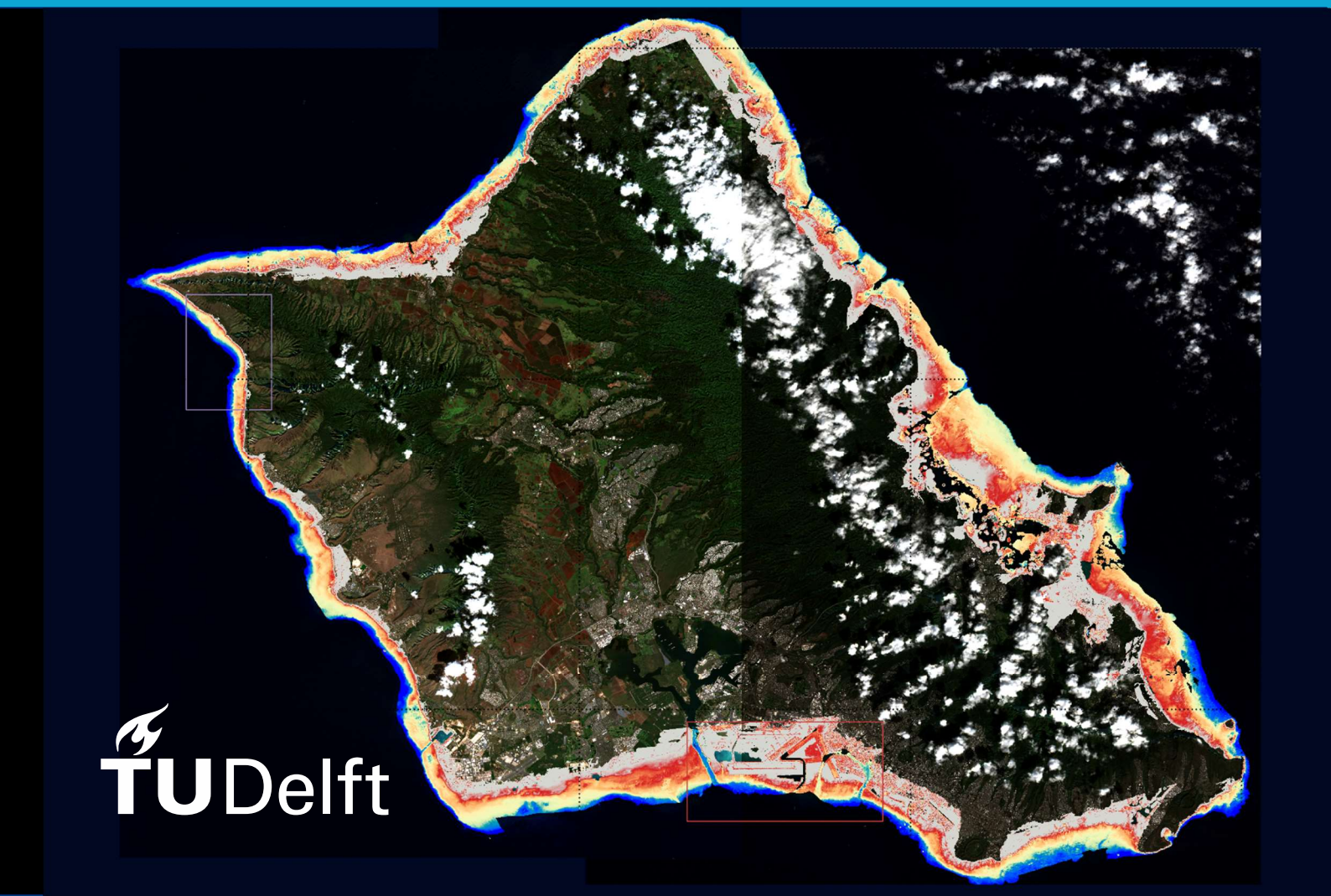
MSc thesis in Geomatics

Convolutional Neural Networks for Satellite-Derived Bathymetry

Yustisi Ardhitasari Lumban Gaol

2021



TU Delft

MSc thesis in Geomatics

Convolutional Neural Networks for Satellite-Derived Bathymetry

Yustisi Ardhitasari Lumban Gaol

2021

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

Cover image shows a Satellite-Derived Bathymetry model in Oahu island, Hawaii. The SDB was calculated from three Sentinel-2 Level-2A images using CNN. Sample data for training was acquired from two areas in the boxes. The depths are visualized in the red to blue colour gradations, indicating coastal depths in the range of 0 to 20 m.

