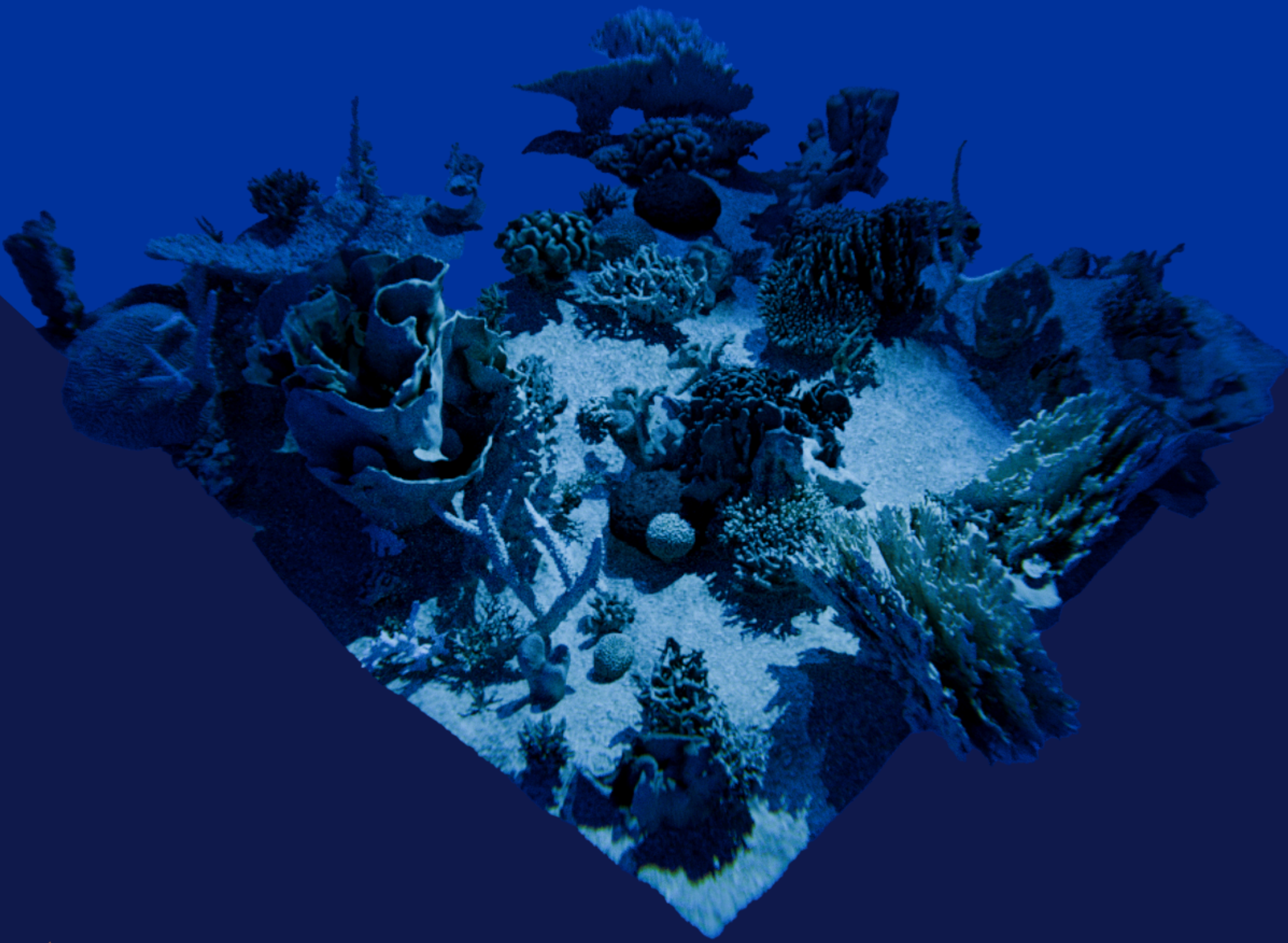
MSc thesis in Geomatics for the Built Environment

Automated Data-Driven Generation of 3D Coral Reef Models: Assessing and Integrating Empirical Data Sources

Gees Brouwer

2024



TU Delft

On cover:
A section of a coral reef generated as part of this thesis, visualized in Blender with added lighting and textures (Image by author)

Automated Data-Driven Generation of 3D Coral Reef Models: Assessing and Integrating Empirical Data Sources

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of

Master of Science in Geomatics for the Built Environment

by

Gees Daniël Brouwer

October 2024

Gees Brouwer: Automated Data-Driven Generation of 3D Coral Reef Models: Assessing and Integrating Empirical Data Sources (2024)

This work is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

The work in this thesis was made in the:

3D geoinformation group
Department of Urbanism
Faculty of Architecture & the Built Environment Delft University of Technology

Supervisors:

Dr. Hugo Ledoux

Dr. Akshay Patil

Co-reader:

Ir. Edward Verbree

ABSTRACT

While three-dimensional coral reef models are valuable for various applications, existing approaches like photogrammetric scanning and manual modeling require substantial time and expertise, limiting their scalability. Previous algorithmic approaches, particularly Agent-Based Models (ABM), have relied heavily on complex ecological simulations and deep domain knowledge. This thesis explores an alternative data-driven approach to automated coral reef modeling, investigating whether empirical data sources can provide a scalable method for generating ecologically plausible 3D models. Rather than simulating ecological processes from first principles, we develop a pipeline that leverages observational data to inform and constrain procedural generation techniques. Through systematic evaluation of available data sources, including the Global Biodiversity Information Facility (GBIF), CoralNet, the Allen Coral Atlas, the Coral Traits Database and the Smithsonian Institution's 3D coral collection, we identified both opportunities and significant limitations in current data availability. The research developed a modular pipeline implemented in Blender that combines procedural terrain generation with the placement of 3D coral models, integrating species occurrence data aggregated over geomorphic zones. To ensure robust data integration across sources and maintain compatibility with evolving taxonomic standards, the pipeline implements automated species name verification through the World Register of Marine Species (WoRMS) (WoRMS - World Register Of Marine Species, n.d.) API. While the resulting pipeline successfully establishes a foundation for automated coral reef modeling, limitations in available structural data necessitated the use of manually configured parameters for critical aspects such as terrain characteristics and population density. The pipeline's modular structure, standardized taxonomy handling, and integration with standardized classification systems position it well for future iterations as improved data sources become available. This research demonstrates the potential of data-driven approaches to coral reef modeling while highlighting the need for more comprehensive, fine-scale structural data to enable fully automated, ecologically plausible modeling of coral reef environments.

ACKNOWLEDGEMENTS

I would like to thank you, the reader, for being part of my P5. Of course I would like to thank many more people, like my loving parents, girlfriend, geomates and friends but I think they deserve an extensive acknowledgement that is written by a happier version of me.

CONTENTS

1 INTRODUCTION

1.1 Background

Coral reefs are among the most diverse and ecologically significant ecosystems on Earth (Fisher et al., 2015), often referred to as the "cities of the sea" (Wicks, 2016) due to their complex structures and the vast number of species they support. Despite covering less than 0.1% of the ocean floor, coral reefs provide habitat for approximately 25% of all marine species (Spalding et al., 2001). Corals, which are actually colonies of small animals called polyps, form the foundation of these ecosystems by secreting calcium carbonate to create a hard skeleton. Different coral species grow into many different shapes—such as branching or massive growth forms, as shown in **Figure 1.1**—each contributing to the overall structure of the reef. This diversity in shape and structure adds to the habitat complexity of the reef and helps support a wide range of marine life.

Figure 1.1: A grooved brain coral (*Diploria labyrinthiformis*), which displays a massive growth form, and a branching coral, possibly *Pseudoplexaura*, at the wreck of the Benwood in the Florida Keys. Image is in the public domain (CC0).

However, these crucial ecosystems face unprecedented threats, primarily due to climate change (Hughes et al., 2017). As global temperatures rise and ocean chemistry alters, coral reefs worldwide are experiencing increased stress, leading to phenomena such as coral bleaching (Hoegh-Guldberg 1999). In light of these challenges, the field of coral reef conservation has become increasingly important, with recent technological advancements playing a significant role in our efforts to understand, monitor, and protect these delicate ecosystems.

1.2 Technological Advancements in Coral Reef Research

Recent technological progress has enhanced coral reef research and conservation efforts by improving both data collection and data analysis capabilities. High-resolution underwater imagery allows researchers to assess biodiversity and monitor the health of extensive reef sections without the need for continuous in-situ presence, thereby minimizing human interference. Orthorectified photomosaics provide comprehensive overviews of reef areas, facilitating the identification of species distribution, colony health, and structural changes over time, with computer vision techniques showing increasing potential to fully automate these processes, as further discussed in Section 2.3. Additionally, photogrammetry techniques allow high-resolution underwater imagery to be transformed into detailed 3D models of coral reefs, as discussed in Section 2.1. These 3D models capture the structural complexity of coral reefs, an important aspect that enables more advanced research, such as simulations and structural analyses.

As technological advancements in data collection and processing continue to evolve, the potential for algorithmic generation of 3D coral reef models has grown, offering new avenues for research that complement traditional scanning or manual modeling methods, as further described in the next Section.

1.3 3D Modeling of Coral Reefs

The creation of 3D models typically follows one of three approaches: scanning (e.g., using photogrammetry techniques), manual modeling, or algorithmic generation. While scanning and manual modeling can produce highly accurate representations, they often require substantial time, effort, and expertise. Recent technological advancements have streamlined these processes, but they still present significant challenges, especially when large-scale or frequent modeling is required. Manual modeling, in particular, demands extensive user input to craft each model, making it less feasible for generating numerous diverse coral reef representations.

Algorithmic or procedural modeling emerges as a promising alternative, offering the potential for rapid generation of coral reef models with minimal user input. This approach is particularly appealing for studies such as Computational Fluid Dynamics (CFD) (Patil et al., 2024), where the quantity and flexibility of 3D data often outweigh the need for exact replication of real-world locations. However, an important question arises: to what extent can algorithmically produced 3D models of coral reefs be ecologically plausible? Coral reefs are shaped by a complex interplay of environmental, biological, and geological factors, and capturing this complexity in an algorithm presents a significant challenge.

Previous work in this domain has explored the use of Agent-Based Models (ABM) to simulate virtual 3D coral reefs (Kubicek et al., 2012). As further discussed in Section 3.1, these approaches typically require extensive ecological knowledge of the dynamic processes that shape reef ecosystems, which must then be simplified and encoded into efficient simulation algorithms. While valuable, such methods demand a deep understanding of coral reef ecology, falling outside of the field in which this thesis is situated: the realm of geomatics.

The limitations of these approaches, particularly their reliance on deep ecological knowledge and complex simulations, suggest an opportunity to explore alternative methods. This thesis investigates whether empirical data sources can provide a scalable, automated approach to generating 3D coral reef models that maintain ecological plausibility. As technological advancements have led to more efficient data collection methods, we can explore whether information can be extracted from these sources in an automated process to inform the generation of realistic 3D coral reef models. Rather than attempting to simulate complex ecological processes from first principles, we explore how existing observational data can be leveraged to inform and constrain our models. For instance, instead of modeling the habitat preferences of different coral species based on theoretical knowledge, we can analyze co-occurrence patterns in empirical datasets to infer realistic species distributions.

This data-driven approach draws inspiration from procedural generation techniques, which use minimal input to generate complex structures based on algorithmic rules, as seen in urban modeling applications discussed in the Related Work section. By establishing adaptable rules derived from empirical data, we aim to create ecologically plausible coral reef models that avoid the need for manual modeling or deep ecological simulations. The specific parameters guiding this procedural modeling will emerge from the empirical data examined in this thesis, with insights into species distributions, colony densities, and spatial patterns shaping the generated structures. This data-driven framework will support scalable, automated coral reef modeling, providing realistic and varied representations of reef environments that reflect the diversity and complexity observed in natural coral reefs.

1.4 Research Questions

The main research question of this thesis is: *How can identified empirical data sources be assessed and integrated into a pipeline for the automated, data-driven generation of ecologically plausible 3D models of coral reefs, given the current state of available data and technology?* To achieve this, we will address the following subquestions:

1. What empirical data sources can be identified as relevant for informing coral reef modeling, and how do they compare in terms of data quantity and potential for scalability?
2. How can relevant information be extracted and processed from the identified data sources to enhance model realism, and what challenges or limitations are encountered during this process?
3. How can the processed data from these sources be integrated into a coherent pipeline for automated coral reef modeling?
4. Given the current state of data quantity, quality, and technology, what can be achieved in developing an automated, data-driven pipeline for coral reef modeling?

1.5 Scope and Scale of the Study

This research focuses on modeling 10 by 10 meter sections of coral reefs, a scale initially chosen to align with the requirements for Computational Fluid Dynamics (CFD) simulations (Patil et al., 2024). This scale strikes an assumed balance between computational feasibility and the representation of reef sections large enough to capture broader patterns while retaining enough detail to model individual corals. Defining this scale early on also provides a concrete framework for assessing the relevance of empirical data sources, ensuring that the selected data is appropriate for modeling small reef sections in a computationally manageable way.

The spatial distribution of species, influenced by global environmental factors like temperature and depth, and more localized conditions like substrate and competition, plays a key role in determining species abundance within a specific area. Therefore, the selected scale enables us to assess species distributions and spatial patterns that are critical for ecologically plausible models of coral reefs.

1.6 Significance and Limitations

It is important to note that while our approach focuses on static, empirical data and does not directly model the dynamic, time-dependent nature of coral reef ecosystems, it offers several advantages. First, it aligns well with the field of geomatics, leveraging spatial data analysis and modeling techniques. Second, it provides a novel perspective on reef modeling that complements existing simulation-based approaches. Third, the information extracted from empirical data sources, even if simplified compared to established ecological research methods, can be valuable in its own right for understanding coral reef distributions and relationships.

Furthermore, we recognize that both the available data sources and our understanding of coral reef ecosystems are continually evolving. As such, our proposed pipeline is designed with future iterations in mind, allowing for the incorporation of improved data sources and more sophisticated modeling techniques as they become available.

While this approach may not directly simulate the temporal dynamics of coral reefs in response to environmental changes, it lays a foundation for future studies and simulations that could contribute to conservation efforts. The static models produced by our pipeline can serve as starting points for dynamic simulations or as baselines for monitoring changes over time. Additionally, the insights gained from analyzing and integrating diverse empirical data sources may inform conservation strategies and highlight areas where additional data collection or research is needed.

1.7 Thesis Outline

This thesis is organized as follows:

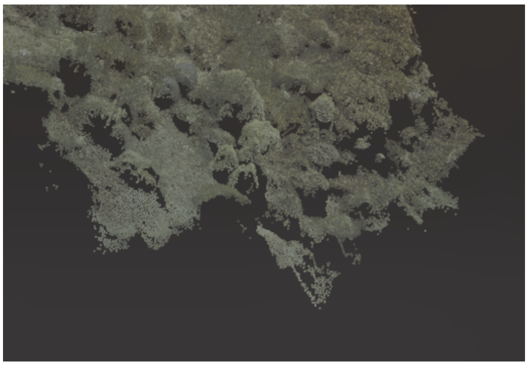
- **Chapter 2: Exploratory Work** and **Chapter 3: Related Work** explore potential empirical data sources for coral reef modeling. These chapters assess the availability, quality, and relevance of existing datasets, highlighting key limitations and identifying the most suitable sources for further analysis.
- **Chapter 4: Methodology** discusses the methods used to extract relevant information from the selected data sources and details the modeling techniques applied to generate coral reef structures, including procedural terrain generation and species distribution.
- **Chapter 5: Results** presents the findings from data extraction and analysis, providing insights into species occurrences and structural modeling. This chapter also describes the construction of the data-driven pipeline, integrating extracted information to generate ecologically informed coral reef models.
- **Chapter 6: Discussion and Conclusion** summarizes the main findings, evaluates the pipeline's limitations, and outlines recommendations for future research to enhance automated coral reef modeling and improve data integration.

2 EXPLORATORY WORK

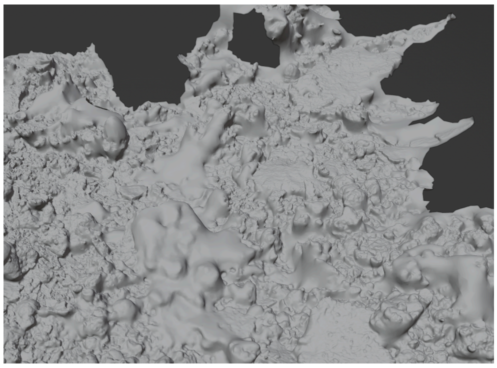
2.1 Initial Exploration with Processing 3D Scanned Coral Reef Models

While capturing high-resolution 3D models of reef sections has become technically straightforward, as detailed in Related Work, this capability has not translated into comprehensive data repositories. Instead, these models exist primarily as isolated instances or supplementary materials in research studies. A 3D model used in Lindhart et al. (2021), a 15x20 meter patch of the fringing reef off Ile Anglaise, was obtained, around 3GB in size. Additionally, a point cloud was requested and obtained from the Royal Netherlands Institute for Sea Research (NIOZ), representing a 3x3 meter patch of coral reef, on the leeward side of Curacao. The point cloud file was around 90GB in size.

