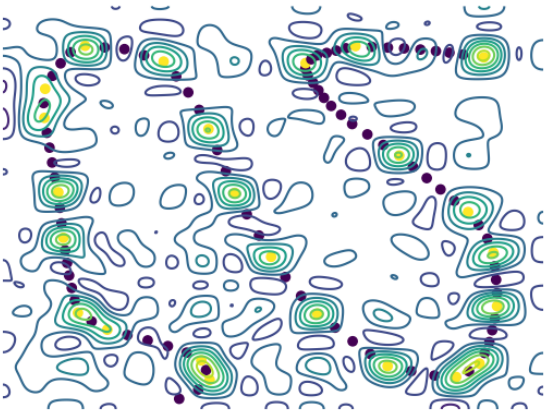
(a) The random prior.



(b) The random prior after a Fourier basis is applied for smoothing. Warmer colors have higher weights.



(c) The output trajectory from the planner.



(d) The final information map of the agent from the trajectory. Locations where sparse sensed measurements were taken are indicated in yellow.

Figure A.8: Outputs for the planner on the random prior from the DCDM

Bibliography

Serge Andréfouët, Frank Muller-Karger, Julie Robinson, Christine Kranenburg, Damaris Torres-Pulliza, Steven Spraggins, and Brock Murch. Global assessment of modern coral reef extent and diversity for regional science and management applications: a view from space. volume 2, 06 2004. URL https://www.researchgate.net/publication/271384472. 1.2

David Arthur and Sergei Vassilvitskii. K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, page 1027–1035, USA, 2007. Society for Industrial and Applied Mathematics. ISBN 9780898716245. 4.1.3

Roger Beeden, Jeffrey Maynard, Marjetta Puotinen, Paul Marshall, Jen Dryden, Jeremy Goldberg, and Gareth Williams. Impacts and recovery from severe tropical cyclone yasi on the great barrier reef. *PLOS ONE*, 10:e0121272, 04 2015. doi: 10.1371/journal.pone.0121272. 1.3

Jonathan Binney and Gaurav S. Sukhatme. Branch and bound for informative path planning. pages 2147–2154. Institute of Electrical and Electronics Engineers Inc., 2012. ISBN 9781467314039. doi: 10.1109/ICRA.2012.6224902. 3.1

Jonathan Binney, Andreas Krause, and Gaurav S. Sukhatme. Informative path planning for an autonomous underwater vehicle. In *2010 IEEE International Conference on Robotics and Automation*, pages 4791–4796, 2010. doi: 10.1109/ROBOT.2010.5509714. 3.1

Jeremy Broderson. Multispectral lighting: A practical option for difficult industrial imaging situations. 2020. URL https://www.photonics.com/Articles/Multispectral_Lighting_A_Practical_Option_for/a66251. 2.1

Cameron B. Browne, Edward Powley, Daniel Whitehouse, Simon M. Lucas, Peter I. Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(1):1–43, 2012. doi: 10.1109/TCIAIG.2012.2186810. 3.1

Alberto Candela, Suhit Kodgule, Kevin Edelson, Srinivasan Vijayarangan, David R.

Thompson, Eldar Noe Dobrea, and David Wettergreen. Planetary rover exploration combining remote and in situ measurements for active spectroscopic mapping. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5986–5993, 2020. doi: 10.1109/ICRA40945.2020.9196973. 3.1

Alberto Candela, Kevin Edelson, Michelle M. Gierach, David R. Thompson, Gail Woodward, and David Wettergreen. Using remote sensing and in situ measurements for efficient mapping and optimal sampling of coral reefs. *Frontiers in Marine Science*, 8, 2021. ISSN 2296-7745. doi: 10.3389/fmars.2021.689489. URL https://www.frontiersin.org/article/10.3389/fmars.2021.689489. 1.2, 1.3, 6

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Mach. Learn.*, 20 (3):273–297, sep 1995. ISSN 0885-6125. doi: 10.1023/A:1022627411411. URL https://doi.org/10.1023/A:1022627411411. 4.2

Shreeja Dahal, Robert Schaeffer, and Eman Abdelfattah. Performance of different classification models on national coral reef monitoring dataset. In *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0662–0666, 2021. doi: 10.1109/CCWC51732.2021.9376135. 1.2, 2.2

Louis Dressel and Mykel J. Kochenderfer. On the optimality of ergodic trajectories for information gathering tasks. In *2018 Annual American Control Conference (ACC)*, pages 1855–1861, 2018. doi: 10.23919/ACC.2018.8430857. 3.2

Kevin Edelson. Ergodic trajectory optimization for information gathering. Master's thesis, Carnegie Mellon University, Pittsburgh, PA, October 2020. 1.2, 3.1, 3.2

Joanne Edkins. Unprecedented global image of coral reefs from allen coral atlas released. 2022 (Accessed April 6, 2023). 1.4

Peter Edmunds and John Bruno. The importance of sampling scale in ecology: Kilometer-wide variation in coral reef communities. *Marine Ecology Progress Series*, 143:165–171, 11 1996. doi: 10.3354/meps143165. 1.2