An electronic version of this thesis is available at http://repository.tudelft.nl and https://github.com/yustisiardhitasari/sdbcnn.git

The work in this thesis was carried out in the:

3D geoinformation group
Delft University of Technology

Supervisors:	Dr. Ken Arroyo Ohori
	Dr. Ravi Ylan Peters
Co-reader:	Dr. Alireza Amiri Simkooei

This research was supported by:

Indonesia Endowment Fund for Education

Abstract

Bathymetric depth data for shallow water regions is essential for coastal management and research purposes. Even combined measurements using echo sounding and Light Detection and Ranging (LiDAR) leave data gaps due to vessels unable to reach nearshore waters, or the green laser not being able to penetrate certain areas, yielding errors. Also, they have high operational costs and require extensive acquisition time. Satellite-Derived Bathymetry (SDB) is an alternative to extract shallow water depths that is able to overcome these problems using multispectral imagery. There are two approaches of SDB: analytical and empirical. The analytical method requires several water properties and assumptions, which might not be known. The empirical method relies on the linear relationship between reflectances and in-situ depths, but the relationship may not be entirely linear due to bottom types variation, water column effect, and noise. Machine learning approaches have been used to address nonlinearity, but it treats pixels independently, whereas there is a spatial correlation that should also influence SDB computations since adjacent pixels are correlated in terms of depth. This characteristic of the local connectivity can be detected by Convolutional Neural Networks (CNN). Therefore, this thesis conducts a study of SDB using CNN.

This research focuses on the following questions: (i) what kind of preprocessing is needed for the data sets; (ii) what kind of CNN architecture can be used; (iii) what is the accuracy of the method; and (iv) to what extent the pretrained model in certain areas can be reused in other areas. In order to represent a variety of reflectance and depth, this study chooses six areas of interest in three different coastal regions as follows: (i) Ponce, Lajas, and San Juan represent Puerto Rico; (ii) Key West represents the Florida Keys; and (iii) Makua and Honolulu represent Hawaii. These locations have different water conditions regarding their bottom types, turbidity, and water column properties.

With several CNN configurations, the optimum accuracy is obtained using three convolutional layers, a window size of 9×9, and the RGBNSS bands. The accuracy of SDB using the CNN approach outperforms the linear transform method while using less training data. CNN also has better accuracy than several previous studies using different methods (the ratio transform, Random Forest, and the radiative transfer). The use of NIR and SWIR bands improves the accuracy, especially in deeper areas. In general, the accuracy of SDB decreases as the depths increases because the proportion of water bottom reflectances decreases with the increasing depth and causes their relationship to become less clear. SDB has a lower accuracy in more turbid water since reflectances are disturbed by the particles that cause the turbidity. Comparisons between different image preprocessing indicate similar SDB accuracy, or even slightly better. This result indicates another benefit of CNN: removing the need to preprocess images since suitable corrections can be automatically performed by CNN given adequate training.

The use of multi-temporal images enhances the variety of training data and makes CNN more reliable, remarked by improved accuracy. However, data variation should be equally distributed to avoid abnormality when computing SDB in a particular image. SDB using the transfer model has several limitations at particular depths or when implemented to a different water condition, making the coastal water characteristics considered when reusing a pretrained model from one area to another.

In summary, surface reflectances are sufficient to be used as input data for CNN training. So, no additional image preprocessing and features specification is needed to define the predictors. CNN can produce better SDB accuracy than several other methods. The accuracy improves by increasing the variety of training data. However, SDB using the transfer model still need to be further investigated. A thorough identification of the proportion of sample data is needed to obtain balanced training data. In this way, it is more likely to produce a more reliable and more stable CNN model for extracting shallow water depths in the new data.

Acknowledgements

I would like to express my sincere gratitude to my supervisors at TU Delft, Ken Arroyo Ohori and Ravi Peters, for their continuous advice, suggestions, guidance, patience, and support during this graduation project. In addition, I would like to thank my co-reader, Alireza Amiri Simkooei, for his constructive feedback and useful questions. I would like to extend my gratitude to BIG for allowing me to continue studying and the Indonesian Endowment Fund for Education (LPDP) for the scholarship that allowed me to conduct this Master's degree.

Contents

List of Figures

List of Tables

Acronyms

1 Introduction

1.1 Background

Coastal zones cover many unique habitats and provide critical human needs. Coastal habitats include estuaries, coastal wetlands, seagrass meadows, coral reefs, mangrove forests, kelp forests, and upwelling areas [EPA, 2018]. They support economic and social activities, especially for people living in coastal communities which represent 37% of the global population in 2017 [UN, 2017]. Hence, to understand coastal zones' characteristics, it is necessary to hold an inventory of the coastal resources, including coastal water depth, which is relatively shallow.

Shallow water depth information is essential to understand the dynamics of the water bed contour characteristics. This data is essential for coastal management and research; including nautical chart determination for safety navigation, coastal spatial planning, onshore buildings and ports construction planning, fishing industry, and coastal disaster mitigation. Water depth data can be obtained through a bathymetry survey. Bathymetry is the study of the floors of the oceans, seas, lakes, rivers, and other areas of water. Bathymetry surveying reveals the terrain of water beds. Figure 1.1 illustrates several bathymetry survey methods.