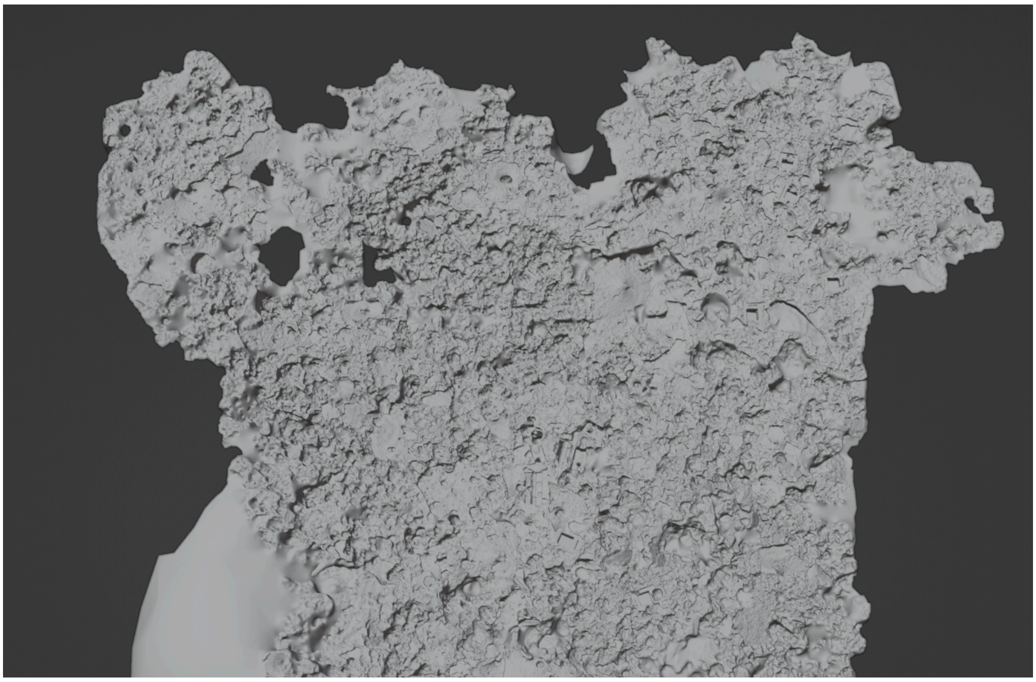
I used the Point Data Abstraction Library (PDAL) to thin the high-density point cloud dataset. A sampling filter was applied, where points within a specified radius of 0.01 units were reduced, effectively decreasing the overall density of the point cloud. The thinning process was performed using a PDAL pipeline configured in a JSON file, which outlined the steps: reading the input PLY file, applying the sampling filter, and saving the output to a new PLY file. Following this, MeshLab was used to perform Poisson surface reconstruction, which resulted in a 3D mesh of a more manageable file size of around 900MB. After manually inspecting both 3D models with Blender, scanning artifacts were observed for both models, with occluded parts being incorrectly merged with the terrain. With the human eye, there was no clear distinction between the spatial frequencies and amplitudes of terrain features and corals. Examples of these observations are shown in **Figure 2.1.** While this could be attributed to the methods used to process the point cloud into a mesh, the limited availability of such data and the time-consuming nature of processing led us to regard further processing efforts as outside the scope of this thesis. Related Work, however, discusses a method for the automated classification of corals in 3D reconstructions.



(a)



(b)



(c)

Figure 2.1: (a) The thinned point cloud showing occluded parts with missing data, probably the cause of (b) unexpected, smooth surfaces between parts of coral and terrain features in the model of Lindhart et al. (2021). (c) The bottom side of the model of Lindhart et al. (2021), demonstrating the complexities of effectively distinguishing terrain features from corals. (Images by author)

2.2 Initial Exploration with SAM for Coral Reef Segmentation

In the early stages of this thesis, as limitations in 3D data availability and its processing complexities became evident, as further discussed in Chapter 6, the use of image-based methods for coral reef analysis emerged as a promising alternative. Segmenting corals from images using the Segment Anything Model (SAM) (Kirillov et al., 2023) was particularly appealing due to its state-of-the-art capabilities in general image segmentation. The publicly available demo interface of SAM (segment-anything.com/demo) was utilized for this experiment. The "Find all the objects in the image automatically" option, which performs a query using a regular grid of foreground points, was applied to segment the images. Since the size of this grid influences both segmentation accuracy and computational load, additional experiments were conducted using the downloadable pre-trained model provided by the SAM developers (github.com/facebookresearch/segment-anything), which allowed for custom grid sizes and configurations. These experiments revealed that the same types of artifacts, as discussed in this chapter, occurred, regardless of the grid size used.

The initial exploration began with images from open-source repositories like Unsplash (unsplash.com) and WikiMedia Commons (commons.wikimedia.org), which offer free-to-use content under broad licenses. However, an inherent bias was observed in the coral reef images available on such platforms: they predominantly feature aesthetically optimized photographs, characterized by vibrant colors, high biodiversity, and carefully composed perspectives. While these images are visually striking, they may not represent typical coral reef conditions. When applying SAM to these high-quality images, two distinct outcomes emerged:

- For corals in the foreground, SAM demonstrated remarkable accuracy in generating segmentation masks, likely aided by the clear definition and vivid colors of the subjects.
- Conversely, when processing larger reef sections farther from the camera, SAM often incorrectly consolidated multiple distinct coral colonies into single segmentation masks.

The issue, of which an example is shown in **Figure 2.2,** stems from the fact that SAM aims to segment any object or region it detects, without distinguishing between foreground and background elements based on distance. In these non-orthographic images, the presence of water surfaces, reef sections further in the background affected by blue-shift, and darker areas made it more challenging to distinguish individual coral colonies. SAM, in these cases, grouped background sections of the reef as a single object, likely due to the visual effects of perspective and blue-shift.

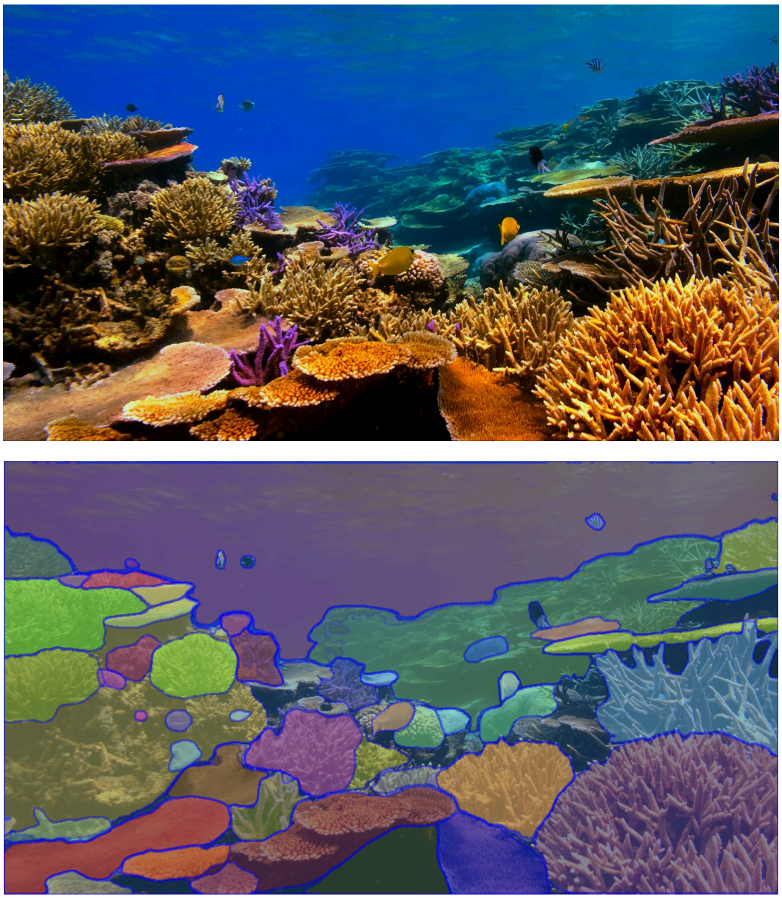


Figure 2.2: The original coral reef image (top) sourced from Unsplash (free to use under the Unsplash License), and its segmented version (bottom) processed using SAM (Segment Anything Model). The segmented image highlights how SAM also segments unintended regions, including the water surface, background reef sections affected by blue-shift, and darker areas, consolidating them into single masks.

Recognizing that public images might not fully represent coral reef environments, the experiment was expanded to include scientifically purposed images from sources like CoralNet (CoralNet, n.d.). These images, often used for training classifiers or monitoring coral reef sites, tend to be more orthographic in perspective and monotonous in color—reflecting realistic reef sections that might be covered with sand, algae, or dead coral skeletons. In these cases, SAM showed promise by distinguishing between different coral morphologies, such as branching and massive corals. However, certain artifacts still occurred, as SAM appears to respond more to texture patterns rather than fully capturing coral structure.

For example, as shown in **Figure 2.3**, SAM struggles in some cases to accurately segment coral colonies. In the bottom left of the segmented image, a section of coral connected to a part further away from the camera is not included in the segmentation mask. Similarly, the coral to the top left of that is incorrectly segmented as two separate corals due to a groove in its texture. The large tabular coral in the center demonstrates small limitations, with parts of its stem not being included in the segmentation mask. These examples suggest that SAM's segmentation is influenced more by consistent visible textures than by an understanding of the overall structure.

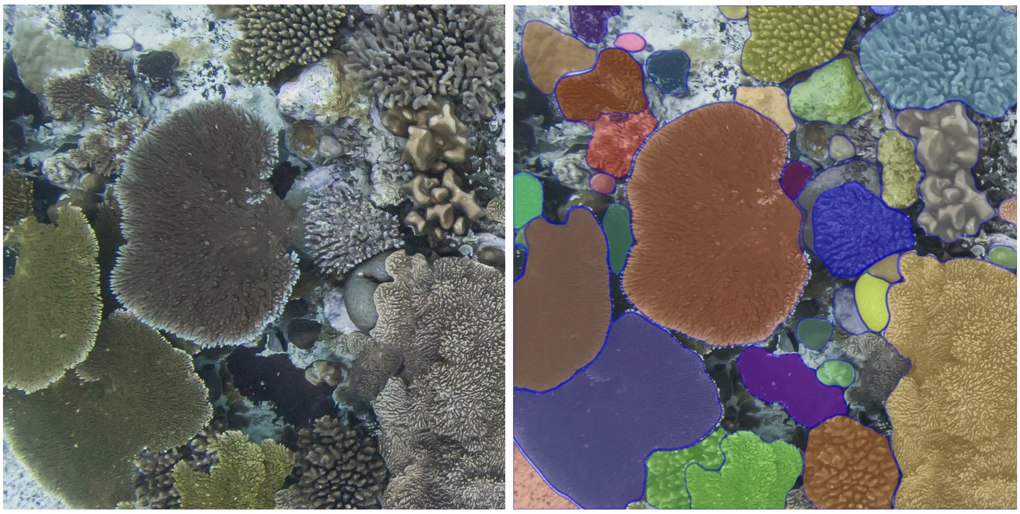


Figure 2.3: Comparison of the original image (left) sourced from CoralNet (*Fiji_Ovalua_GSR*) and its segmented version (right) processed using SAM. The figure illustrates several segmentation artifacts: in the bottom left, a distant coral section is excluded from the segmentation mask, while the coral above it is mistakenly split into two separate segments due to a groove in its texture. Additionally, part of the stem of the large tabular coral in the center is not included in the segmentation mask, demonstrating how SAM tends to respond to texture patterns rather than fully capturing coral structure.

As the authors of SAM note, *"While SAM can perform many tasks, it is unclear how to design simple prompts that implement semantic and panoptic segmentation."* (Kirillov et al., 2023) Given that SAM was explored as a potential straightforward solution without delving into complex computer vision methodologies, further exploratory work for automatically filtering out these mis-segmented regions were regarded as outside the scope of this thesis. However, in Related Work, image segmentation methodologies specifically developed for segmenting corals are discussed.

3 RELATED WORK

3.1 From 3D Reef Modeling to Procedural Generation: Parallels to the Urban Domain

Advancements in 3D reconstruction techniques, such as Simultaneous Localization and Mapping (SLAM) (Johnson-Roberson et al., 2010) and Structure-from-Motion (SfM) (Burns et al., 2015), have made it relatively straightforward to capture the complex structures of coral reefs using accessible tools like video surveys. These techniques remove some of the many logistical barriers of underwater data collection, allowing researchers to create detailed 3D models of coral reef environments without requiring highly specialized equipment or intensive fieldwork. Despite these advances, the development of comprehensive 3D datasets of coral reefs has yet to be realized. Small-scale examples of reef reconstructions exist (Lindhart et al., 2021), but large-scale datasets are limited by logistical challenges in remote and hard-to-access areas, the technical constraints of processing large files, and the comparatively lower demand for such models outside of research contexts.

Given the objectives of this research and its grounding in the field of geomatics, it is useful to draw parallels with other domains that face similar challenges but have seen greater progress, such as urban environments. Cities, like coral reefs, are highly complex systems that are difficult to capture comprehensively in 3D due to their size and structural intricacies. However, in the urban domain, there has been considerable advancement in 3D modeling, driven by a broad range of applications including urban planning, infrastructure development, and smart city simulations, resulting in comprehensive datasets and advanced standards (Gröger and Plümer 2012).

Procedural generation has been adopted in urban environments, allowing researchers to generate synthetic cities that serve as testbeds for studying urban phenomena. As Kim et al. (2018) highlight, the use of procedural city generation has moved beyond the gaming industry to become a tool for scientific research, offering a flexible and standardized environment for testing theories and technologies. For instance, in the development of self-driving cars, procedural generation enables the creation of diverse and plausible urban environments that allow for rigorous testing of algorithms under a wide range of conditions. These synthetic environments can simulate different city layouts, road systems, and traffic conditions that may not be fully captured by existing real-world datasets.

The value of procedural generation in the urban domain lies in its ability to provide controlled, diverse, and flexible environments for testing, independent of the limitations imposed by real-world data. While real-world cities are well-documented, their data can be inconsistent, incomplete, or tied to specific geographic locations, limiting the scope for large-scale, varied simulations. Procedural generation fills this gap by generating cities that, while synthetic, are structurally and behaviorally plausible, enabling a broader range of experimentation and analysis.

Similarly, coral reef research could benefit from the application of procedural generation to create synthetic reef environments. Even with large-scale datasets of real coral reefs, they would only capture specific locations and conditions, limiting their broader ecological utility. Procedurally generated reefs, however, could offer ecologically plausible but flexible environments that serve as testbeds for various ecological studies. These synthetic reef environments would allow researchers to simulate a wide variety of scenarios under controlled conditions, overcoming the limitations of real-world data collection and enabling scalable, replicable studies.

For example, Computational Fluid Dynamics (CFD) simulations demonstrate how synthetic models can significantly contribute to understanding water flow around coral structures. The study by Patil et al. (2024) highlights how detailed, adjustable 3D models of coral reefs improve simulations of water flow and nutrient distribution, which is valuable for assessing coral health and ecosystem sustainability. This underscores the value of synthetic models in advancing research on environmental impacts and conservation efforts.

3.2 Procedural Generation vs. Agent-Based Modeling

Procedural generation offers a method to produce flexible, detailed coral reef models based on spatial patterns from real-world data, but it operates on a static temporal scale. In contrast, simulations, such as those using Agent-Based Modeling (ABM) (Kubicek et al., 2012) (Schneekloth, 2019), are inherently linked to temporal dynamics, allowing researchers to study how coral reefs evolve over time in response to environmental factors, offering a clearer picture of how reefs might respond to future climate scenarios. In this sense, simulation appears to make a more immediate and direct contribution to coral reef conservation by predicting and managing ecosystem changes over time.

While ABM provides valuable insights into the dynamic processes shaping coral reefs, it also requires a deep understanding of environmental factors, which are complex and typically need to be simplified for computational feasibility. Additionally, ABM models rely on site-specific assumptions that make them less applicable to other coral reef ecosystems without significant adaptation (Kubicek et al., 2012). These assumptions can oversimplify interactions within the reef ecosystem, limiting the model's broader utility.

On the other hand, procedural generation has not yet been applied to coral reef research but offers a promising way to address the spatial limitations present in ABM approaches. While procedural generation does not simulate ecological processes over time, it may excel at producing structurally detailed and scalable 3D models of coral reefs.

3.3 Advances in Automated Image and 3D Model Classification for Coral Reefs

Species-specific data collection in coral reef environments has long presented significant challenges to researchers due to the complex nature of underwater environments. As Clinton et al. (2017) noted, *"Application of individual-level demographic data has been limited in subtidal marine environments largely due to the logistical constraints of obtaining data at the appropriate scale."* However, recent advancements in digital imaging have begun to address these constraints.