Alberto Candela Garza. *Bayesian Models for Science-Driven Robotic Exploration*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, September 2021. 1.2, 1.3, 2.2, 2.2, 2.2, 4.2, 6

Shivam Gautam, Bishwamoy Sinha Roy, Alberto Candela, and David Wettergreen. Science-aware exploration using entropy-based planning. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3819–3825, 2017. doi: 10.1109/IROS.2017.8206232. 3.2

Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler, and Xiao Xiang Zhu. A survey of uncertainty in deep neural networks, 2022. 4.3

GISGeography. Multispectral vs hyperspectral imagery explained. 2022 (Accessed March 20, 2023). 2.1

James Goodman and Susan Ustin. Classification of benthic composition in a coral reef environment using spectral unmixing. *Journal of Applied Remote Sensing*, 1, 12 2007. doi: 10.1117/1.2815907. 1.2

John D. Hedley, Chris M. Roelfsema, Iliana Chollett, Alastair R. Harborne, Scott F. Heron, Scarla Weeks, William J. Skirving, Alan E. Strong, C. Mark Eakin, Tyler R. L. Christensen, Victor Ticzon, Sonia Bejarano, and Peter J. Mumby. Remote sensing of coral reefs for monitoring and management: A review. *Remote Sensing*, 8(2), 2016. ISSN 2072-4292. doi: 10.3390/rs8020118. URL https://www.mdpi.com/2072-4292/8/2/118. 1.2

John D. Hedley, Chris Roelfsema, Vittorio Brando, Claudia Giardino, Tiit Kutser, Stuart Phinn, Peter J. Mumby, Omar Barrilero, Jean Laporte, and Benjamin Koetz. Coral reef applications of sentinel-2: Coverage, characteristics, bathymetry and benthic mapping with comparison to landsat 8. *Remote Sensing of Environment*, 216:598–614, 10 2018. ISSN 00344257. doi: 10.1016/j.rse.2018.07.014. 1.2

Eric Hochberg and M. Atkinson. Spectral discrimination of coral reef benthic communities. *Coral Reefs*, 19:164–171, 07 2000. doi: 10.1007/s003380000087. 1.2

Eric J. Hochberg and Michelle M. Gierach. Airborne imaging spectroscopy of coral reef condition. *AGU Fall Meeting Abstracts*, 2018:GC11A–07, December 2018. 4.1.1

Eric J. Hochberg and Michelle M. Gierach. Missing the reef for the corals: Unexpected trends between coral reef condition and the environment at the ecosystem scale. *Frontiers in Marine Science*, 8, 8 2021. ISSN 22967745. doi: 10.3389/fmars.2021.727038. 1.1

Eric J. Hochberg, Marlin J. Atkinson, and Serge Andréfouët. Spectral reflectance of coral reef bottom-types worldwide and implications for coral reef remote sensing. *Remote Sensing of Environment*, 85(2):159–173, 2003. ISSN 0034-4257. doi: https://doi.org/10.1016/S0034-4257(02)00201-8. URL https://www.sciencedirect.com/science/article/pii/S0034425702002018. 1.2

Ove Hoegh-Guldberg, Elvira S. Poloczanska, William Skirving, and Sophie Dove. Coral reef ecosystems under climate change and ocean acidification. *Frontiers in Marine Science*, 4, 5 2017. ISSN 22967745. doi: 10.3389/fmars.2017.00158. 1.1

Geoffrey A. Hollinger and Gaurav S. Sukhatme. Sampling-based robotic information gathering algorithms. *The International Journal of Robotics Research*, 33(9):1271–1287, 2014. doi: 10.1177/0278364914533443. URL https://doi.org/10.1177/0278364914533443. 3.1

ISRS. Isrs consensus statement on climate change and coral bleaching, conf. of the parties to the united nations framework convention on climate change. December

2015. 1.1

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4.4, 5.1, 5.2

Suhit Kodgule. Active sampling for planetary rover exploration. Master's thesis, Carnegie Mellon University, Pittsburgh, PA, August 2019. 3.1

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951. ISSN 00034851. URL http://www.jstor.org/stable/2236703. 2.2

Zhongping Lee and Kendall Carder. Effect of spectral band numbers on the retrieval of water column and bottom properties from ocean color data. *Applied optics*, 41: 2191–201, 05 2002. doi: 10.1364/AO.41.002191. 4.1.1

Jiwei Li, David E. Knapp, Nicholas S. Fabina, Emma V. Kennedy, Kirk Larsen, Mitchell B. Lyons, Nicholas J. Murray, Stuart R. Phinn, Chris M. Roelfsema, and Gregory P. Asner. A global coral reef probability map generated using convolutional neural networks. *Coral Reefs*, 39(6):1805–1815, December 2020. ISSN 0722-4028. doi: 10.1007/s00338-020-02005-6. URL https://doi.org/10.1007/s00338-020-02005-6. 1.2