The most traditional method (leadline) measures depth manually using ropes and weights attached to them [NOAA, 2017b]. The ropes are lowered until they touch the bottom of the water. This method leaves huge data gaps due to sporadic sampling and thus misses many details in between. With the development of shipboard echo sounders, water depth is easier to measure. The echo sounding method produces accurate depth values using a transducer installed in the ship. The transducer transmits a signal into the water and records the time interval until the signal returns. Depth is obtained by multiplying half of the time interval by the speed of sound in the water. The transducer measures the depth as the ship moves, creating the sounding line. Single Beam Echo Sounding (SBES) still generates gaps between sounding lines, while MBES can measure denser depth values. However, these methods have high operational costs and require extensive measurement time, and the survey vessels cannot reach nearshore water areas [Gao, 2009]. The vessels' incapability to measure very shallow water depth, around 5 m below sea level, will cause data gaps along the shoreline area.

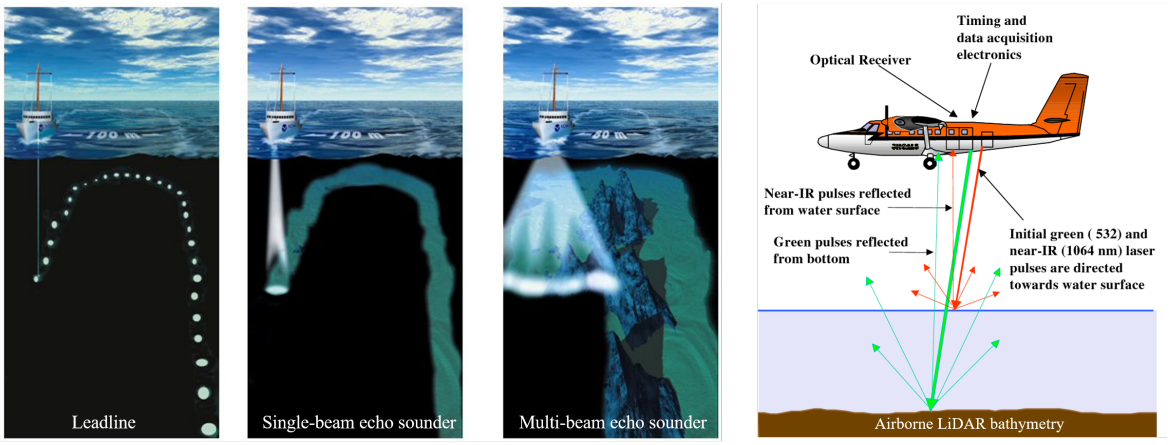


Figure 1.1: Different bathymetry survey techniques [NOAA, 2016; LaRocque and West, 1999].

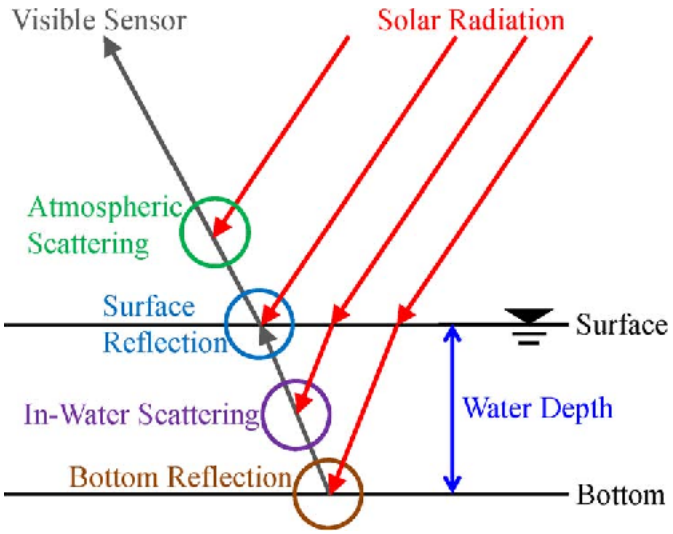


Figure 1.2: Light attenuation in shallow water [Kanno et al., 2012].

Another survey method is airborne LiDAR bathymetry, where a green laser is used, which is able to penetrate shallow water columns over clear water areas. LiDAR surveys usually combine bathymetry with topography, the study of terrain surfaces on land. They are suitable for rapid mapping, e.g. post-disaster events, due to the ability to measure coastline areas in a short time. Although bathymetric LiDAR can produce accurate water depth data, the survey plan must consider numerous factors, including the weather for flying, air traffic controls, waves, tides, ground control accessibility, turbidity, and seafloor type [Quadros, 2016]. Not accounting for these factors could lead to an error or inability to capture a certain area in the measurements. Also, the operational cost is very high, so measurements using this method are still infrequent in large areas.

Since those survey methods are inefficient to measure large water areas, especially in shallow water, optical remote sensing images are a promising alternative data source to extract water depth information. The availability of satellite imagery worldwide with different spatial and temporal resolution data sets makes it popular in various applications, including bathymetric modelling. SDB is a way to model water depth in shallow water areas using multispectral imagery. The basic principle of SDB is based on the attenuation of light in the water column, which is a function of wavelength (Figure 1.2). Shorter wavelengths attenuate less than longer wavelengths. Longer wavelengths penetrate the water column and interact with the water bed before the signal arrives at the sensor. Once the light touches the water bed, the bottom reflectance value can be obtained from satellite images after removing other components in the atmosphere, water surface, and water column. The bottom reflectance is then used to extract water depth values.