Progress in digital imaging technology enables the creation of orthorectified photomosaics of subtidal benthic marine environments. These high-detail, large-area images capture extensive sections of coral reefs, providing researchers with a large set of data for analysis. As noted by Edwards et al. (2017), *"the high detail and large spatial extent of photomosaic surveys can capture thousands of individual coral colonies identifiable to species, enabling the extraction of spatially explicit information on benthic communities previously only available through intense field campaigns."*

One frequently assessed metric in coral reef studies is percent cover, which involves classifying species and analyzing their spatial extents. Traditionally, this metric has been assessed by annotating randomly selected points in an image and determining the relative frequency of various classes. However, this process is time-consuming and labor-intensive. To address these challenges, there have taken several steps towards automating the computational procedure and related tasks. Initial efforts focused on developing programs that automatically select random locations in images, facilitate image viewing, and track annotation frequency (Kohler and Gill, 2006). These approaches have been widely adopted by the research community, streamlining the annotation process.

More recently, significant progress has been made in applying machine learning techniques; training automated classifiers such as Support Vector Machines (SVMs) and Convolutional Neural Networks (CNNs) to automate the classification of coral reef images. Beijbom et al. trained SVMs to classify points in reef images based on local texture and color features. Their method achieved promising results, with accuracy rates exceeding 60% for most classes at the genus or functional group level. Building on this work, CoralNet (CoralNet, n.d.) has been developed, an online tool that incorporates these classification frameworks. Initially using SVMs, the platform has since been updated to employ Convolutional Neural Networks (CNNs), resulting in improved system performance and accuracy (Williams et al., 2019). The current CNN-based approach in CoralNet relies on classifying randomly sampled points in images. While this method has proven effective for estimating percent cover and has significantly increased the efficiency of data analysis, it has limitations. Notably, it cannot distinguish individual coral colonies, as it classifies discrete points rather than delineating entire organisms.

Unlike random point sampling, which provides sparse coverage and struggles to distinguish individual coral colonies, image segmentation analyzes all pixels in an image, offering more detailed and spatially explicit information about reef compositions. This approach holds promise for tracking changes in individual colonies over time by capturing finer details, such as colony boundaries and species distribution. Recent studies (Ziqiang et al., 2023) (Pavoni et al., 2019) have shown that dense segmentation can improve coral coverage and population estimates by accounting for the irregular shapes and dense growth patterns that traditional methods often miss. However, segmentation still faces challenges in accurately classifying coral species due to subtle morphological differences, overlapping colonies, and the variability of underwater environments. While this technique represents a significant step forward, it is still relatively new in coral research, and further refinement is needed to generalize across different conditions and achieve fine-grained species classification.

In 2020, Hopkinson et al. became the first to apply automated classification techniques to three-dimensional (3D) reconstructions of coral reefs using convolutional neural networks (CNNs). Prior to this, automated classification in coral reef research focused on 2D image analysis, with no existing methods applied to 3D models. Their approach combines Structure-from-Motion (SfM) to generate 3D reconstructions from images with CNN-based classification to assign species or functional groups to individual mesh elements within the reconstructions. The methodology demonstrated high classification accuracy (~96%) in patches of reef approximately 10x10 meters in size. Their CNN model, ResNet152, was trained using image patches extracted from multiple views of each mesh element, allowing the classification of coral species such as *Acropora palmata, Porites astreoides*, and *Siderastrea siderea*, as well as broader functional groups like “Sea Rods,” which contains multiple genera of octocorals. This approach bears similarities to techniques used for labeling buildings in urban environments (Lotte et al., 2018), again demonstrating the cross-disciplinary nature of these advancements. Despite its promising accuracy, this technique presents significant limitations that render it unsuitable for our research objective, which involve global-scale modeling of species abundance and spatial distribution patterns.

First, while the method works well for classifying small, localized reef patches, scaling up to cover larger areas, such as entire regions or global datasets, remains challenging. The authors acknowledge that "*finer mesh reconstructions could be generated with existing programs but require more computational power to process*" making it difficult to apply this technique on a larger scale. Furthermore, the classification process relies heavily on manually annotated training data, a bottleneck that the authors themselves describe as "time-consuming," limiting its potential for extensive datasets.

3.4 Assessment of Occurrence Databases

Occurrence databases like the Global Biodiversity Information Facility (GBIF) are of great value for biodiversity research, providing large-scale datasets that aggregate species records from around the world (GBIF, n.d.). However, as various studies have demonstrated, including Beck et al. (2014), Kusumoto et al. (2020), and Marcer et al. (2022), there are significant challenges when using these datasets, especially for spatially complex ecosystems like coral reefs.

One of the primary issues with GBIF is spatial bias, where certain regions are over-represented due to concentrated sampling efforts, while others are under-sampled or missing entirely. Beck et al. (2014) highlighted this problem in their study of butterfly distributions, showing that uneven sampling effort leads to biased species distribution models (SDMs). The study found that subsampling over-sampled regions improved the predictive power of SDMs by reducing the impact of clustering in high-density areas. This same issue affects coral reefs, as noted by Kusumoto et al. (2020), who demonstrated that coral diversity is often misrepresented due to inadequate sampling in regions like the western Indian Ocean, despite potentially high species richness. Their study utilized rarefaction techniques to correct for these biases, but it also underscored the persistent issue of incomplete sampling coverage.

Another important limitation is coordinate uncertainty within GBIF records. Marcer et al. (2022) conducted a comprehensive review of georeferencing quality, finding that a large portion of GBIF records lacked precise coordinates or had high levels of uncertainty. This poses a significant challenge for fine-scale ecological modeling, particularly in ecosystems like coral reefs, where species distributions are tightly linked to specific environmental conditions and spatial structures. Without accurate geospatial data, the reliability of models that rely on occurrence records is compromised.

3.5 Macro- and Micro-scale Spatial Patterns of Coral Reefs

Understanding the spatial patterns of coral ecosystems requires data at multiple scales. While satellite imagery is effective for capturing broad reef structures, obtaining individual-level data from remote sensing is not possible. As the creators of Reef Cover state, "*Despite its name, the Allen Coral Atlas is actually a Coral Reef (and not a Coral) Atlas: remote sensing can detect reefs well but is less good at detecting individual corals*." (Kennedy et al., 2020) However, Reef Cover classifies reef structures into geomorphic zones, which act as reliable proxies for estimating species abundance based on physical attributes such as slope, depth, wave energy and substrate type.

These classified geomorphic zones are publicly available through the Allen Coral Atlas, a global platform that provides high-resolution maps covering over 348,361 km^2 of coral reefs (Allen Coral Atlas | Atlas, n.d.). The Atlas integrates satellite data with the Reef Cover classification system, offering a comprehensive resource for researchers and conservationists. While there may still be some regions, particularly deeper or turbid waters, that are not fully mapped (Lyons et al., 2024), the Atlas is a prominent tool, capturing a significant portion of the coral reefs known globally. Its extensive coverage makes it one of the leading references for coral reef mapping, supporting efforts to assess reef health, biodiversity, and the impacts of climate change.

While Reef Cover and the Allen Coral Atlas provide macro-level insights into coral reefs, high-resolution studies have revealed finer-scale spatial patterns that are missed by broader satellite-based systems; a study by Edwards et al. used self-captured large-area photomosaics to conduct a cluster analysis on 16 reef sections of 10x10 meter, identifying significant local variation in coral assemblages. The study found that Complete Spatial Randomness (CSR) could be rejected for almost all species, with most exhibiting clustered spatial patterns due to factors such as dispersal limitations, fragmentation (Furby et al., 2017) and habitat availability.

4 METHODOLOGY

4.1 Baseline pipeline

The baseline pipeline for this methodology is developed to provide a modular and flexible framework for procedural coral reef generation, ensuring that the main components such as terrain generation, coral colony modeling, and species distribution can be developed independently and improved as more advanced data and techniques become available. This modular approach is central to creating a pipeline that can adapt to the ever-evolving nature of coral reef datasets.

The methodology for the baseline pipeline adapts elements from (Patil et al., 2024), which provided a basic system for generating coral reef models using a 2D plane as terrain and Complete Spatial Randomness (CSR) to distribute randomly selected 3D models of coral colonies. However, several improvements were made to refine both the terrain generation and the spatial placement of coral colonies, resulting in a more visually realistic and functional foundation for procedural coral reef modeling.

4.2 Terrain Generation Using Fractional Brownian Motion (fBm)

The terrain in this baseline pipeline is generated using fractional Brownian motion (fBm), an established procedural technique introduced by Benoit Mandelbrot for modeling rough and self-similar surfaces (Mandelbrot, 1983). fBm is well-suited for simulating natural environments, such as underwater landscapes, due to its ability to create fractal-like topographical features with varying levels of detail. This method is widely used in procedural generation because it efficiently produces terrain that is visually complex and realistic.

In the pipeline, the 2.5D terrain is generated by subdividing a 2D plane into a grid of nodes. Each node represents a point on the terrain's surface, and these nodes are displaced vertically based on the fractal noise produced by the fBm algorithm. The fBm method in this case is built upon Perlin noise, a gradient noise function developed by Ken Perlin (Perlin, 1985), which serves as the foundational noise layer. fBm creates complex terrain by combining multiple layers (or octaves) of Perlin noise at different frequencies and amplitudes. This process adds finer details on top of broader features, generating elevation variations that range from large hills to small ridges, mimicking natural seafloor topography.

The subdivision level of the grid is an input parameter in the pipeline, which controls the resolution of the terrain. This parameter allows flexibility in generating terrains with varying levels of detail depending on the desired scale and computational resources.

Although fBm produces visually realistic terrain, it is important to note that this method is not based on real-world bathymetric data. Instead, it serves as a placeholder for terrain generation in the pipeline. Given that terrain generation is a well-established area of research, this component can be replaced with more sophisticated, data-driven methodologies based on real-world bathymetric or ecological data in the future, as discussed in Chapter 6.

To generate the coral colony positions on the terrain, a Complete Spatial Randomness (CSR) function was applied to select random 2D locations for coral colonies. Since the terrain produced by fractional Brownian motion is 2.5D—with elevation variations but no vertical walls or overhangs—the spatial distribution patterns of coral colonies are treated as being 2D. This approach simplifies the distribution of colonies, mapping them onto the 2.5D terrain without considering how complex 3D topographical features might influence colony placement.

While this baseline approach treats spatial distribution as fundamentally 2D, we acknowledge that real-world terrain is fully 3D and that topography can influence where coral colonies establish themselves. This simplification is a limitation of the current modular framework and will be addressed in Chapter 6, particularly when considering potential future improvements that incorporate more detailed, ecologically informed spatial distribution models.

4.3 Smithsonian Institution 3D Model Collection Assessment

The Smithsonian National Museum of Natural History and Digitization Program Office, in collaboration with *The Hydrous*, provide a valuable dataset of high-resolution 3D coral models. The collection includes 89 watertight .stl models of coral species, making it a valuable resource for integrating realistic coral colonies into the pipeline. These models are accessible for download through separate web pages on the Smithsonian Institution's website or via SketchFab, but not as a bulk download; accessing and preparing these models for procedural generation required several steps, given the individual download process and inconsistent metadata.

To automate the process of obtaining these models, I developed a custom Node.js script using the Puppeteer library. This script automated interactions with the Smithsonian's web pages, downloading the models in the desired format and collecting metadata, such as species names and descriptions. The collected metadata was stored in a JSON file, with each model's filename serving as the key, ensuring that the 3D models were organized for easy integration into the pipeline.

Some of the models included a pedestal or metal rods to support the structure in open air. To address this, the open-source software Blender was used to manually inspect the models and remove these extraneous objects. Blender's boolean modifiers were applied to cleanly subtract the unwanted elements, ensuring no missing faces were left behind.

Following these checks, a taxonomic verification process was performed to ensure that the species names associated with each model were accurate and standardized, as detailed in the next subsection.

4.4 Taxonomic Verification Using WoRMS API

In biological taxonomy, species names can change over time as new information becomes available. These changes may result from new research, genetic analysis, or reclassification efforts. As a result, a single species might be known by different names in the literature. To ensure that the pipeline can accommodate the evolving nature of taxonomy and consistently use accurate species names aligned with current taxonomic standards, an automated taxonomic verification process was developed using the *World Register of Marine Species* (WoRMS) API. This step was necessary to address potential discrepancies, such as outdated or synonymous species names, that could lead to inconsistencies when integrating other data sources in the pipeline.

To clarify the different taxonomic statuses returned by the WoRMS API during the automated verification process, the following terms are used to categorize species names and their validity within the database:

- `'accepted'`: The current, valid name of a species as recognized by taxonomic authorities, in this case, WoRMS (WoRMS - World Register Of Marine Species, n.d.) .
- `'junior subjective synonym'`: A name that refers to the same species as the accepted name but was published later. The "subjective" aspect means that the synonymy is based on the taxonomist's interpretation rather than on definitive genetic or type specimen evidence.
- `'superseded combination`': This refers to an older pairing of a species with a genus that is no longer valid because the species has since been reassigned to a different genus.

- `'unaccepted (synonymy)'`: A name that has been deemed invalid because it refers to the same species as the accepted name, and thus is a synonym.

Taxonomic complexities often result in multiple names being associated with the same species, as observed in the metadata of the 3D models sourced from *the Smithsonian Institution*, described in Section 4.3. I developed a script to query the WoRMS API for each species name associated with the models. The script uses the `/AphiaRecordsByMatchNames` endpoint, which uses the TAXAMATCH fuzzy matching algorithm by Tony Rees to handle variations in species names, including spelling discrepancies and alternative names. The script was designed to handle various scenarios: models labeled with a single species name, multiple species names, or combinations of species and genus names, both current and historical (e.g., 'Pleurocorallium niveum (Bayer, 1956)' and 'Corallium sp.'). It also accounts for species names with or without the author, the discovery year, and the use of brackets. For each species name, the script sends a request to the WoRMS API, which returns taxonomic information such as the species' status (e.g., accepted, synonym, superseded combination), as described above. When processing models associated with multiple species names, the script queries each name independently and checks if it resolves to multiple accepted names. If so, a warning is triggered, indicating a potential issue with the matching process or the original labeling of the model. This built-in warning system allows for manual review of potential errors in taxonomic matching. Ultimately, the script iterates through all species names, storing the accepted name for each model to ensure consistency and accuracy. The results of this process are described in Section 5.2.

In addition to verifying the taxonomic names of the 3D models, this process is further used to assess how well other datasets adhere to taxonomic standards. Specifically, we evaluate what portion of the 51 species represented in the 3D models is also present in the other datasets, as well as how many of the 645 unaccepted synonyms from the 3D models are included. The process and results of this assessment are detailed in Sections 5.3, 5.4, and 5.5.

4.5 CoralNet Assessment

CoralNet is an extensive repository of coral reef imagery, containing 3,772,653 images and 165,352,550 classified point annotations, which provide detailed species-level classification for points across coral reef images. CoralNet uses a set of 8,658 labels (CoralNet, n.d.)(**https://coralnet.ucsd.edu/label/list**), which include labels for coral health indicators such as bleaching and mortality. These labels were assessed with the taxonomic process described in Section 4.4.

Although CoralNet is inconsistent with georeferencing, providing only a single optional coordinate for entire image collections without indicating uncertainty, we experimented with extracting species co-occurrence frequencies. When large collections of images with multiple classified species are available, the frequencies of all possible species pairs can be analyzed to infer patterns of species abundance and spatial distribution.

To extract species co-occurrence data from CoralNet, an automated process was developed to iteratively analyze the publicly available sources (CoralNet, n.d.) (accessible at **https://coralnet.ucsd.edu/source/about**). Out of the 4,458 total sources listed on their platform, 1,078 are public. Each source represents a collection of images, though these vary in content; some include images of the same reef patch, while others appear to contain unrelated collections. For each image, random point annotations are made, and species classification is performed on 150x150 pixel patches surrounding each point.

Given CoralNet's data structure, no direct method exists to obtain all species classifications for a single image. Therefore, a custom solution was developed that relied on executing JavaScript code directly in the browser. This script interacted with the platform's user interface, automating the retrieval of data by systematically querying CoralNet's search form for species classifications. By leveraging this browser-based automation, the script iterated through the available patches in each source and performed searches for each species label.

The process involved executing a search query for every species label associated with a given source. Once results were returned, the script extracted and stored relevant data, including the image name, source name, and species classification for each point annotation. This method allowed for efficient data extraction without the need for manual interaction, although handling the platform's slow response times posed a challenge. To mitigate extended running times, the search process was restricted to labels matching species included in the pipeline (see Section 5.1). Specifically, search queries were limited to labels containing both the genus and species name of the 79 species represented in the 3D models. This ensured that label variations, such as terms like 'intermediate' or 'bleached' (e.g., Acropora cervicornis intermediate bleached), were still recognized and classified as the base species (Acropora cervicornis).

After gathering the data, a second script analyzed the aggregated results to identify species co-occurrences within images containing multiple distinct species. The results of this co-occurrence analysis, including the frequency of unique species combinations and their relevance to spatial distribution and abundance patterns, are discussed in detail in Section 5.5.

4.6 GBIF Occurrence Data Assessment

The procedural pipeline aims to generate 3D models of coral reefs that reflect realistic occurrence proportions of each coral species, ensuring ecological plausibility. However, the few datasets with global coverage, GBIF and the Allen Coral Atlas, come with limitations, as discussed in the **related work**. After filtering for the classes Anthozoa and Hydrozoa, 2712813 occurrences from GBIF are used to estimate species abundance across coral reef ecosystems (GBIF, n.d.). While this is a vast collection, it provides only sparse coverage compared to the estimated 384,000 km² of coral reef area mapped by the Allen Coral Atlas, resulting in an average of only 9.29 coral colonies per square kilometer (Allen Coral Atlas | Atlas, n.d.).