Mitchell B. Lyons, Chris M. Roelfsema, Emma V. Kennedy, Eva M. Kovacs, Rodney Borrego-Acevedo, Kathryn Markey, Meredith Roe, Doddy M. Yuwono, Daniel L. Harris, Stuart R. Phinn, Gregory P. Asner, Jiwei Li, David E. Knapp, Nicholas S. Fabina, Kirk Larsen, Dimosthenis Traganos, and Nicholas J. Murray. Mapping the world's coral reefs using a global multiscale earth observation framework. *Remote Sensing in Ecology and Conservation*, 6:557–568, 12 2020. ISSN 20563485. doi: 10.1002/rse2.157. 1.2

Stéphane Maritorena, André Morel, and Bernard Gentili. Diffuse reflectance of oceanic shallow waters: Influence of water depth and bottom albedo. *Limnology and Oceanography*, 39(7):1689–1703, 1994. doi: https://doi.org/10.4319/lo.1994.39.7.1689. URL https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.4319/lo.1994.39.7.1689. 4.1.1

George Mathew and Igor Mezić. Metrics for ergodicity and design of ergodic dynamics for multi-agent systems. *Physica D: Nonlinear Phenomena*, 240(4):432–442, 2011. ISSN 0167-2789. doi: https://doi.org/10.1016/j.physd.2010.10.010. URL https://www.sciencedirect.com/science/article/pii/S016727891000285X. 3.2, 3.2, 4.4

Ian Miller, Michelle Jonker, and Germiithya Coleman. *Crown-of-thorns starfish and coral surveys using the manta tow technique*. Australian Institute of Marine Science, Townsville, Australia, 2018. 1.1, 1.2

Lauren M. Miller and Todd D. Murphey. Trajectory optimization for continuous

ergodic exploration. In *2013 American Control Conference*, pages 4196–4201, 2013. doi: 10.1109/ACC.2013.6580484. 3.2

Fredrik Moberg and Carl Folke. Ecological goods and services of coral reef ecosystems. *Ecological Economics*, 29(2):215–233, 1999. ISSN 0921-8009. doi: https://doi.org/10.1016/S0921-8009(99)00009-9. URL https://www.sciencedirect.com/science/article/pii/S0921800999000099. 1.1

Carlo Orrieri, Gianmario Tessitore, and Petr Veverka. Ergodic maximum principle for stochastic systems. *Applied Mathematics and Optimization*, 79:567–591, 6 2019. ISSN 14320606. doi: 10.1007/s00245-017-9448-7. 3.2

Sooho Park. Learning for informative path planning. 2008. URL https://dspace.mit.edu/handle/1721.1/45887. 3.1

Ananya Rao, Ian Abraham, Guillaume Sartoretti, and Howie Choset. Sparse sensing in ergodic optimization. 2022. URL https://www.marmotlab.org/publications/43-DARS2022-SparseErgodic.pdf. 3.2, 3.3, 3.3, 4.4, 5.2

Marjorie Reaka. The global biodiversity of coral reefs: a comparison with rainforests. 01 1997. URL https://www.researchgate.net/publication/239063261/. 1.1

Chris M. Roelfsema, Eva M. Kovacs, Juan Carlos Ortiz, David P. Callaghan, Karlo Hock, Mathieu Mongin, Kasper Johansen, Peter J. Mumby, Magnus Wettle, Mike Ronan, Petra Lundgren, Emma V. Kennedy, and Stuart R. Phinn. Habitat maps to enhance monitoring and management of the great barrier reef. *Coral Reefs*, 39:1039–1054, 2020. doi: https://doi.org/10.1007/s00338-020-01929-3. URL https://link.springer.com/article/10.1007/s00338-020-01929-3. 1.2

David Thompson, Eric Hochberg, Gregory Asner, Robert Green, David Knapp, Bo-Cai Gao, Rodrigo Garcia, Michelle Gierach, Zhongping Lee, Stephane Maritorena, and Ronald Fick. Airborne mapping of benthic reflectance spectra with bayesian linear mixtures. *Remote Sensing of Environment*, 200:18–30, 10 2017. doi: 10.1016/j.rse.2017.07.030. 4.1, 4.1.1

UNEP-WCMC, WorldFish Centre, WRI, and TNC. Global distribution of warm-water coral reefs, compiled from multiple sources including the Millennium Coral Reef Mapping Project. Version 4.1. Includes contributions from IMaRS-USF and IRD (2005), IMaRS-USF (2005) and Spalding et al. (2001). 2021. doi: https://doi.org/10.34892/t2wk-5t34. URL https://data.unep-wcmc.org/datasets/1. 1.2

Clive R Wilkinson. *Status of Coral Reefs of the World*. Global Coral Reef Monitoring Network and Reef and Rainforest Research Centre, Townsville, Australia, 2008. 1.1

Jingping Xu, Jianhua Zhao, Fei Wang, Yanlong Chen, and Zhongping Lee. Detection of coral reef bleaching based on sentinel-2 multi-temporal imagery: Simulation and case study. *Frontiers in Marine Science*, 8, 2021. ISSN 2296-7745. doi:

10.3389/fmars.2021.584263. URL https://www.frontiersin.org/articles/10.3389/fmars.2021.584263. 1.2