SDB has been known to efficiently fill data gaps that occur with the survey method based on echo sounding. SDB is also a low-cost technique since remote sensing images are used instead of field surveys. Even more, it has few environmental impacts and risks to personnel or equipment since the model can be derived without accessing sea water areas directly. Also, water depth information can be generated relatively quickly compared to in-situ surveys. Because of this, many researchers continuously study SDB to improve the accuracy of the model and make it more efficient.

1.2 Problem definition

Although SDB has been often used to produce seamless bathymetry data, it requires in-situ data in the training process to determine model coefficients which represent the relationship between reflectance and depth. This can be a major limitation of SDB, since shallow water in-situ data is not available for many areas. Many studies focus on improving the accuracy of the model, but they still require in-situ

coverage. At the same time, LiDAR bathymetry data, which is available in several locations exposes the relationship between depth and reflectance value in more detail than echo-sounding data. It can be used to study the transfer model, where a pre-trained model is reused outside in-situ coverage, so the creation of SDB models without in-situ data has become possible.

A number of different SDB techniques have been applied and evaluated. Two well-known methods exist, which are linear transform [Lyzenga, 1985; Lyzenga et al., 2006] and ratio transform [Stumpf et al., 2003]. Much research on SDB uses these methods [Hamylton et al., 2015; Kabiri, 2017; Traganos et al., 2018]. Some studies modified Lyzenga85's method to improve the accuracy [Lyzenga et al., 2006; Kanno et al., 2012; Kanno and Tanaka, 2012; Kanno et al., 2013], or used a different regression technique, such as Geographically Weighted Regression [Vinayaraj et al., 2016]. These methods rely on a linear relationship between water reflectance and depth, while it may not be entirely linear due to bottom types variation and noise. Available spatial regression techniques are capable of addressing the spatial heterogeneity, but they cannot undertake the bottom variations. Existing machine learning techniques for SDB, such as Random Forest [Manessa et al., 2016; Sagawa et al., 2019; Tonion et al., 2020] and Support Vector Machines [Misra et al., 2018; Manessa et al., 2018; Tonion et al., 2020], can learn nonlinearity due to bottom variations, but are still not capable of considering the spatial correlation. Based on literature reviews, there are two critical factors to be considered when building an SDB model: bottom type variations and the spatial correlation between pixels. Bottom type variations cause the non-linearity since different objects underwater response the light differently and thus, the reflectance differ over the same depth. Meanwhile, the spatial correlation assumes that adjacent pixels have relationship to depths. Therefore, research on another method that considers both factors is needed and will be carried out in this thesis.

1.3 Objectives and research questions

This research aims to develop convolutional neural networks that extract accurate water depth in shallow water areas. In order to evaluate the results, their accuracy will be compared to another method that is well known. This research will use open data for both satellite imagery and in-situ depth, which will be used for training and evaluation, to be used for further study. The research objective has been used to formulate the main research question:

To what extent can convolutional neural networks be used for accurate shallow water depth extraction using Sentinel-2 satellite images?

To answer the main research question, the following sub-questions are deemed relevant:

- What kind of preprocessing is needed for the data sets?
- What kind of CNN architecture can be used for SDB?
- What is the accuracy of the method (compared to the linear transform method, in different depth ranges, in different locations)?
- To what extent can the pretrained model be reused?

1.4 Scope and limitations

This project will focus on shallow water depth extraction, limited to the depth range from 0 to 20m. There will be a minimum of two different locations. This project will use open data sets and software. Sentinel-2 Level 2A will be used as the source for input images. As in-situ depth data, this thesis will use bathymetry acquired from LiDAR and MBES. Convolutional neural networks will be the method to be researched and another well-known method, e.g. linear transform, will be used for comparison and evaluation of the results. When the algorithm achieves a satisfactory result, that is when the accuracy is better than the other method, the project may study the transfer learning possibilities by reusing the pre-trained model to a new location.

1.5 Outline

The report is structured as follows;

Chapter 1 briefly describes the background, problems, objective and research questions, as well as scope and limitations.

Chapter 2 provides an overview of related work on SDB, including a brief description of several existing methods. An introduction to CNN, reflectance disturbance in coastal waters, as well as satellite image products, is presented. This chapter is intended to provide sufficient theoretical foundation required to understand the following chapters.

Chapter 3 explains the study areas, data sets, and the workflow from data acquisition to the evaluation of the SDB model.

Chapter 4 demonstrates the establishment of CNN architecture and the accuracy of the SDB model in different depth ranges, locations, the temporal aspect of satellite images, and the transfer model.

Chapter 5 gives a summary of the main results, conclusions, and recommendations for future work.

2 Literature review

This chapter provides an overview of the theoretical background related to this thesis. Section 2.1 describes the existing methods on SDB in general, alongside a review on related work. Then, Section 2.2 describe general concepts of CNN. Subsequently, Section 2.3 briefly discusses several types of factors affecting the calculations of SDB in reflectance values on coastal waters. An explanation of Sentinel-2 image products together with its related application for SDB are provided in Section 2.4.