Given this sparsity and the arbitrary, non-systematic sampling locations of GBIF data, applying a straightforward plot sampling approach (Borchers et al. 2002), where small, randomly selected 10x10 meter areas are sampled to assess species composition, would not provide reliable estimates. Plot sampling is effective when it can be assumed that each sample covers a representative portion of the population and that species are uniformly distributed. However, GBIF's sampling locations and missing (indications of) coordinate uncertainty values make this assumption invalid.

Instead, the pipeline aggregates species occurrences across larger zones with similar environmental conditions, defined using the geomorphic zones of the Allen Coral Atlas. This aggregated approach compensates for the limitations of plot sampling under these data conditions, allows for a more reliable estimate of relative species abundance across broader areas and allows us to explore the correlation between reef geomorphology and species distribution.

The Allen Coral Atlas provides access to its mapped coral reef zones via downloadable GeoPackage files, which include 30 mapped areas that can be accessed directly from their website (Allen Coral Atlas | Atlas, n.d.). Similarly, GBIF offers its occurrence records in a tab-delimited CSV format, available for download through their platform (after creating an account) (GBIF, n.d.). Both datasets are loaded into a Python environment using the geopandas library. An R-tree spatial index is used to efficiently check for intersections between the species occurrences and the geomorphic zones. The results are aggregated for each geomorphic zone class and stored in a JSON file.

4.7 Geomorphic Zone Overlap Analysis

To ensure that the generated models accurately reflect the global distribution of environmental gradients, which are assumed to correlate with species abundance as represented by the Allen Coral Atlas (as described in Section 4.6), the pipeline assigns species abundance probabilities based on the likelihood that a randomly sampled 10x10 meter area would occur within each geomorphic zone. These probabilities could straightforwardly be derived by analyzing the total area occupied by each geomorphic zone within the coral reef regions, ensuring that larger zones are more likely to be sampled.

In addition to single-zone overlaps, we should account for the possibility that a 10x10 meter area could span multiple geomorphic zones, particularly at their boundaries. This is determined by subdividing the 10x10 meter area into four 5x5 meter quadrants, allowing the script to capture whether the area crosses into adjacent zones. For each quadrant, the geomorphic zone with the largest overlap is selected, and a record is made of the zones' composition across all quadrants. The statistics are aggregated across a large number of randomly generated rectangles to estimate the relative frequency of single-zone overlaps versus multi-zone overlaps.

To ensure that the sampling process converges on a stable estimate of these probabilities, the pipeline iteratively generates 1000 random 10x10 meter areas and tracks changes in the zone overlap frequencies. Convergence is assessed by comparing the relative frequency of zone overlaps across iterations, and once the difference falls below a predefined threshold (0.01%) for each possible overlap composition, the process halts. This allows the pipeline to produce a robust estimate of the probability distribution of geomorphic zone overlaps, ensuring that species abundance distributions are assigned in a way that reflects the natural variability and spatial structure of coral reef geomorphology.

This process was implemented using the same method as described for species occurrence overlap analysis in Section 4.6, using geopandas and an r-tree index to efficiently track and calculate spatial intersections.

4.8 CoralTraits DB Assessment

In Chapter 5, we demonstrated that the prominent spatial datasets we assessed were insufficient to extract fine-scale spatial distribution patterns of coral colonies. However, general ecological traits available in the CoralTraits database, such as larval dispersal methods and growth rates, provide a useful proxy for inferring clustering behavior (Coral Trait Database | Home, n.d.). Edwards et al. (2017) highlight that most coral taxa exhibit non-random clustering patterns on

10x10 meter plots, influenced by life-history strategies, including fragmentation and dispersal limitations. This finding informed the design of our stochastic clustering algorithm.

We implemented a stochastic algorithm capable of placing coral colonies with varying clustering behaviors, as aligned with the findings of Edwards et al. (2017). The algorithm operates on a scale from species approaching Complete Spatial Randomness (CSR), such as *Pocillopora eydouxi* (which did not significantly depart from randomness in Edwards et al.'s variance-to-mean ratio analysis), to species exhibiting high clustering tendencies, such as 'brooding corals' and species prone to fragmentation. Edwards et al. describe how fragmentation, often driven by natural forces like wave action or animal activity, allows coral fragments to reattach and form new colonies under favorable conditions. Our algorithm reflects this range of behaviors, from CSR to highly clustered patterns. The pseudocode for this algorithm can be found in **Appendix (Pseudocode A.1)**.

While CoralTraits provides only three possible values for the 'Mode of Larval Development' trait (brooder, broadcast spawner, or surface brooder), these values serve as essential indicators for clustering behavior. Brooding species, for example, are expected to form denser clusters as their larvae tend to settle closer to the parent colony. Although the available data lacks finer granularity, we regard these three values as 'temporary constants' on a scale from 0 (CSR) to 1 (maximum clustering), with the understanding that this scale may be refined as more detailed information becomes available or extracted from fine-scale spatial datasets.

Edwards et al. (2017) also noted density-dependent clustering behavior in coral colonies, which is reflected in our implementation. In nature, clusters formed by a single species do not maintain a constant variance-to-mean ratio, as the availability of space and habitat filtering cause natural variability. To mirror this, our stochastic algorithm iteratively places colonies based on the derived species-specific abundance data (as described in Section 4.6). The algorithm attempts to position colonies near same-species parent colonies when available, within a radius determined by the species' clustering factor. If no nearby colony is found or the species has a low clustering factor, the colony is placed randomly. This iterative process naturally leads to the formation of multiple clusters as colonies compete for available space, simulating the natural clustering dynamics observed in coral reefs.

To address edge effects, where clusters near the edges or corners of the 10x10 meter terrain could become artificially denser due to placement constraints, our algorithm allows colonies to be placed 'out of bounds' during the generation process. This ensures that the terrain's boundaries do not distort the clustering behavior. While this means that the calculated species abundance (as described in Section 4.6) might not be exactly matched, the inherent limitations in the available species abundance data across geomorphic zones make this a minor issue.

This implementation balances ecological realism with the limitations of the currently available data. By adopting a flexible clustering approach, we account for the lack of detailed data on fine-scale species distribution, while still generating plausible reef structures that reflect general ecological trends. As more precise spatial data becomes accessible, this algorithm can be refined to further enhance its accuracy and representation of coral reef dynamics.

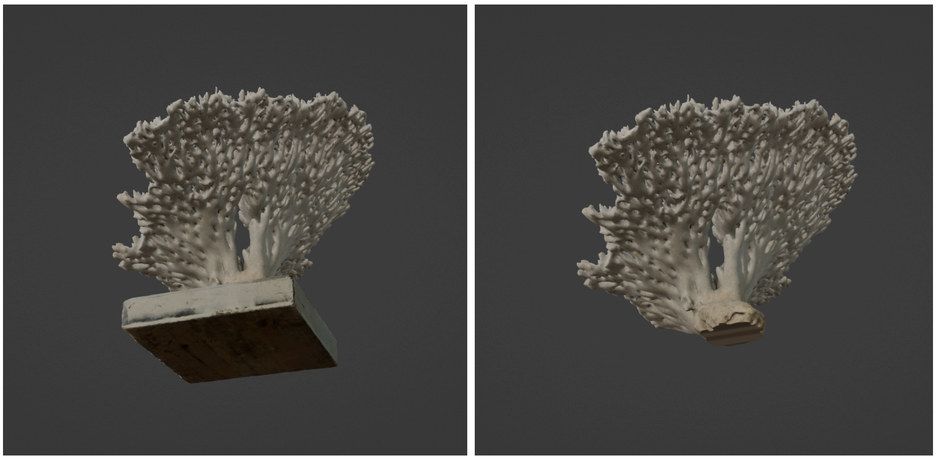
5 RESULTS

5.1 Smithsonian Institution 3D Model Collection Assessment

The filtering process of the 89 initially downloaded 3D models, as described in Section 4.3, resulted in excluding 10 models because they were not related to coral colonies; the excluded models included non-coral specimens, such as a sea dragon carcass and a Bignose Shark jaw. Additionally, 8 models were excluded due to incomplete geometry, where only one side of the colony had been scanned, resulting in one-sided surfaces that lacked complete structural representation. A further 6 models were removed because they represented fragments, such as broken branches of branching corals. While fragments like these are realistic components of coral reef sections, representing fragments was considered outside the scope of this study, which focuses on complete coral colonies. One model was also excluded because it consisted of a small colony merged with a colony from another species (Acropora Secale and Tubipora Musica). Although merging of coral species can occur in nature, the modular approach used in this study focuses on representing single coral colonies.

After this filtering process, 64 models remained, though there were several duplicate models of the same species; the species *Pocillopora Damicornis* appeared four times, and *Acropora Cervicornis*, *Goniastrea Favulus*, and *Pocillopora Molokensis* each appeared twice. This results in a total of 56 unique species represented by the collection of 64 models. These duplicates are not considered problematic, as coral colonies of the same species can exhibit significant structural variation. As discussed in Section 1.6, static models are a limitation because coral species display a wide range of morphologies, so having multiple models for a single species increases visual variability. In fact, the success of the pipeline will benefit from a wider and more variable collection of 3D models in the future, as discussed in Chapter 6.

An example of the result of the manual removal of pedestal with the open-source software Blender is shown in **Figure 5.1.**



(a) **(b)**

Figure 5.1 3D model of *Madrepora spicifera* **(a)** with a pedestal and **(b)** without (Image by author)

5.2 Taxonomic Verification Using WoRMS API

The previous Section 5.1, describes how 64 3D models were collected from the Smithsonian Institution, representing 56 unique species. The taxonomic verification process as described in Section 4.4 was applied to the metadata of these models, yielding the following results:

- 30 models were labeled with a single species name, of which 29 used accepted species names. One model had a superseded combination, but the accepted species name was successfully retrieved using the API.
- 7 models were labeled with both a senior name, i.e. the accepted species name, and a junior subjective synonym.
- 16 models were labeled with an accepted species name and its corresponding superseded combination, reflecting an earlier taxonomic classification.
- 1 model was labeled with both a junior subjective synonym and a superseded combination, both of which corresponded to the same accepted species name, which was retrieved from the API.
- 1 model was labeled with an accepted species name and an unaccepted older genus name.
- 1 model included a superseded combination along with both the name of the previous genus and the newer, accepted genus.
- 1 model contained two junior subjective synonyms.
- 1 model was labeled with two superseded combination names.
- 1 model was labeled with both an accepted species name and an "unaccepted (synonymy)" designation.

For five models, multiple species names were provided that were resolved into two different accepted species names or were both marked as 'taxon inquirendum' when queried via the WoRMS API (WoRMS - World Register Of Marine Species, n.d.) . These cases reflect the complexities of taxonomic classification, where some species are recognized under different names or have unresolved taxonomic identities:

- *Gemmipora brassica* and *Turbinaria brassica*
- *Manopora scabricula* and *Montipora scabricula*

In both cases, the names were flagged as '**taxon inquirendum'**, meaning their classification is uncertain or unresolved. This designation is used when the original descriptions or subsequent research are insufficient to confidently assign the species to a particular taxon. Further taxonomic investigation is required to determine their proper placement, and thus these names are not considered fully accepted.

- *Acropora hyacinthus* and *Acropora humilis*
- *Montipora danae* and *Montipora tuberculosa*
- *Stylaster brochi* and *Stylaster elassotomus*

In these cases, both names were verified as accepted species. Although they represent distinct species, the presence of both names associated with the same model suggests ambiguity in species identification. Further clarification is necessary to determine which species the model accurately represents.

5.3 CoralNet

The taxonomic verification process, described in Section 4.4**,** for the 51 species as represented by the 3D models obtained by the process described in Section 5.1**,** found 9 species (including their synonymized names) to be not included in CoralNet's label set. These species, along with their synonyms, are:

- *Pleurocorallium niveum (Corallium niveum)*
- *Crypthelia viridis*
- *Distichopora borealis*
- *Distichopora violacea (Distichopora cinnabarina, Distichopora fisheri, Distichopora fulvacea, Distichopora rosea, Millepora violacea, Stylaster violaceus)*
- *Leptoseris gardineri (Pavona gardineri)*
- *Homophyllia australis (Caryophyllia australis, Culicia magna, Cylicia magna, Isophyllia australis, Parascolymia australis, Scolymia australis)*
- *Stylaster alaskanus (Stylaster alaskana, Stylaster cancellatus)*

- *Stylaster sanguineus (Stylaster elegans, Stylaster tenuis)*
- *Stylaster verrillii (Allopora moseleyi, Allopora verrillii, Stylaster norvegicus f. pacificus)*

For the remaining 42 species in the collection, CoralNet's set of labels also included the synonymized names of 8 species. Table 1 shows the original species names alongside their synonyms as listed in CoralNet, as well as the taxonomic statuses of these synonyms according to WoRMS, as explained in Section 4.4.

Accepted name	**Synonym**	**Taxonomic status of synonym**
Stylophora pistillata	Stylophora mordax	junior subjective synonym
Psammocora columna	Coscinaraea columna	superseded combination
Acropora hyacinthus	Acropora surculosa	junior subjective synonym
Acropora humilis	Acropora ocellata	junior subjective synonym
Acropora muricata	Acropora formosa	junior subjective synonym
Acropora aspera	Acropora hebes	junior subjective synonym
Helioseris cucullata	Leptoseris cucullata	superseded combination
Astrea curta	Montastraea curta	superseded combination

CoralNet's label set also included multiple variations for a species, typically reflecting different statuses such as bleaching or death. For example, the species Acropora cervicornis was associated with the following label variations:

- *"Acropora cervicornis - bleached"*
- *"Acropora cervicornis intermediate bleached"*
- *"Acropora cervicornis - outplanted"*
- *"Acropora cervicornis recently dead"*
- *"Bleached Acropora cervicornis - Outplanted"*
- *"Acropora cervicornis dead"*

In total, 119 labels corresponding to the 42 species in the collection were identified in CoralNet's label set. This represents only 1.38% of the total 8,613 labels available in the CoralNet system (CoralNet, n.d.). While this is a small subset of the total labels, it is important to note that CoralNet provides a “popularity” metric for each label, which could indicate how frequently each label is used within the dataset. Although there is no documentation available on how the popularity metric is calculated, with values ranging from 0 to 100%, it appears to reflect the frequency of use for each label rather than being relative to the total number of classifications.

Upon reviewing the popularity of the 119 labels corresponding to the species in this study, it was observed that most labels were not significantly low in popularity (e.g., Acropora cervicornis had a popularity score of 83%). This suggests that the 42 species assessed in this thesis are not particularly niche or underrepresented within the CoralNet dataset, despite the limited number of total labels.

The co-occurrence analysis for the 42 species yielded a total of 91 cases where two different species were classified within the same image. No cases of more than two species co-occurring in a single image were found. With 42 species, there are `(42 x 41) / 2 = 861` possible unique species combinations (pairwise combinations excluding same-species pairs). However, out of the 91 co-occurrences, only 21 unique species combinations were observed across the dataset. While co-occurrences of the same species (which could represent different coral colonies) are valuable, the use of random point classification in CoralNet makes it impossible to determine whether the same colony has been classified multiple times.

Despite CoralNet containing more than 3 million images, the number of co-occurrences was surprisingly low. Of the 21 unique species combinations, 6 occurred only once, and only two combinations occurred more than 10 times;
Pocillopora meandrina and *Porites lobata* occurred 31 times and *Diploria labyrinthiformis* and *Pseudodiploria strigosa* occurred 15 times. The small number of unique co-occurrences and their low frequencies make the data insufficient for drawing any statistically significant conclusions about species co-occurrence, spatial distribution, or abundance patterns.

5.4 GBIF Occurrence Data Assessment

The GBIF species occurrence dataset demonstrated complete taxonomic coverage, as all of the accepted species names for the 51 unique species identified through the process described in Sections 4.3, 4.4, 5.1, and 5.2 were included. None of the 645 unaccepted synonyms were included. Filtering of the georeferenced occurrence records for these 51 species resulted in 290213 results, which account for 10.70% of the total 2,712,813 occurrence records belonging to the relevant taxonomic classes *Anthozoa* and *Hydrozoa*.

The methodology described in Section 4.7 was implemented for all of the 30 distinct mapped coral reef areas to estimate the probability distribution of geomorphic zone overlaps in 10x10 meter areas. As expected, for each mapped area, the frequencies of single-zone overlaps mirrored the overall proportions of each zone in the mapped regions. If our interest were solely in these single-zone overlaps, we could have straightforwardly derived these results based on the total area of each zone. The focus here is on the more complex cases where multiple zones overlap within the 10x10 meter area, which occurred on average for 6.68% of the samples for all mapped areas, ranging from 2.08% (*Northern Caribbean + Florida + Bahamas*) to 9.36% (*South China Sea*). The process took 41 iterations to converge for the *Northern Caribbean + Florida + Bahamas* mapped area, and 94 iterations for the *South China Sea* area (1000 samples per iteration). An overview of the multi-zone overlap percentages and iterations to convergence for each of the 30 mapped areas can be seen in **Table A.1** in **Appendix**.

The analysis of multi-zone overlaps also revealed patterns that align with the expected geomorphic relationships outlined in the Reef Cover's classification system (Kennedy et al. 2020). The most frequent zone combinations involved adjacent, shallow geomorphic zones, such as "Inner Reef Flat" bordering "Outer Reef Flat" and "Inner Reef Flat" bordering "Shallow Lagoon". These combinations accounted for 14.23% and 10.40% of all multi-zone overlaps. These frequent combinations suggest a gradual transition between these zones, which is consistent with the geomorphic processes of sediment deposition and wave exposure that shape these areas. In contrast, deeper or more structurally distinct zones, such as Deep Lagoon and Reef Crest, rarely bordered each other, making up only 0.04% of the overlaps. This is likely due to the steeper gradients and sharper transitions that separate these zones. Similarly, Patch Reefs, representing more isolated features, showed relatively low overlap frequencies, with combinations such as Patch Reefs bordering Reef Slope accounting for just 0.02% of overlaps. These results emphasize the importance of accounting for geomorphic transitions in modeling reef structures, as shallow zones exhibit high connectivity, while deeper or more distinct zones tend to form more isolated boundaries.

Table 1 shows the 5 most occurring multi-zone overlaps, including the percentages in which they occurred, relative to the total number of multi-zone overlaps.

Geomorphic Zone Combination	**Zone Bordering Percentages** (relative to total multi-zone overlaps)
Inner Reef Flat & Outer Reef Flat	14.24%
Inner Reef Flat & Shallow Lagoon	10.40%
Back Reef Slope & Deep Lagoon	5.11%
Reef Slope & Sheltered Reef Slope	4.54%
Back Reef Slope & Outer Reef Flat	4.05%

The multi-zone overlap analysis, specifically for *the Great Barrier Reef and Torres Strait* mapped area, revealed interesting directional patterns when comparing the frequencies of overlapping geomorphic zones across the quadrants (top left, top right, bottom right, bottom left) of the 10x10 meter sample areas. While patterns emerged, they did not strictly align with the expected East/North East orientation of the Great Barrier Reef, which generally runs parallel to the Australian coastline. For example, the 'Reef Slope' zone showed the highest overlap frequency in the bottom left quadrant (37.24% of the total multi-zone overlaps with 'Reef Slope'), aligning with the idea that the reef slopes away from the coast. However, the high frequency in the bottom right quadrant (31.11%) was less expected. Upon closer inspection using the Allen Coral Atlas, the complexity of the reef system becomes clearer. The Great Barrier Reef is not a single, continuous structure but a series of smaller, isolated reef patches and formations, each with varying orientations. These smaller reefs, combined with local variations in geomorphic zones such as Reef Flats, Lagoons, and Reef Slopes, explain why directional patterns may deviate from the overall reef alignment. The scattered nature of these reef zones contributes to the irregular patterns observed, as different parts of the reef are shaped by distinct local geomorphic and oceanographic conditions. As a result, while the general orientation of the Great Barrier Reef follows the coastline, the detailed spatial distribution of the reef's zones introduces complexity into the multi-zone overlap directional patterns.

The analysis of overlapping species occurrences across the 30 mapped areas revealed significant disproportionality between species records and the total areas of each region. For example, while the Great Barrier Reef and Torres Strait, with an area of 14836.014 km², had the highest number of overlapping species occurrences at 16356, many other mapped areas exhibited much lower counts. Regions like Western Africa (1652.743 km²) recorded just 8 overlapping species, and the South China Sea (1204.245 km²) had only 48 occurrences. Such low counts make these regions unsuitable for extracting reliable species abundance data. Even the Northern Caribbean, Florida & Bahamas (167666.842 km²), which showed a relatively lower occurrence count of 11949 compared to its size, still provides a substantial dataset from which useful abundance estimates can be drawn, especially in comparison to the regions with extremely low counts. A full overview of species occurrences for each mapped area and their geomorphic zones is shown in the **Appendix, Table A.3**.

In the case of the Northern Caribbean, Florida & Bahamas, the lower species record count may be influenced by its geomorphic characteristics, with 67.48% of the area consisting of Deep Lagoon zones. These deeper zones are typically harder to survey and monitor, leading to fewer recorded species compared to shallower, more accessible zones. This phenomenon, where certain habitat types are underrepresented in biodiversity datasets due to their inaccessibility, has been discussed in the literature (Beck et al. 2014). Additionally, regions like Western Africa and the South China Sea, which exhibit disproportionately low species counts, likely suffer from national-level disparities in biodiversity data collection. As Beck et al. (2014) highlight, countries with fewer resources or less infrastructure for biodiversity monitoring contribute fewer records to global databases like GBIF. This lack of data can create a false impression of lower biodiversity in these regions, when in reality, the scarcity of records is due to insufficient data collection rather than actual species absence.

A surprising result was that out of the total 290213 species occurrences examined, only 85838 overlapped with the geomorphic zones from the Allen Coral Atlas, despite the Atlas' claim of having all coral reefs globally mapped . Upon manual inspection of the non-overlapping records, unexpected clusters were found in regions not typically associated with coral reefs, such as 744 occurrences on the south shores of Japan, 314 in the sea between Alaska and Russia, and 311 on the south shore of Australia. Even more unexpectedly, 20 occurrences were scattered across Europe, all on land and in remote areas, far from museums, research institutions, or zoos. These occurrences appeared arbitrary and did not follow any discernible pattern in proximity to the sea. Similar land-based occurrences were found globally, all within 1000 km of shorelines.

These results raise concerns about the accuracy of certain geospatial records in the GBIF dataset. While some non-overlapping records could be explained by the Allen Coral Atlas not covering marginal reef zones, the presence of coral occurrence records in regions like Alaska, Europe, and remote land areas globally suggests potential geolocation errors or misclassified records. The full list of the 30 mapped areas, including their area sizes and occurrence counts, can be found in the **Appendix**.

5.5 CoralTraits

To evaluate the coverage of the species in the collection, the full species list from the CoralTraits database was obtained in CSV format from their website (Coral Trait Database | Species, n.d.). An analysis was conducted to evaluate the coverage of the CoralTraits database in relation to the 51 species from the coral model collection (Section 5.1). The CoralTraits database focuses on traits of scleractinian corals, which belong to the class Anthozoa and does not include species from the class Hydrozoa. As a result, the 7 species in the collection belonging to Hydrozoa were not covered by CoralTraits. For the remaining 44 species from the class Anthozoa, a comparison was made using the accepted species names according to the World Register of Marine Species (WORMS) (WoRMS - World Register Of Marine Species, n.d.) . 42 species were found in the CoralTraits database with their accepted names, except for two species:

- *Pocillopora grandis* was not included in the database, but the junior subjective synonym *Pocillopora eydouxi*, was listed.
- *Psammocora columna* was also not included, but the superseded combination *Coscinaraea columna*, was found.

Of the 42 Anthozoa species that were found in CoralTraits with their accepted names, 12 species were also listed with one or more synonyms. Among these synonymized species, two cases were noted for special comments in the CoralTraits database:

- For *Leptoseris cucullata*, the superseded combination of *Helioseris cucullata*, CoralTraits (n.d.) included the note: “WoRMS says should be synonymized with *Helioseris cucullata*.”
- For *Porites excavata*, a junior subjective synonym of *Porites lobata*, the CoralTraits (n.d.) database marked it as a “Nomen dubium,” indicating uncertainty regarding its taxonomic status.

For each of the 44 species, the "Mode of Larval Development" trait was retrieved using an automated process. This trait is important for determining clustering behavior, as species classified as 'brooders' tend to exhibit more localized clustering patterns, while 'spawners' often disperse their larvae more widely, resulting in broader distributions.

Out of the 42 species, the "Mode of Larval Development" was available for 27 species. However, for 15 species, this trait was missing in the CoralTraits database. Among the species with available data, 24 were classified as 'spawner,' while only 3 were classified as 'brooder,' indicating a significant skew towards species that rely on broader dispersal mechanisms. This uneven distribution of larval development modes influence the clustering patterns observed in the generated coral reef sections. With only 3 species identified as 'brooder,' the prevalence of dense, localized clustering behavior is limited in the generated reef models. In contrast, the 24 species classified as 'spawner' are more likely to exhibit random or dispersed patterns.

5.6 Resulting Pipeline

As discussed in Section 4.1**,** the pipeline is built on three main components: terrain generation, geomorphic zone selection, and stochastic species placement, each of which is designed to allow future improvements based on more detailed data sources. The terrain generation in the resulting pipeline utilizes fractional Brownian motion (fBm), as discussed in Section 4.2, to create a natural, fractal-like landscape. The subdivision level of the grid, which controls the terrain's resolution, is an input parameter, defaulting to 500 vertices for each dimension. This ensures flexibility in generating terrains with varying levels of detail depending on the desired scale and computational resources. Unlike the baseline pipeline, which used a simple flat plane, fBm is now used to simulate underwater topography, providing a suitable foundation for placing coral colonies in a 2.5D environment. Visually, the fBm-generated terrain appears more realistic compared to the earlier Perlin noise-based terrain, offering a greater degree of complexity and natural variation. **Figure 5.5** illustrates the progression from the flat plane used in the baseline, to the terrain generated by Perlin noise, and finally the fBm-generated terrain in the resulting pipeline.

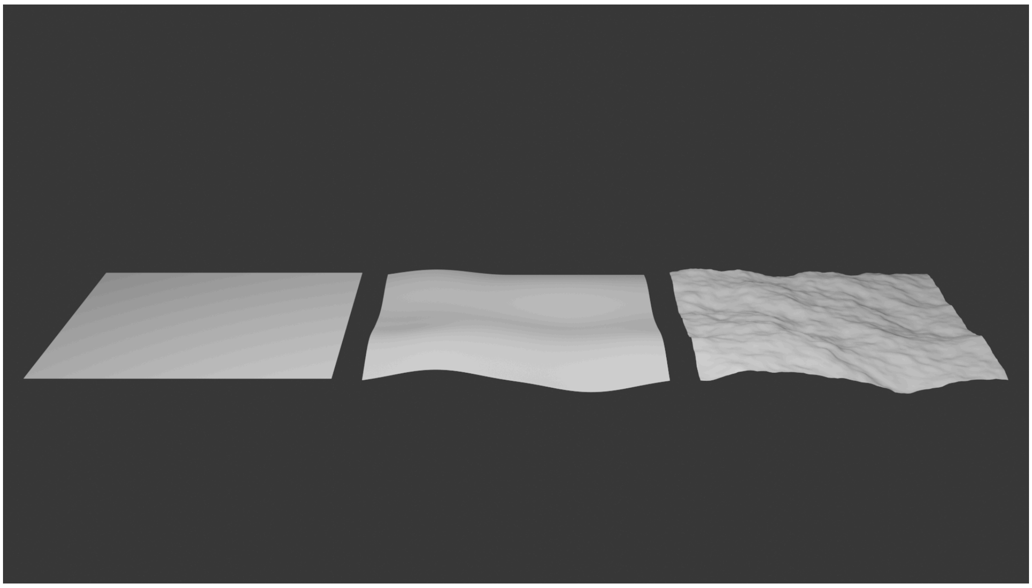


Figure 5.5: (1) Flat 2D plane (2) Terrain generated with Perlin noise (3) Terrain generated with Fractional Brownian Motion (fBm) (Image by author)

The geomorphic zone overlap analysis, detailed in Section 4.7, ensures that species abundance patterns correspond to realistic environmental gradients. The pipeline statistically selects a mapped coral reef area, weighted by the total area of each zone, and then determines the appropriate geomorphic zone(s) for the 10x10 meter terrain patch. As highlighted in Section 5.4, on average, 6.68% of the generated terrain patches overlap with multiple geomorphic zones. In the current implementation, this overlap influences the species abundance numbers for the patch. Future versions of the pipeline could also use this data to inform more realistic terrain generation, incorporating slopes and directional gradients to better represent the physical structure of coral reefs.

Because detailed percent cover or total abundance data for individual 10x10 meter terrain patches could not be extracted from assessed data sources, the total number of coral colonies is treated as an input parameter. Based on Edwards et al. (2017), this input parameter defaults to 2500 colonies per patch. The stochastic placement algorithm, discussed in Section 4.8, then distributes these colonies across the terrain using the species abundance data calculated from the overlap analysis, discussed in Section 4.6. Each colony is assigned a species, and an appropriate 3D model is selected for placement based on the Smithsonian 3D models, described in Section 5.1.

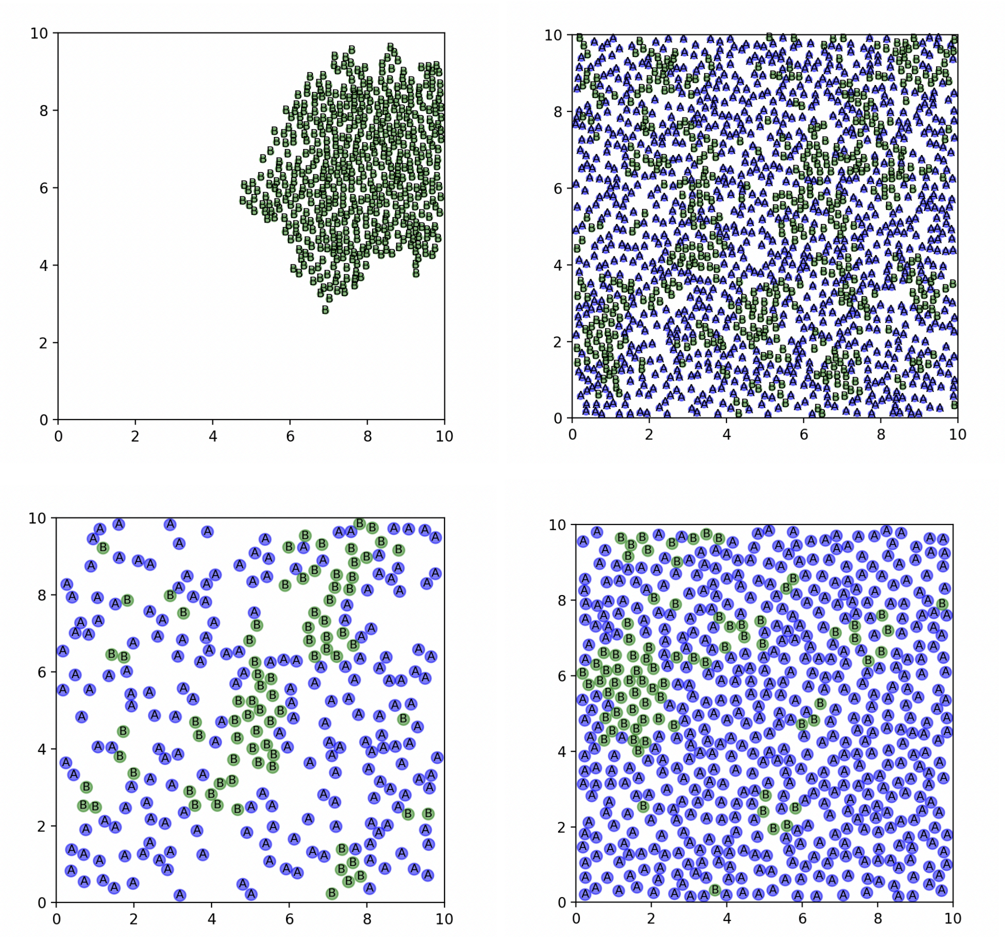


Figure 5.2: Density based clustering, (1) multiple, smaller clusters are formed when less space is available, (2) "*dispersion is linearly related to abundance*" (Edwards et al., 2017) (Images by author)

To simulate realistic spatial patterns, the stochastic algorithm uses a species-specific clustering factor, as discussed in Section 4.8. Currently, a clustering factor of 0.1 (indicating low clustering) is assigned to all of the species we currently have in our pipeline, except for the three 'brooder' species identified in CoralTraits (Coral Trait Database | Species, n.d.), which are assigned a higher clustering factor of 0.8. **Figure 5.2** shows examples of 2D distributions generated by the stochastic placement algorithm, demonstrating both random and clustered species distributions. A realistic scenario for a "Sheltered Reef Slope" zone in the Southwestern Pacific is visualized in **Figure 5.3**, showing a diverse array of species with small clusters of the three 'brooder' species. The 3D model produced by the pipeline for the same scenario is shown in **Figure 5.4.**

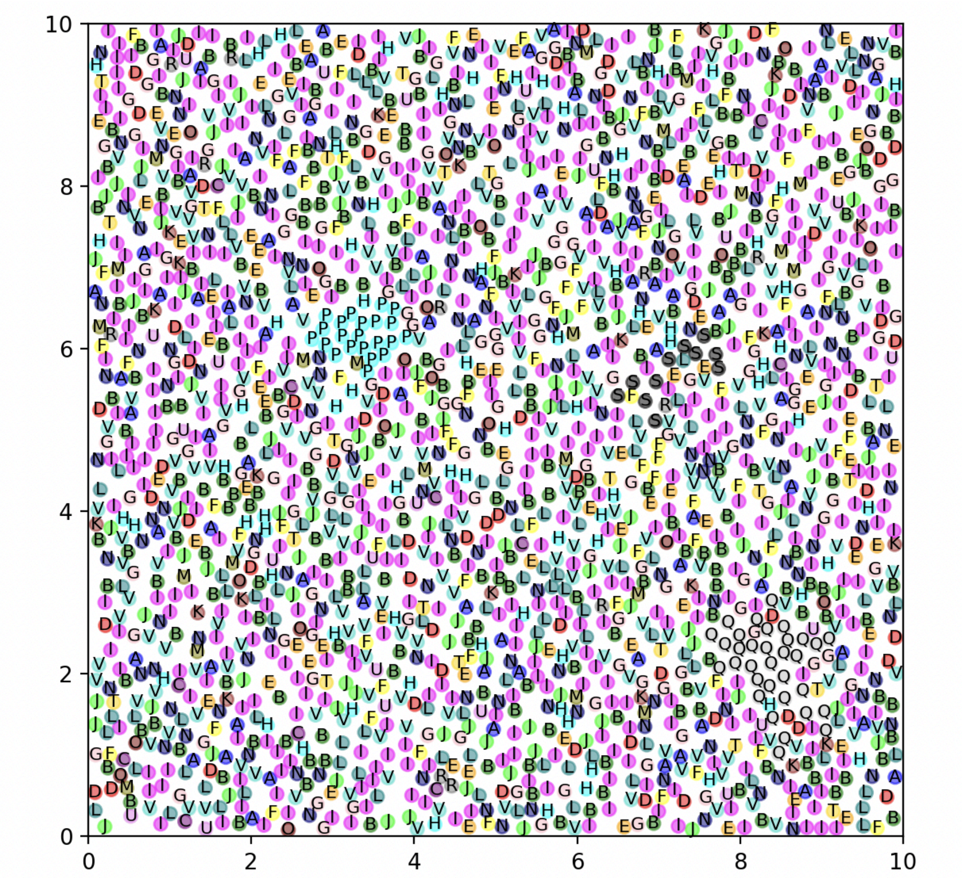


Figure 5.3: A realistic scenario for a "Sheltered Reef Slope" zone in the Southwestern Pacific, with colony centroids 'P', 'Q' and 'S' showing clustered distribution patterns due to being 'brooders' (Image by author).