2.1 Related work

Since its initial development, many algorithms have been developed to improve the accuracy of SDB. Algorithms such as a physical-based or radiative transfer model, simple or multi-linear regression, spatial regression, and machine learning have been studied to generate a shallow water depth model. These algorithms can be classified as analytical and empirical. The analytical methods, known as the radiative transfer model, are based on light penetration in water. It requires several optical properties of shallow water region such as the attenuation coefficient, backscatter coefficient, coefficient of suspended and dissolved materials, and bottom reflectance [Spitzer and Dirks, 1986; Gao, 2009].

In comparison, the empirical methods find the relationship between reflected radiation and in-situ depth empirically without considering light transmission in water. The empirical methods assume that the total water reflectance is more correlated with depth than water column interaction. Most of the developed algorithms are based on the empirical method since it does not require a lot of physical parameters of water as in the analytical method. Many researchers use a combination of both methods [Lyzenga et al., 2006; Kanno and Tanaka, 2012; Vinayaraj et al., 2016]. The combined method uses an empiric process to derive the predictors and to extract water depth.

2.1.1 General SDB methods

SDB has been introduced by Polcyn et al. [1970] where depth values are retrieved based on the ratio of reflected radiation in at least two visible spectral bands. This method assumes that the ratio of reflectance values between the two bands are the same over different water bed types. This method was then modified by Lyzenga [1978] by including the scattering effect of the water, but the algorithm became more complex and challenging to implement due to the many parameters that are needed. Lyzenga [1985] proposed a **linear transform** method where a transformation was used to linearize the relationship between the transformed radiance and the water depth. The bottom reflectance can be obtained using deep water pixels, assuming no bottom reflectance in the deep water area. It can be expressed as

$$X(\lambda)_i = log(L(\lambda)_i - mean(L_\infty(\lambda)_i)) \tag{2.1}$$

where $X(\lambda)_i$ is transformed radiance of band λ at ith point, $L(\lambda)_i$ and $L_\infty(\lambda)_i$ are the pixel values in shallow and deep water respectively. Then, linear regression is used to calculate the coefficients that are needed to retrieve the depth. With two different bands, the equation can be written as follows.

$$D = \beta_0 + \beta_1 * X(\lambda)_1 + \beta_2 * X(\lambda)_2 \tag{2.2}$$

where D is the in-situ depth for calibration, β_0 is the intercept, β_1 and β_2 are the coefficient of regression. This linear transform method showed that the use of more bands in the regression could provide better results; the accuracy improved up to 0.20 m when the NIR band was added to the red, green, and

blue combination [Vinayaraj, 2017]. However, it did not consider the bottom reflectance variations, and hence the calculated coefficients only apply locally; different areas might end up with different results. Then, Stumpf et al. [2003] proposed a new approach called **ratio transform**.

$$D = \beta_0 + \beta_1 * \frac{log(nL(\lambda)_b)}{log(nL(\lambda)_g)} \tag{2.3}$$

This method relied on the assumption that the ratio between blue and green is the same everywhere. So, it compensates for the variability of the bottom type. This method only needs two coefficients. Nevertheless, the ratio between bands is noisy, and it causes this method cannot always calculate water depths well.

Afterwards, much research on SDB applied the linear and ratio transform methods, e.g. Hamylton et al. [2015]; Kabiri [2017]; Traganos et al. [2018], to different locations. Some of them modified the Lyzenga85' method [Lyzenga, 1985] to improve its accuracy [Lyzenga et al., 2006; Kanno et al., 2012; Kanno and Tanaka, 2012]. Kanno et al. [2012] stated that the improved methods had better accuracy when sufficient training data were available. It becomes a drawback when the data is limited. Others used different regression techniques to address the spatial heterogeneity into the model, such as Geographically Weighted Regression [Vinayaraj et al., 2016]. However, the calibration process for this method is computationally extensive and overshooting/undershooting of prediction becomes a major problem since the model depends on the radius used to compute the coefficients of regression. A larger radius will produce a more general prediction with fewer variations than the smaller one. Consequently, this method is limited by the number of data points or areas.

2.1.2 Machine learning for SDB

The different regression techniques mentioned above only focus on the linear relationship between water reflectance and depth. However, the variation of bottom types and noise in the satellite images cause the relationship to be not entirely linear. As alternatives, some studies have started to implement machine learning for SDB to consider the non-linear relationship. Machine learning has been known to be able to learn a complicated relationship between input and target variables [Auret and Aldrich, 2012; Sarker, 2021]. In the SDB case, machine learning therefore needs to learn the relationship between water reflectance and depth to fit the model. In the last five years, machine learning approaches such as Random Forest [Manessa et al., 2016; Sagawa et al., 2019; Tonion et al., 2020] and Support Vector Machine [Misra et al., 2018; Manessa et al., 2018; Tonion et al., 2020] have been used and showed promising results, but their spatial extent is limited to locations that have in-situ coverage of depths acquired through other methods. Those methods require tuning several parameters and generating many features to make the algorithm learn better. Also, they do not consider the spatial component. In SDB, spatial correlation influences the model results, assuming that pixels that are close together have a particular reflectance pattern that is linked to the depth. This characteristic of the local connectivity can be captured by CNN.