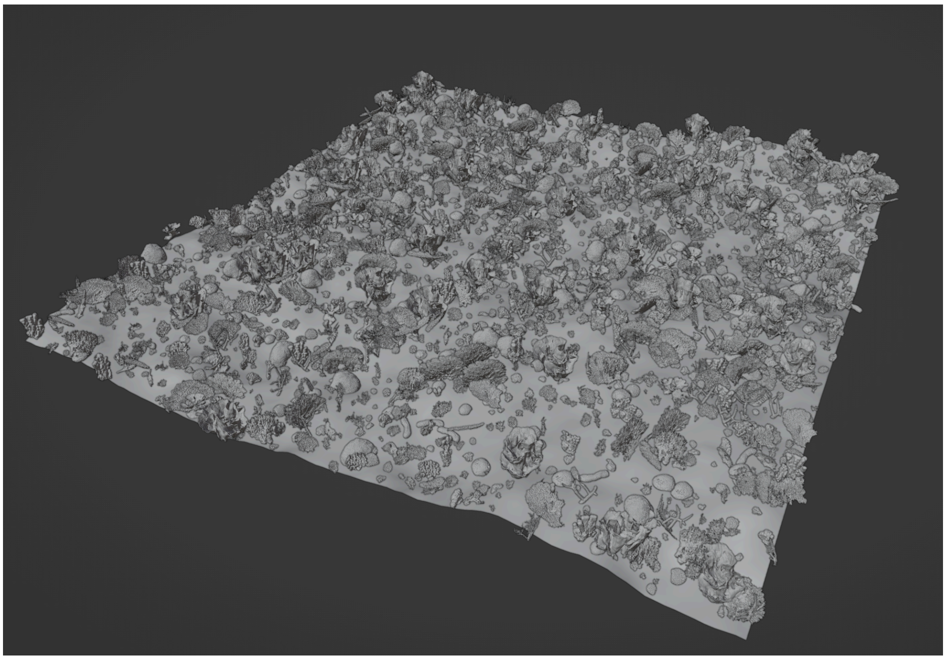


Figure 5.4: A generated model representing a "Sheltered Reef Slope" zone in the Southwestern Pacific (Image by author).

The pipeline was implemented as a Python script designed to be executed within Blender, using the `bpy` module to control Blender's 3D modeling environment. This script can be run in background mode using the command `/path/to/blender --python --background script.py`, enabling automated model generation without the Blender GUI. Each generated model is exported as a Blender file, where the terrain and corals are separate, editable objects. This setup allows for easy visualization, adjustments, and further manual modeling efforts within Blender. The complete source code for the pipeline is publicly available for review and use in a GitHub repository: https://github.com/GGBRW/CoralGenerator. Instructions for exporting models from these Blender files as watertight `.stl` files are provided in the Appendix.

6 DISCUSSION AND CONCLUSION

This thesis aimed to assess and integrate empirical data sources for the automated, data-driven generation of 3D models of coral reefs. By addressing the main research question through a structured set of subquestions, we developed a pipeline that leverages available data and technology to create realistic and scalable coral reef models.

6.1 Identifying Empirical Data Sources

In addressing the first subquestion, Exploratory Work and Related Work explored various empirical data sources, discussing their initial relevance, data quantity, accessibility, and potential for scalability. These chapters provided information about the strengths and limitations of each source, allowing us to evaluate their suitability for coral reef modeling.

Initially, using 3D scanned coral reefs as input for generating 3D coral reef models seems almost trivial—essentially an extrapolation task, where the aim is to produce something that closely resembles the original scans. However, introducing realistic variations within these models requires a deeper semantic understanding, which is far less straightforward. In this context, 'semantic' 3D models refer to models that not only capture the geometry but also encode meaningful information about the structure and biodiversity of the coral ecosystem, similar to the concept of semantic city models in (Groger et al., 2012 CityGML).

Unfortunately, no comprehensive sources of semantic 3D models for coral reefs were identified. The option of employing the 3D automated classification method by (Hopkinson et al., 2020), as described in Chapter 3, was also disregarded due to the scarcity of 3D scans and the complexity of processing them at scale. As further detailed in Chapter 2, only a few isolated 3D coral scans were found, but their limited scale, fragmented nature, and the computational challenges they presented made them unsuitable as a primary data source for this thesis. This situation reflects a tension within the thesis: while the goal is to overcome the limitations of existing 3D data sources to generate realistic reef models, the reliance on this same sparse data constrains the ability to create fully realistic and varied alternatives.

Images of coral reefs were considered as a potentially rich source of data, given their expected abundance and the possibility of extracting detailed semantic information through segmentation. Image segmentation technologies are rapidly advancing, with newer methods such as Segment Anything Model 2 (SAM 2) (Ravi et al., 2024) and Depth Pro (Bochkovskii et al. 2024) being released during the final stages of this thesis. However, no comprehensive sources of semantically segmented coral reef images were identified. The exploratory efforts detailed in Chapter 2 involved attempting to perform straightforward segmentation using the general-purpose model Segment Anything (SAM) (Kirillov et al., 2023). These efforts revealed that segmentation of coral images is more complex than initially anticipated, with numerous edge cases and inconsistent results that undermine the feasibility of reliably extracting structural and biodiversity information from segmented corals within the scope of this thesis.

- **CoralNet** was a significant data source identified, offering over 3.9 million images and nearly 189 million point annotations (CoralNet, n.d.). While CoralNet provides extensive species-level information, its utility for extracting structural data is limited. The use of point classification is less promising for structural extraction compared to 3D data and segmented images, as it does not effectively distinguish between separate corals. For instance, a single coral colony might be classified multiple times by different point annotations, leading to ambiguities in spatial arrangements. Despite these limitations, CoralNet remains a relevant source for extracting species information, as its vast repository facilitates large-scale species identification, which is essential for understanding biodiversity patterns within coral reef ecosystems.
- The **Coral Trait Database** was identified as a valuable source of empirical measurements, a source that *"aims to bring together physiological, morphological, ecological, phylogenetic and biogeographic trait information into a single repository"* (Madin et al. 2016) for over 5,000 coral species. This database is highly relevant for modeling species-specific traits, contributing to the ecological realism of the generated 3D models. However, an initial exploration of the CoralTraits website's traits overview (Coral Trait Database | Home, z.d.) (can be found on www.coraltraits.org/traits) revealed significant limitations. Not only were key morphological traits missing for many species, but the structural information provided was often limited to broad classifications and generalized measurements. For example, branch-related traits were absent for *Acropora cervicornis*, one of the most studied branching corals. These gaps, coupled with the lack of detailed structural data, undermine the database's utility for creating precise, species-level models of coral structures.
- **The Smithsonian National Museum of Natural History, in collaboration with The Hydrous** offers a collection of 89 high-resolution 3D models described as "coral and reef-dwelling specimens" (*Corals | 3D Digitization*, n.d.). These models come with species names and detailed metadata and are available for download in multiple formats, including watertight .stl meshes, making them accessible for integration into coral reef modeling. However, since this is a museum collection, further

analysis is needed to determine whether the 3D models realistically represent coral colonies as they appear in natural reef environments. The historical nature of some specimens raises questions about their relevance to contemporary coral reef structures. Despite these considerations, the models provide valuable species-specific structural data, offering a foundational resource for incorporating detailed coral representations into the modeling pipeline.

- The **Allen Coral Atlas** emerged as a comprehensive source, offering global maps of coral reef environments covering approximately 423,589 km² (Allen Coral Atlas | Atlas, n.d.). Its extensive coverage and adoption of the Reef Cover classification system (Kennedy et al., 2020) provide valuable habitat context, crucial for understanding large-scale environmental factors influencing coral distribution, as noted in Related Work. The Reef Cover classification has the potential to become a standard in coral reef research. However, Related Work discussed the Atlas lacking the finer spatial detail needed for modeling individual coral colonies or species-level distributions, limiting its direct application for detailed 3D models.
- The **Global Biodiversity Information Facility (GBIF) (GBIF, n.d.)** was identified, given its vast repository of species occurrence data. GBIF's extensive coverage includes millions of records across a wide range of coral species globally. However, the direct relevance of GBIF for the scale chosen in this thesis (100m²) is low due to significant spatial biases and coordinate uncertainties, as discussed in Related Work. These issues result in uneven sampling densities and inaccuracies that complicate the extraction of precise spatial information necessary for fine-scale modeling. Consequently, while GBIF offers valuable large-scale species data, its limitations in spatial precision render it less suitable for detailed, small-scale (100m²) coral reef modeling without substantial data processing and aggregation.

These data sources, though valuable, each have limitations that affect their scalability, representativeness, or ability to contribute to detailed modeling efforts. As a result, the selection and processing of these sources in the pipeline needed to account for these limitations, leading to the exclusion or adjustment of certain data during further analysis and modeling.

6.2 Extraction and Processing of Relevant Information

The second subquestion focused on how relevant information could be extracted and processed from the identified data sources, as well as the challenges that emerged in doing so. While the first subquestion focused on identifying which data sources could be useful, this subquestion addresses the actual steps taken to make the data usable and whether the information extracted is relevant for further use in the modeling pipeline.

- **CoralNet** was identified as a useful source for species information, particularly for extracting species co-occurrence data. The assumption was that sufficient combinations of species would be annotated within the same image, providing insights into how species coexist in close spatial proximity. The goal was to extract these co-occurrence frequencies of species within these images to infer probabilities of species being generated together in the same 3D model, contributing to the ecological realism of the models, as described in Chapter 4. However, as detailed in Chapter 5, the analysis revealed unexpectedly few cases where multiple species were annotated within the same image. The scarcity of co-occurrences undermined CoralNet's potential for this task, leading to the conclusion that this data source could not provide the necessary species interaction patterns for further use.
- To extract useful species occurrence information from **GBIF**, the thesis adopted an approach of aggregating occurrence records over larger geomorphic zones. This method mirrors the "subsampling routine" proposed by Beck et al. (2014), which is discussed in Chapter 3. By aggregating data across broader spatial scales, it became possible to address the spatial biases and uneven sampling densities inherent in GBIF's raw occurrence data, particularly when dealing with fine-scale modeling such as the $100m^2$ coral reef sections targeted in this thesis. To define these larger zones, the **Allen Coral Atlas** was selected as a reliable source for geomorphic zone classification. The Atlas provides geopackage files for 30 mapped regions, offering a global coverage of coral reef environments with geomorphic zones classified according to the **Reef Cover classification system**. This allowed for a systematic aggregation of GBIF occurrence records within these predefined geomorphic zones, ensuring that species occurrence data could be analyzed in a spatially explicit manner.
- **The Smithsonian National Museum of Natural History, in collaboration with The Hydrous** provided high-resolution scans of coral specimens, which required a filtering process to extract models relevant for representing different coral species in the modeling pipeline. This filtering, as described in Chapter 4, reduced the set of available models to those suitable for structural representation. Chapter 5 revealed that after filtering, models remained for 51 different species, although most species were represented by only a single model. The representativeness of each model for its species is questionable, given that each species is only represented by one model. In combination with this, the observed age of the specimens in the metadata raises further concerns. While the models

provide high structural detail, they are therefore not suitable for extracting information about realistic structural variation within coral species.

6.3 Exploring the Integration of Extracted Data

The third subquestion explored how the information extracted from the identified data sources could be integrated into a modeling pipeline for coral reef generation. Given the limitations of the available data, various approaches were considered to determine how best to utilize each source for constructing a coherent and effective pipeline.

While answering the first two subquestions, it was observed that species names across the identified empirical data sources often included synonyms—names that are not accepted in the current field of taxonomy. This inconsistency presented a challenge for integrating information from different sources, as it could lead to mismatches in species identification. Additionally, taxonomy is dynamic and continues to evolve, meaning that species names might change or become obsolete over time. To ensure compatibility between the extracted information from the various empirical data sources and to future-proof the pipeline for further iterations, the automated detection and elimination of species synonyms became necessary.

As described in Chapter 4, an automated process was developed using the **World Register of Marine Species (WoRMS) API**. This process identifies species synonyms and retrieves their accepted species names during the extraction of information from the empirical data sources. In Chapter 5, we demonstrated that this automated process successfully identified and resolved synonyms, ensuring that the extracted data across the pipeline maintained consistency with the current accepted taxonomy. By integrating this step into the pipeline, the process supports accurate species identification and alignment between data sources, while also providing flexibility for future updates as taxonomic classifications evolve.

The **Smithsonian Institution's models** were found to use a significant number of unaccepted species names, which is not surprising given that much of the metadata showed the scanned corals were collected in the 19th century (*Corals | 3D Digitization*, n.d.). This inconsistency in taxonomy eventually led to the exclusion of five models, as their associated species names were uncertain or ambiguous, as discussed in Chapter 5. The automated synonym detection process played an important role in identifying these discrepancies, allowing for the correction or exclusion of models that could not be confidently aligned with current taxonomic standards.

- For the **Smithsonian National Museum of Natural History, in collaboration with The Hydrous**, their collection of 3D scanned models provided high-resolution structural details for various coral species (*Corals | 3D Digitization*, n.d.). However, as discussed in Chapter 5, most species were represented by only a single model. This made the source unsuitable for extracting information about realistic structural variation within coral species' morphology, as multiple models or detailed measurements are required to capture natural variation. Since no other identified sources could provide this kind of information, procedural modeling of individual coral structures was not further explored in this thesis. Instead, the Smithsonian models were selected as 'building blocks' for the generation process of coral reef models, adapting the approach used by Patil et al. (2024). This method involved distributing the models across a terrain model and applying randomized rotations around the Z-axis to introduce limited structural variation. While this approach could not replicate realistic morphological diversity, like growth or health stages of corals, it provided a practical way to use the available high-resolution models within the constraints of the data.
- For **Allen Coral Atlas**, the geomorphic zone maps played an important role in integrating spatially explicit information into a modeling pipeline. As described in Chapter 5, the Atlas was used to extract and integrate realistic weighted choices for geographic areas and geomorphic zone types that correspond to the coral reef models. This approach ensures that the models reflect realistic spatial patterns and species compositions tied to different reef environments as classified by the **Reef Cover classification system**. Rather than generating random or uniform distributions, the models are based on real-world geographic variation and reef characteristics. This enabled for the realistic integration of extracted species occurrence information from **GBIF**. By aggregating GBIF occurrence records over the same geographic zones, the pipeline could accurately reflect the variations in relative species occurrences.

6.4 Constructing and Evaluating the Modeling Pipeline

The previous subquestions led to the extraction of information directly relevant for the integration in a modeling pipeline, Chapter 5 describes how such a pipeline was constructed, allowing us to explore its benefits and limitations.

- A fundamental component of generating coral reef models is the terrain on which the reef develops. This terrain acts as a 'canvas' for the model, with morphological characteristics such as elevation variations setting the foundation for coral growth and distribution. Due to the absence of such data, we had to set plausible values manually as input to a procedural terrain generation technique. This introduces a degree of uncertainty, as the natural variability of reef terrains—often crucial to growth patterns and species distribution—cannot be fully replicated without this specific data. Furthermore, the procedural technique implemented, fractional Brownian motion (fBm), generates only 2.5D terrains. This constraint, compounded by limited 3D data, precluded the exploration of fully 3D features (e.g., vertical overhangs, caves) and their potential effects on coral growth.

- Representing corals as 'building blocks' using scanned 3D models provided by **the Smithsonian National Museum of Natural History, in collaboration with The Hydrous**, introduces significant structural homogeneity to the generated models (*Corals | 3D Digitization*, n.d.). Most species are represented by only one model, which limits the natural variability within species. The pipeline, however, includes a feature to randomly select from multiple models where available, currently applicable to only four species, which hints at its capacity to incorporate more intra-species variation if additional models become available in the future.
- Information about the natural density of corals across reef terrains could not be extracted from the available data sources, so the total number of corals in each generated model is set as an input parameter. To address distribution, a stochastic algorithm was developed to place corals across the terrain, where clustering factors can be applied to specific coral species, producing natural-looking clusters. This approach aligns with findings by Edwards et al. (2017), who observed that coral colonies often form clusters. The 'Mode of larval development' data extracted from CoralTraits was used as input for this clustering algorithm, though the suitability of this empirical qualitative data is uncertain within the scope of this thesis. While the mode of larval development is supported by empirical observations, its integration here may reflect more of a theoretical framework for spatial distribution rather than direct empirical input for clustering patterns.
- The pipeline's probabilistic species abundance method uses species occurrence aggregates over Allen Coral Atlas geomorphic zones, which can cover thousands of kilometers of coral reef. Due to spatial bias—consistent with findings by Beck et al. (2014)—certain zones are significantly underrepresented, making the averaged species compositions within these zones less reliable. Moreover, the lack of finer spatial data means that distribution patterns within these zones are not accurately reflected. For instance, a species that may densely dominate smaller areas within a geomorphic zone could be inaccurately modeled as dispersed evenly across all generated models within that zone.

6.5 Overall Conclusion

This thesis reveals significant limitations in the available empirical data for coral reefs, particularly regarding the lack of structural data necessary for generating fully automated, ecologically plausible models. Assessment of the empirical data sources showed very limited information on structural variability within coral reefs, with only structural differences between species being represented. As a result, important aspects such as terrain characteristics, coral population density, and species distribution required manually configured parameters within the constructed pipeline. Given that such foundational parameters need manual configuration, accessible visualization and the ability to adjust models were prioritized over achieving a fully autonomous, automated pipeline. Consequently, the pipeline was implemented in Blender and made available in an open repository (https://github.com/GGBRW/CoralGenerator). This implementation enables the generated models to be inspected, edited, and further refined using Blender's powerful tools for handling complex 3D models. While this approach may not directly fulfill the initial goal of performing Computational Fluid Dynamics (CFD) simulations—one of the primary motivations of this thesis—guidelines for exporting watertight .stl meshes are provided in the Appendix, supporting further use of the models in CFD or other simulation applications.

While the current pipeline has limitations in generating fully ecologically plausible structures, it successfully establishes a foundation for future iterations. The pipeline incorporates several elements that could support more advanced modeling as improved data sources become available.

- First, models are given a geographic and ecological context by automatically selecting a geographic location and geomorphic zone for each model. This approach uses the Reef Cover classification system from the Allen Coral Atlas, a system with the potential to become a standard in coral reef research, potentially enabling the seamless integration of future data.
- The pipeline's modular structure—with distinct steps for generating terrain, representing individual coral structures, selecting species, and spatially distributing corals—provides a flexible framework. This design allows for future iterations to integrate improved data sources and incorporate more sophisticated modeling techniques for each stage.
- Lastly, recognizing the evolving nature of taxonomy and its potential impact on data compatibility was an important insight. Taxonomic changes, such as species name updates, can lead to integration issues with new or legacy data. To address this, a solution was implemented using the WoRMS API to retrieve accepted names for coral species. This ensures

the pipeline remains compatible with current taxonomy, supporting future data updates and alignment with contemporary classification standards.

6.6 Limitations and Future Work

This thesis represents an initial exploration into the automated modeling of coral reef ecosystems using empirical data, and faced several limitations, particularly due to constraints in data availability, performance considerations, and the complex ecological patterns involved in coral reef modeling.

- For the data sources GBIF, CoralNet, and Coral Traits Database, data extraction and taxonomic assessments were performed using only the 51 species that were extracted from the Smithsonian National Museum of Natural History, in collaboration with The Hydrous' 3D model collection. To ensure that these species were not overly specialized or niche, we assessed their representativeness for these broader data sources, validating that they provide a sufficiently general basis for meaningful conclusions. However, this small sample size remains a limitation, influenced by performance considerations and the need for results that align directly with the pipeline's current capabilities. Future work could expand species representation to enhance the accuracy and ecological realism of the models.
- Terrain plays a foundational role in coral reef structure, and terrains slopes and aspects influence coral growth patterns due to their effects on water flow and light availability. However, the only empirical data source available for such terrain details was 3D data, which was limited and not integrated into the pipeline. Early in the thesis, the importance of realistic terrain modeling was overshadowed by the focus on species-specific data, and bathymetric data was deemed unsuitable due to its insufficient quality and resolution for reef modeling. Future work should consider integrating high-resolution bathymetric data, allowing terrain features such as slope and aspect to directly influence coral distribution patterns and better reflect natural conditions.
- The integration of GBIF data with Allen Coral Atlas geomorphic zones allowed for a valuable automated and flexible extraction of species distribution data, but this process represents a limited adaptation of Species Distribution Modeling (SDM). In ecology, SDM involves predictive modeling that incorporates environmental factors to map potential species distributions. Consulting an ecology expert is recommended to better refine this process, identify potential enhancements, and address ecological factors that could strengthen the pipeline's species distribution approach.

- Another data limitation emerged from inaccuracies in GBIF records, with many occurrences showing significant coordinate errors or missing uncertainty information. This resulted in numerous records that placed coral species on land or in unlikely environments. Our findings are consistent with recent recommendations by Marcer et al. (2022), who advocate for improved georeferencing, inter-institutional collaboration, and better transparency in occurrence data. Further refinement of GBIF records would be beneficial for future coral reef modeling.
- The lack of fine-scale data on coral clustering patterns required using the CoralTraits Database's 'Mode of Larval Development' as a proxy for clustering behavior. The algorithm developed in this thesis assigns simplified 'clustering factor' values, simulating coral colonies competing for available space. While this approach aligns with general findings on clustering (Edwards et al., 2017), it does not perfectly replicate specific clustering tendencies for each species. If detailed clustering data becomes available, adopting deterministic spatial point processes, such as the Neyman-Scott or Thomas processes, is recommended to better represent species-specific clustering patterns.

6.7 Final Remarks

The challenges encountered in this research reflect broader issues within coral reef science regarding data availability and quality.
By fostering collaboration and emphasizing data quality improvements, researchers can significantly enhance our capacity to model and understand coral reef ecosystems. Such advancements are important for informing conservation strategies and managing coral reefs in the face of ongoing environmental threats.

REFERENCES

Edwards, C. B., Eynaud, Y., Williams, G. J., Pedersen, N. E., Zgliczynski, B. J., Gleason, A. C. R., … Sandin, S. A. (2017). Large-area imaging reveals biologically driven non-random spatial patterns of corals at a remote reef. *Coral Reefs*, *36*, 1291–1305.

Hughes, T. P., Barnes, M. L., Bellwood, D. R., Cinner, J. E., Cumming, G. S., Jackson, J. B. C., … Others. (2017). Coral reefs in the Anthropocene. *Nature*, *546*(7656), 82–90.

Madin, J. S., Anderson, K. D., Andreasen, M. H., Bridge, T. C. L., Cairns, S. D., Connolly, S. R., … Others. (2016). The Coral Trait Database, a curated database of trait information for coral species from the global oceans. *Scientific Data*, *3*(1), 1–22.

Hoegh-Guldberg, O. (1999). Climate change, coral bleaching and the future of the world's coral reefs. *Marine and Freshwater Research*, *50*(8), 839–866.

Thamarai, M., & Aruna, S. P. (2023). Stressed Coral Reef Identification Using Deep Learning CNN Techniques. *Journal of Electronic & Information Systems*, *5*(2), 1–9.

Houk, P., Camacho, R., Johnson, S., McLean, M., Maxin, S., Anson, J., … Others. (2015). The Micronesia Challenge: assessing the relative contribution of stressors on coral reefs to facilitate science-to-management feedback. *PlOS One*, *10*(6), e0130823.

Watanabe, A., & Nakamura, T. (2019). Carbon dynamics in coral reefs. *Blue Carbon in Shallow Coastal Ecosystems: Carbon Dynamics, Policy, and Implementation*, 273–293.

Borchers, D. L., Buckland, S. T., Zucchini, W., & Borchers, D. L. (2002). *Estimating animal abundance: closed populations*.

Fisher, R., O'Leary, R. A., Low-Choy, S., Mengersen, K., Knowlton, N., Brainard, R. E., & Caley, M. J. (2015). Species richness on coral reefs and the pursuit of convergent global estimates. *Current Biology*, *25*(4), 500–505.

Smith, J. E., Brainard, R., Carter, A., Grillo, S., Edwards, C., Harris, J., … Others. (2016). Re-evaluating the health of coral reef communities: baselines and evidence for human impacts across the central Pacific. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1822), 20151985.

Wicks, M. (2016). *Science comics: Coral reefs: Cities of the ocean*. First Second.

Gracias, N. R., Van Der Zwaan, S., Bernardino, A., & Santos-Victor, J. (2003). Mosaic-based navigation for autonomous underwater vehicles. *IEEE Journal of Oceanic Engineering*, *28*(4), 609–624.

Lirman, D., Gracias, N. R., Gintert, B. E., Gleason, A. C. R., Reid, R. P., Negahdaripour, S., & Kramer, P. (2007). Development and application of a video-mosaic survey technology to document the status of coral reef communities. *Environmental Monitoring and Assessment*, *125*, 59–73.

Spalding, M., Ravilious, C., & Green, E. P. (2001). *World atlas of coral reefs*. Univ of California Press.

Pavoni, G., Corsini, M., Callieri, M., Palma, M., Scopigno, R., & Others. (2019). Semantic segmentation of benthic communities from ortho-mosaic maps. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*(2), 151–158.

Ziqiang, Z., Yaofeng, X., Haixin, L., Zhibin, Y., & Yeung, S.-K. (2023). Coralvos: Dataset and benchmark for coral video segmentation. *arXiv Preprint arXiv:2310. 01946*.

Kohler, K. E., & Gill, S. M. (2006). Coral Point Count with Excel extensions (CPCe): A Visual Basic program for the determination of coral and substrate coverage using random point count methodology. *Computers & Geosciences*, *32*(9), 1259–1269.

Williams, I. D., Couch, C. S., Beijbom, O., Oliver, T. A., Vargas-Angel, B., Schumacher, B. D., & Brainard, R. E. (2019). Leveraging automated image analysis tools to transform our capacity to assess status and trends of coral reefs. *Frontiers in Marine Science*, *6*, 222.

Hopkinson, B. M., King, A. C., Owen, D. P., Johnson-Roberson, M., Long, M. H., & Bhandarkar, S. M. (2020). Automated classification of three-dimensional reconstructions of coral reefs using convolutional neural networks. *PloS One*, *15*(3), e0230671.

Runting, R. K., Phinn, S., Xie, Z., Venter, O., & Watson, J. E. M. (2020). Opportunities for big data in conservation and sustainability. *Nature Communications*, *11*(1), 2003.

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, *19*, 10–15.

Kennedy, E., Roelfsema, C., Lyons, M., Kovacs, E., Borrego-Acevedo, R., Roe, M., … Others. (2020). Reef Cover: a coral reef classification to guide global habitat mapping from remote sensing. *BioRxiv*, 2020–2009.

Bochkovskii, A., Delaunoy, A., Germain, H., Santos, M., Zhou, Y., Richter, S. R., & Koltun, V. (2024). Depth Pro: Sharp Monocular Metric Depth in Less Than a Second. *arXiv Preprint arXiv:2410. 02073*.

Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., … Others. (2024). Sam 2: Segment anything in images and videos. *arXiv Preprint arXiv:2408. 00714*.

Furby, K. A., Smith, J. E., & Sandin, S. A. (2017). Porites superfusa mortality and recovery from a bleaching event at Palmyra Atoll, USA. *PeerJ*, *5*, e3204.

Lyons, M. B., Murray, N. J., Kennedy, E. V., Kovacs, E. M., Castro-Sanguino, C., Phinn, S. R., … Others. (2024). New global area estimates for coral reefs from high-resolution mapping. *Cell Reports Sustainability*, *1*(2).

Marcer, A., Haston, E., Groom, Q., Ariño, A. H., Chapman, A. D., Bakken, T., … Others. (2021). Quality issues in georeferencing: From physical collections to digital data repositories for ecological research. *Diversity and Distributions*, *27*(3), 564–567.

Kusumoto, B., Costello, M. J., Kubota, Y., Shiono, T., Wei, C.-L., Yasuhara, M., & Chao, A. (2020). Global distribution of coral diversity: Biodiversity knowledge gradients related to spatial resolution. *Ecological Research*, *35*(2), 315–326.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., … Others. (2023). Segment anything. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4015–4026.

Marcer, A., Chapman, A. D., Wieczorek, J. R., Xavier Picó, F., Uribe, F., Waller, J., & Ariño, A. H. (2022). Uncertainty matters: ascertaining where specimens in natural history collections come from and its implications for predicting species distributions. *Ecography*, *2022*(9), e06025.

Patil, A., & Garcia-Sanchez, F. (2024, in preparation). Turbulent kinetic energy dissipation in wave boundary layers over artificially generated coral-colonies. *Journal of Geophysical Research: Oceans*.

Mandelbrot, B. B. (1983). The fractal geometry of nature/Revised and enlarged edition. *New York*.

Gröger, G., & Plümer, L. (2012). CityGML--Interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, *71*, 12–33.

Lindhart, M., Monismith, S. G., Khrizman, A., Mucciarone, D., & Dunbar, R. (2021). How consistent are estimates of roughness parameters on a rough coral reef? *Journal of Geophysical Research: Oceans*, *126*(12), e2021JC017825.

Perlin, K. (1985). An image synthesizer. *ACM Siggraph Computer Graphics*, *19*(3), 287–296.

Lotte, R. G., Haala, N., Karpina, M., Aragão, L. E. O. e. C. de, & Shimabukuro, Y. E. (2018). 3D façade labeling over complex scenarios: A case study using convolutional neural network and structure-from-motion. *Remote Sensing*, *10*(9), 1435.

Johnson-Roberson, M., Pizarro, O., Williams, S. B., & Mahon, I. (2010). Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. *Journal of Field Robotics*, *27*(1), 21–51.

Burns, J. H. R., Delparte, D., Gates, R. D., & Takabayashi, M. (2015). Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3D ecological characteristics of coral reefs. *PeerJ*, *3*, e1077.

Kim, J.-S., Kavak, H., & Crooks, A. (2018). Procedural city generation beyond game development. *SIGSPATIAL Special*, *10*(2), 34–41.

Kubicek, A., Muhando, C., & Reuter, H. (2012). Simulations of long-term community dynamics in coral reefs-how perturbations shape trajectories. *PLoS Computational Biology*, *8*(11), e1002791.

Schneekloth, F. (2019). *Design and implementation of a coral reef simulation with focus on scientific value and useablility*. https://cgvr.informatik.uni-bremen.de/theses/finishedtheses/dai_coralreef/Siccom_Bachelorthesis_Fabian_Schneekloth.pdf

Allen Coral Atlas | Atlas. (n.d.). https://allencoralatlas.org/atlas/#1.00/0.0000/-145.0000

Coral Trait Database | Home. (n.d.). https://www.coraltraits.org/

Coral Trait Database | Species. (n.d.). https://www.coraltraits.org/species

CoralNet. (n.d.). https://coralnet.ucsd.edu/

Corals | 3D Digitization. (n.d.). https://3d.si.edu/corals

GBIF. (n.d.). https://gbif.org/

WoRMS - World Register of Marine Species. (n.d.). https://www.marinespecies.org/

APPENDICES

Reproducibility

The datasets used, such as the Smithsonian 3D coral models, GBIF occurrence records, and Allen Coral Atlas geomorphic zones, are publicly available and referenced within the methodology. Python scripts developed for data processing, including taxonomic verification with the WoRMS API and spatial analysis with geopandas can be accessed via a public repository (**https://github.com/GGBRW/CoralGenerator**). Parameters controlling pipeline components, such as the subdivision level for terrain generation and species abundance estimation, are configurable to allow replication of experiments under varying conditions.