Although CNN has been used in image analysis, it has been implemented mostly for classification purposes. In this thesis, CNN is used for regression purposes. There is only a limited number of literature that studies CNN for regression. Sonogashira et al. [2020] used CNN to generate a high-resolution bathymetry map from low-resolution one. This study used CNN to automatically learn the geometric features of ocean floors and recover their details. As for SDB, CNN was firstly used in Ai et al. [2020] to model shallow water depth using three different satellite images with a spatial resolution ranging from medium (8 m) to high resolution (2 m). The result showed that CNN could retrieve water depth with the same resolution as the input images. Ai et al. [2020] used a single convolution layer with batch normalization. Unfortunately, the study contains no further descriptions regarding other parameters. The study also used restricted data, so it cannot be reproduced, and the data used may not be representative enough. A comprehensive study on the feasibility of CNN to extract shallow water depth from satellite images is needed to find out whether CNN is capable of producing a robust SDB model or not.

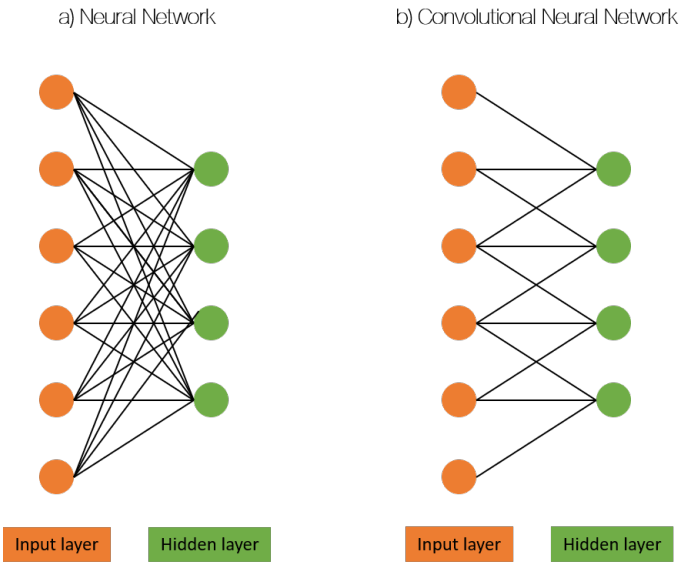


Figure 2.1: An illustration of two variant of layers.

2.2 Convolutional Neural Networks

CNN is a promising approach in machine learning when applied to image analysis. The regular Neural Networks (NN) do not scale well to full images because the networks end up with an enormous number of parameter weights. For example, a small image with dimensions of 200x200x3 needs 120,000 weights to train in a simple NN. Therefore, the alternative approach is to use a convolution layer. As shown in Figure 2.1b, instead of having a fully connected layer (Figure 2.1a) where every input neuron is connected to all the hidden neurons, a neuron only is connected to a specific region of interest (usually nearby pixels) in the layer before it. By doing this, the number of connections are enormously reduced. A typical CNN architecture uses pixel values as the input layer, followed by a combination of convolutional layers, activation functions, pooling layers, and fully-connected layers (Figure 2.2).

There are various architectures of CNN that have been successfully used for classification and segmentation tasks. A classical classification network, AlexNet [Krizhevsky et al., 2012], is considered as a leading architecture in computer vision. AlexNet contains five convolutional layers, followed by three fully connected layers. It has different kernel sizes in each layer, which are 11×11, 5×5, and 3×3. Data augmentation and dropout layers were applied to reduce overfitting. Another architecture, ResNet [He et al., 2016], used shortcut connections to improve the accuracy of a considerably deep layer network. Mulder [2020] used different CNN architectures for segmentation, including U-Net [Ronneberger et al., 2015]. U-Net used a contracting path with typical convolution layers, followed by an expansive path with upsampling and convolution layers. Based on those studies, classification and segmentation have a diverse series of CNN architectures available that have been shown to be able to predict various image objects.

Wang et al. [2021] provides an interactive tool[1] to visualize how every layer in CNN works. It describes the basic CNN architecture, which is the convolve operation, Rectified Linear Unit (ReLU) function, and max-pooling. A general overview of some components in CNN architecture is discussed here.

Input layer. The input layer takes an image with a particular dimension: width, height, and depth. For the input, the depth is the number of channels that are used to train.