Table A.1

MAPPED AREA	AREA (km2)	Percentage of Multi-zone Overlap	ITERATIONS TO CONVERGENCE
Andaman Sea:	4937.334	7.11%	79
Bermuda:	513.907	4.04%	80
Brazil:	422.727	2.34%	41
Central Indian Ocean:	3830.054	7.95%	88
Central South Pacific:	3353.81	8.41%	65
Coral Sea:	927.871	6.32%	83
Eastern Africa & Madagascar:	11000.818	7.39%	96
Eastern Micronesia:	4952.924	6.44%	67
Eastern Papua New Guinea & Solomon Islands:	12326.962	8.42%	70
Eastern Tropical Pacific:	624.47	8.55%	54
Great Barrier Reef and Torres Strait:	14836.014	7.53%	82
Hawaiian Islands:	1633.944	6.12%	71
Mesoamerica:	6874.919	7.02%	62
Northeastern Asia:	1400.629	8.45%	65
Northern Caribbean, Florida & Bahamas:	167666.842	2.07%	41

Northwestern Arabian Sea:	9500.612	4.67%	76
Philippines:	18386.735	8.18%	66
Red Sea & Gulf of Aden:	14714.188	8.18%	63
South China Sea:	1441.654	9.27%	84
Southeast Asian Archipelago:	29574.352	7.23%	50
Southeastern Asia:	4825.253	6.66%	52
Southeastern Caribbean:	3796.823	7.00%	83
Southern Asia:	2594.305	3.14%	51
Southwestern Pacific:	11665.773	7.98%	74
Subtropical Eastern Australia:	145.96	4.12%	59
Timor & Arafura Seas:	10230.219	3.40%	72
Western Africa:	1652.743	4.51%	59
Western Australia:	3451.72	4.48%	75
Western Indian Ocean:	2635.213	8.01%	93
Western Micronesia:	2226.876	7.55%	81

Appendix A.2: Exporting Models as Watertight .stl Files

To export the generated coral reef models as watertight **.stl** files from the Blender project files, follow these steps:

1. Open the exported Blender file that contains the model with separate terrain and coral objects.
2. Select the terrain and each coral object you wish to export. Ensure each object is manifold (i.e., watertight) to guarantee compatibility with .stl format requirements. You can use Blender's built-in 3D Print Toolbox add-on to check for non-manifold edges and repair them if needed:

    - Enable the `3D Print Toolbox` add-on by navigating to *Edit > Preferences > Add-ons*, then search for and activate "3D Print Toolbox."

    - In the 3D Print Toolbox panel, select each object and click *Check All* to identify any non-manifold geometry. Use the repair tools provided in the toolbox to fix any issues.

3. Once all objects are watertight, proceed with exporting:

    - Select the terrain and coral objects.

    - Go to *File > Export > Stl (.stl)*.

    - In the *Export STL* options panel on the left, ensure *Selection Only* is checked to export only the selected objects.

    - Adjust any other export settings as needed (e.g., scale adjustments), then click *Export STL*.

Table A.3

MAPPED AREA + GEOMORPH. ZONE	OVERLAPPING OCCURRENCES
Southeastern Caribbean, Reef Slope:	6053
Southeastern Caribbean, Back Reef Slope:	4573
Great Barrier Reef and Torres Strait, Outer Reef Flat:	4300
Hawaiian Islands, Reef Slope:	4052
Northern Caribbean, Florida & Bahamas, Outer Reef Flat:	2921
Hawaiian Islands, Deep Lagoon:	2621
Southeastern Caribbean, Sheltered Reef Slope:	2412
Great Barrier Reef and Torres Strait, Sheltered Reef Slope:	2394
Hawaiian Islands, Sheltered Reef Slope:	2362
Northern Caribbean, Florida & Bahamas, Back Reef Slope:	2234
Eastern Micronesia, Reef Slope:	2176
Great Barrier Reef and Torres Strait, Deep Lagoon:	2173
Eastern Tropical Pacific, Back Reef Slope:	2158
Eastern Tropical Pacific, Outer Reef Flat:	2075
Eastern Tropical Pacific, Terrestrial Reef Flat:	2061
Mesoamerica, Reef Slope:	1838
Northern Caribbean, Florida & Bahamas, Shallow Lagoon:	1772
Great Barrier Reef and Torres Strait, Reef Slope:	1746
Great Barrier Reef and Torres Strait, Back Reef Slope:	1720
Hawaiian Islands, Plateau:	1545
Great Barrier Reef and Torres Strait, Plateau:	1448
Hawaiian Islands, Inner Reef Flat:	1348
Mesoamerica, Back Reef Slope:	1288
Northern Caribbean, Florida & Bahamas, Sheltered Reef Slope:	1268

Mesoamerica, Deep Lagoon:	1221
Northern Caribbean, Florida & Bahamas, Deep Lagoon:	1039
Hawaiian Islands, Back Reef Slope:	889
Northern Caribbean, Florida & Bahamas, Inner Reef Flat:	873
Great Barrier Reef and Torres Strait, Terrestrial Reef Flat:	872
Southeastern Caribbean, Terrestrial Reef Flat:	868
Eastern Micronesia, Back Reef Slope:	798
Mesoamerica, Outer Reef Flat:	726
Northeastern Asia, Outer Reef Flat:	712
Northern Caribbean, Florida & Bahamas, Plateau:	702
Southeastern Caribbean, Deep Lagoon:	680
Great Barrier Reef and Torres Strait, Inner Reef Flat:	676
Great Barrier Reef and Torres Strait, Reef Crest:	672
Northern Caribbean, Florida & Bahamas, Reef Slope:	637
Hawaiian Islands, Shallow Lagoon:	606
Southeastern Caribbean, Shallow Lagoon:	555
Southeastern Caribbean, Inner Reef Flat:	535
Southwestern Pacific, Reef Slope:	524
Hawaiian Islands, Outer Reef Flat:	522
Mesoamerica, Inner Reef Flat:	443
Western Australia, Outer Reef Flat:	404
Northeastern Asia, Reef Slope:	391
Mesoamerica, Plateau:	390
Timor & Arafura Seas, Inner Reef Flat:	374
Mesoamerica, Terrestrial Reef Flat:	372
Mesoamerica, Shallow Lagoon:	357
Northeastern Asia, Terrestrial Reef Flat:	356
Great Barrier Reef and Torres Strait, Shallow Lagoon:	355
Northeastern Asia, Back Reef Slope:	346
Northeastern Asia, Inner Reef Flat:	337
Mesoamerica, Sheltered Reef Slope:	317

Northern Caribbean, Florida & Bahamas, Reef Crest:	285
Southeastern Caribbean, Plateau:	284
Eastern Micronesia, Outer Reef Flat:	279
Southeastern Caribbean, Outer Reef Flat:	269
Southeast Asian Archipelago, Outer Reef Flat:	246
Western Australia, Shallow Lagoon:	240
Northern Caribbean, Florida & Bahamas, Terrestrial Reef Flat:	218
Subtropical Eastern Australia, Inner Reef Flat:	210
Western Micronesia, Sheltered Reef Slope:	203
Southwestern Pacific, Sheltered Reef Slope:	184
Timor & Arafura Seas, Deep Lagoon:	170
Southwestern Pacific, Back Reef Slope:	168
Southwestern Pacific, Outer Reef Flat:	168
Eastern Micronesia, Reef Crest:	158
Eastern Africa & Madagascar, Back Reef Slope:	155
Philippines, Sheltered Reef Slope:	148
Western Indian Ocean, Outer Reef Flat:	148
Hawaiian Islands, Reef Crest:	143
Northeastern Asia, Sheltered Reef Slope:	143
Southwestern Pacific, Shallow Lagoon:	135
Western Indian Ocean, Back Reef Slope:	127
Southwestern Pacific, Plateau:	123
Western Australia, Back Reef Slope:	121
Timor & Arafura Seas, Reef Slope:	119
Southwestern Pacific, Reef Crest:	114
Bermuda, Deep Lagoon:	110
Western Australia, Inner Reef Flat:	103
Eastern Papua New Guinea & Solomon Islands, Sheltered Reef Slope:	99
Southeast Asian Archipelago, Terrestrial Reef Flat:	99
Timor & Arafura Seas, Back Reef Slope:	97
Southeast Asian Archipelago, Plateau:	94
Timor & Arafura Seas, Outer Reef Flat:	93

Northeastern Asia, Shallow Lagoon:	91
Southeastern Caribbean, Reef Crest:	91
Coral Sea, Shallow Lagoon:	89
Coral Sea, Back Reef Slope:	89
Western Australia, Reef Slope:	89
Southwestern Pacific, Terrestrial Reef Flat:	88
Southwestern Pacific, Inner Reef Flat:	86
Southeast Asian Archipelago, Deep Lagoon:	83
Timor & Arafura Seas, Terrestrial Reef Flat:	82
Western Micronesia, Reef Slope:	70
Eastern Africa & Madagascar, Inner Reef Flat:	67
Red Sea & Gulf of Aden, Back Reef Slope:	67
Eastern Papua New Guinea & Solomon Islands, Reef Crest:	60
Andaman Sea, Sheltered Reef Slope:	59
Western Indian Ocean, Inner Reef Flat:	59
Western Indian Ocean, Sheltered Reef Slope:	59
Western Micronesia, Terrestrial Reef Flat:	59
Philippines, Inner Reef Flat:	58
Western Micronesia, Outer Reef Flat:	58
Southeast Asian Archipelago, Reef Slope:	55
Western Indian Ocean, Reef Slope:	54
Eastern Africa & Madagascar, Reef Slope:	53
Red Sea & Gulf of Aden, Reef Slope:	53
Red Sea & Gulf of Aden, Outer Reef Flat:	51
Timor & Arafura Seas, Shallow Lagoon:	51
Central Indian Ocean, Inner Reef Flat:	50
Southeast Asian Archipelago, Inner Reef Flat:	50
Central Indian Ocean, Reef Slope:	46
Coral Sea, Reef Crest:	46
Southeast Asian Archipelago, Sheltered Reef Slope:	45
Eastern Micronesia, Plateau:	44
South China Sea, Plateau:	44
Western Indian Ocean, Shallow Lagoon:	42
Central South Pacific, Outer Reef Flat:	41

Southwestern Pacific, Deep Lagoon:	41
Western Indian Ocean, Terrestrial Reef Flat:	39
Northeastern Asia, Deep Lagoon:	38
Central Indian Ocean, Shallow Lagoon:	37
Eastern Micronesia, Shallow Lagoon:	37
Southeast Asian Archipelago, Reef Crest:	37
Southeastern Asia, Reef Slope:	35
Eastern Africa & Madagascar, Deep Lagoon:	34
Western Australia, Sheltered Reef Slope:	34
Western Micronesia, Back Reef Slope:	34
Eastern Africa & Madagascar, Shallow Lagoon:	33
Southeast Asian Archipelago, Back Reef Slope:	33
Northwestern Arabian Sea, Back Reef Slope:	31
Philippines, Terrestrial Reef Flat:	31
Coral Sea, Reef Slope:	28
Brazil, Back Reef Slope:	27
Philippines, Outer Reef Flat:	26
Southeast Asian Archipelago, Shallow Lagoon:	25
Philippines, Back Reef Slope:	24
Northeastern Asia, Reef Crest:	23
Southeastern Asia, Sheltered Reef Slope:	23
Eastern Papua New Guinea & Solomon Islands, Deep Lagoon:	22
Red Sea & Gulf of Aden, Deep Lagoon:	22
Western Australia, Patch Reefs:	22
Southeastern Asia, Terrestrial Reef Flat:	21
Eastern Tropical Pacific, Reef Slope:	20
Western Micronesia, Reef Crest:	20
Central Indian Ocean, Back Reef Slope:	18
Central South Pacific, Inner Reef Flat:	17
Eastern Papua New Guinea & Solomon Islands, Outer Reef Flat:	17
Subtropical Eastern Australia, Terrestrial Reef Flat:	17
Bermuda, Inner Reef Flat:	16
Andaman Sea, Inner Reef Flat:	15

Bermuda, Shallow Lagoon:	15
Central South Pacific, Deep Lagoon:	15
Eastern Africa & Madagascar, Terrestrial Reef Flat:	15
Mesoamerica, Reef Crest:	15
Subtropical Eastern Australia, Reef Slope:	14
Western Micronesia, Deep Lagoon:	14
Central South Pacific, Terrestrial Reef Flat:	13
Eastern Africa & Madagascar, Outer Reef Flat:	13
Central South Pacific, Shallow Lagoon:	12
Western Indian Ocean, Deep Lagoon:	12
Bermuda, Outer Reef Flat:	11
Central South Pacific, Back Reef Slope:	11
Northwestern Arabian Sea, Shallow Lagoon:	11
Southeastern Asia, Outer Reef Flat:	11
Southern Asia, Shallow Lagoon:	11
Western Australia, Terrestrial Reef Flat:	11
Central Indian Ocean, Deep Lagoon:	10
Central South Pacific, Reef Slope:	10
Eastern Papua New Guinea & Solomon Islands, Terrestrial Reef Flat:	10
Timor & Arafura Seas, Reef Crest:	10
Western Indian Ocean, Reef Crest:	10
Central Indian Ocean, Outer Reef Flat:	9
Red Sea & Gulf of Aden, Inner Reef Flat:	9
Timor & Arafura Seas, Plateau:	9
Western Australia, Plateau:	9
Central South Pacific, Sheltered Reef Slope:	8
Red Sea & Gulf of Aden, Shallow Lagoon:	8
Western Africa, Sheltered Reef Slope:	8
Northeastern Asia, Plateau:	7
Andaman Sea, Terrestrial Reef Flat:	6
Andaman Sea, Deep Lagoon:	6
Bermuda, Back Reef Slope:	6
Eastern Papua New Guinea & Solomon Islands, Plateau:	6
Hawaiian Islands, Terrestrial Reef Flat:	6

Southern Asia, Reef Slope:	6
Andaman Sea, Back Reef Slope:	5
Eastern Micronesia, Sheltered Reef Slope:	5
Red Sea & Gulf of Aden, Terrestrial Reef Flat:	5
Red Sea & Gulf of Aden, Plateau:	5
Red Sea & Gulf of Aden, Reef Crest:	5
Northwestern Arabian Sea, Plateau:	4
Philippines, Shallow Lagoon:	4
Southeastern Asia, Inner Reef Flat:	4
Subtropical Eastern Australia, Outer Reef Flat:	4
Subtropical Eastern Australia, Back Reef Slope:	4
Western Micronesia, Inner Reef Flat:	4
Central South Pacific, Reef Crest:	3
Coral Sea, Inner Reef Flat:	3
Eastern Africa & Madagascar, Sheltered Reef Slope:	3
Eastern Africa & Madagascar, Reef Crest:	3
Eastern Micronesia, Deep Lagoon:	3
Eastern Papua New Guinea & Solomon Islands, Back Reef Slope:	3
Philippines, Plateau:	3
Red Sea & Gulf of Aden, Sheltered Reef Slope:	3
South China Sea, Outer Reef Flat:	3
Southeastern Asia, Back Reef Slope:	3
Brazil, Outer Reef Flat:	2
Central Indian Ocean, Sheltered Reef Slope:	2
Eastern Micronesia, Inner Reef Flat:	2
Eastern Papua New Guinea & Solomon Islands, Shallow Lagoon:	2
Eastern Tropical Pacific, Plateau:	2
Eastern Tropical Pacific, Shallow Lagoon:	2
Northwestern Arabian Sea, Reef Slope:	2
Philippines, Patch Reefs:	2
Southern Asia, Inner Reef Flat:	2
Subtropical Eastern Australia, Reef Crest:	2

Timor & Arafura Seas, Sheltered Reef Slope:	2
Western Australia, Reef Crest:	2
Western Micronesia, Plateau:	2
Western Micronesia, Shallow Lagoon:	2
Andaman Sea, Reef Slope:	1
Andaman Sea, Outer Reef Flat:	1
Andaman Sea, Plateau:	1
Northwestern Arabian Sea, Deep Lagoon:	1
Philippines, Deep Lagoon:	1
South China Sea, Back Reef Slope:	1
Southern Asia, Outer Reef Flat:	1
Subtropical Eastern Australia, Shallow Lagoon:	1

Pseudocode A.1

```
// Define CoralSpecies class
class CoralSpecies:
    name
    radius
    clustering_factor
    color

// Function to generate colony positions with clustering behavior
function generate_position_clustering(colonies, species, clustering_factor,
colony_radius, max_attempts):
    attempts = 0
    repeat until position is valid or attempts exceed max_attempts:
        if random chance < clustering_factor:
            // Choose an existing colony of the same species
            base_colony = select random colony of same species
            calculate position around base_colony within a certain distance
        else:
            // Generate a random position
            x = random value within terrain limits
            y = random value within terrain limits

        if position (x, y) is valid:
            return position (x, y)

    // Fallback: return random position if clustering fails
    return random valid position

// Function to check if a colony overlaps with others
function check_overlap(existing_colonies, new_x, new_y, new_radius):
    for each colony in existing_colonies:
        if distance between (new_x, new_y) and colony is less than allowed
radius:
```

```
            return true  // overlap detected
    return false  // no overlap

// Function to place coral colonies on the terrain
function place_coral_colonies(species_list, colony_counts):
    initialize empty list of colonies
    for each species in species_list:
        for number of colonies to place:
            repeat until valid position is found:
                (x, y) = generate_position_clustering(existing_colonies,
species, clustering_factor, colony_radius)
                if no overlap with other colonies:
                    create new colony with position (x, y)
                    add colony to the list

    return list of colonies

// Function to visualize the terrain and colonies
function visualize_terrain(colonies):
    for each colony in colonies:
        draw circle at (colony.x, colony.y) with radius and color
    display the terrain

// Main logic
species_list = define list of CoralSpecies
colony_counts = define list of how many colonies per species
colonies = place_coral_colonies(species_list, colony_counts)
visualize_terrain(colonies)
```