Convolutional layer. The convolutional layer applies a convolutional filter that translates only a small region of every pixel and its neighbourhood. The computation is based on a dot product between the filter weights and a small region connected to the input volume. The output has the same width and

[1] https://poloclub.github.io/cnn-explainer/

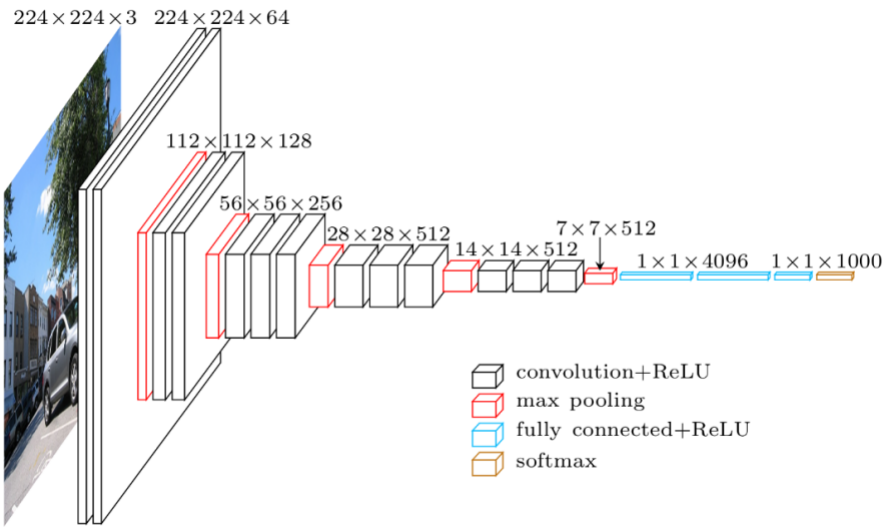


Figure 2.2: A typical CNN architecture for classification [Simonyan and Zisserman, 2015].

height as the input, and depth refers to the number of filters used. A convolutional layer requires a set of parameters, such as the number of filters, filter extent (dimensions to convolve), stride (how many steps to skip), padding (expanding the volume with zeroes to keep the original volume).

Activation function. The activations introduce non-linear transformation to the network. There are several activations available, but the most common function is ReLU. It applies a function $f(x) = max(0, x)$ to each pixel with a threshold of zero. The minimum value is zero, so negative values are converted to zero and then deactivated. ReLU is computationally efficient as few neurons are generally used per time. However, due to the lower threshold, the negative elements are saturated as all weights will not be updated during back propagation. Activation is also needed at the end of the network to compute the final output. For classification tasks, a sigmoid is usually used. It computes the probability of a pixel belonging to a certain class. Meanwhile, to obtain continuous values in the regression task, linear activation is employed.

Pooling layer. A pooling layer performs a downsampling operation along the spatial dimensions, width and height, resulting in a volume with smaller dimensions than the original input. Because of this, it also reduces the number of parameters and computations in the network.

Fully-connected layer. A fully connected layer works as a simple NN where each neuron has a connection to all neurons in the previous layer. The dimensions are 1x1xN, where N is the number of desired outputs, e.g. the number of classes.

Model regularization. In addition to the previous architectures, dropout and batch normalization may be added to perform regularization in order to avoid overfitting. The dropout layer randomly omits a certain proportion of neurons. Batch normalization reduces the internal covariate shift due to the randomness in the input data as well as the parameter initialization. Adding batch normalization and dropout layer into the CNN network may produce a stabler model.

2.3 Reflectance disturbances on coastal waters

When light strikes matter, three forms of interaction may occur as shown in Figure 2.3: reflection, absorption, or transmission. The degree to which these interactions occur depends on different parameters: the wavelength, type of matter, and incidence angle. As shown in Figure 1.2, in a shallow water area, the light captured by the sensor interacts with diverse substances in the air, water surface, water

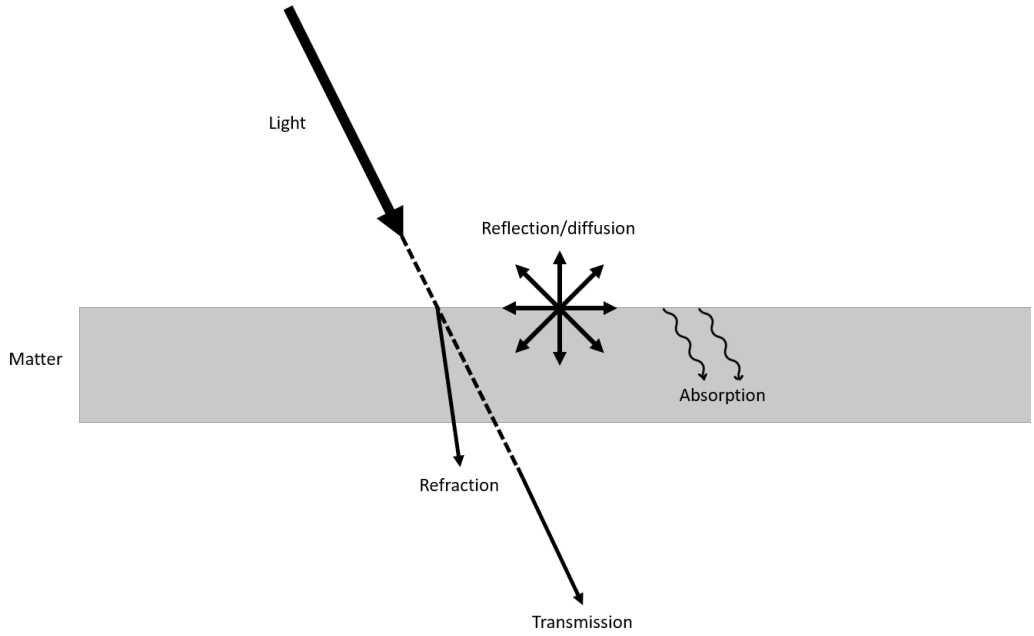


Figure 2.3: An illustration of interaction between light and matter.

column, and bottom waters. Table 2.1 provides a summary of several factors that affect the energy of light reaching the sensor, followed by a description of each layer.

Table 2.1: Factors affecting the energy reaching the sensor.

Factor	Layer	Interaction
Atmospheric contents	Atmosphere	Scattering or absorption
Cloud	Atmosphere	Scattering
Sun elevation angle	Water surface	Specular reflection
Dissolved organic matter	Water column	Absorption
Suspended sediments	Water column	Scattering
Phytoplankton / algae	Water column	Scattering or absorption
Coral	Water bottom	Absorption
Sand	Water bottom	Scattering

Atmosphere. The atmosphere contains water vapour, ozone, nitrogen, oxygen, carbon dioxide, and even dust. The interaction of light with the atmosphere diffuses the radiation and cause a scattering effect or diffusion. The scattering depends on the size of particles and the wavelength. Small particles, such as nitrogen, oxygen, or dust, scatter the energy more in the shorter than longer wavelengths, whereas fog and clouds equally scatter the visible bands. In thick clouds, light is completely blocked, so the energy intercepted by the sensor is the reflection from the cloud. Besides that, absorption by particles such as water vapour, carbon dioxide, and ozone in the atmosphere results in a loss of energy, which diminishes the energy reaching the water surface. The level of absorption depends on the wavelength. The spectral bandwidths of the satellite are set to capture the wavelengths which pass the atmosphere.

Water surface. When light strikes the water surface, its energy is specularly reflected, absorbed, and backscattered to the surface, making it available for remote detection. The specular reflection phenomenon, when the water surface is mirror-like, is called sun glint. The amount of energy that is reflected depends on the sun-elevation angle, where the highest percentage occurs when the angle equal to 0^o (horizon) and lowest when the angle equal to 90^o [Moore, 1980]. The sun glint effect causes a bright pixel in the image. Since most of the energy is reflected on the water surface, reflectances from

underwater becomes less clear.

Water column. The light that is not specularly reflected on the water surface continues to descend into the water column and is affected by absorption and scattering. The water colour and turbidity are two sources that affect the energy received by the sensor. They primarily affect the visible and NIR wavelengths. A dark water colour diminishes the energy received by the sensor because more of the light energy is absorbed in the water column. Some wavelengths are absorbed more than others. For example, NIR's reflectance in clean, deep water is almost zero because the water column absorbs almost all NIR light. On the contrary, turbidity increases the energy reaching the sensor because the particles in the water column backscatter more light. Shallow water with high turbidity causes the bottom water's reflectance to be obscured by the particles that produce turbidity. Meanwhile, phytoplankton in coastal waters with a clear condition, less obstruction occurs in the water column, so the reflection of light that is backscattered to the sensor is more representative of the bottom of the waters.

Water bottom. The bottom type in coastal waters is classified into hard bottom, soft sediment, and coral reefs. They include sand, rock, seagrass, soft coral reefs, and hard coral reefs. Although this does not constitute as a disturbance in the reflectance, it is noted that variation in the water bottoms cause the energy received by the sensor to vary at the same depth.

2.4 Sentinel-2 image products

The European Space Agency (ESA) launched the Sentinel-2A mission on 23 June 2015. In the following two years, ESA launched its twin satellite, the Sentinel-2B, on 7 March 2017. Having two satellites in the same orbit makes the Sentinel-2 satellite program able to capture the same footprints every five days. Each satellite in the Sentinel-2 mission carries the MSI payload that samples 13 spectral bands with spatial resolution ranging from 10 m to 60 m (Table 2.2).

Table 2.2: Sentinel-2 MSI bands.

Spatial resolution (m)	Band number	Band name	Central wavelength (nm)	Bandwidth (nm)
10	2	Blue	490	65
	3	Green	560	35
	4	Red	665	30
	8	Near Infrared (NIR)	842	115
20	5	Vegetation Red Edge (VRE1)	705	15
	6	Vegetation Red Edge (VRE2)	740	15
	7	Vegetation Red Edge (VRE3)	783	20
	8a	Near Infrared narrow (NIRn)	865	20
	11	Short Wave Infrared (SWIR1)	1,610	90
	12	Short Wave Infrared (SWIR2)	2,190	180
60	1	Coastal aerosol	443	20
	9	Water vapour	940	20
	10	Cirrus	1,375	30

The Sentinel-2 data products are provided based on the processing levels as follows [European Space Agency, 2015]:

- *Level-0*, performed in real-time during the data reception process in order to package the MSI and satellite ancillary raw data acquired by the ground segment. The archiving process, together with the relevant parameters and metadata, is also executed at this level. A Level-0 data product is raw data at full resolution and is not accessible to users. The requirements in Level-0 data products are necessary to facilitate onward processing to Level-1A;

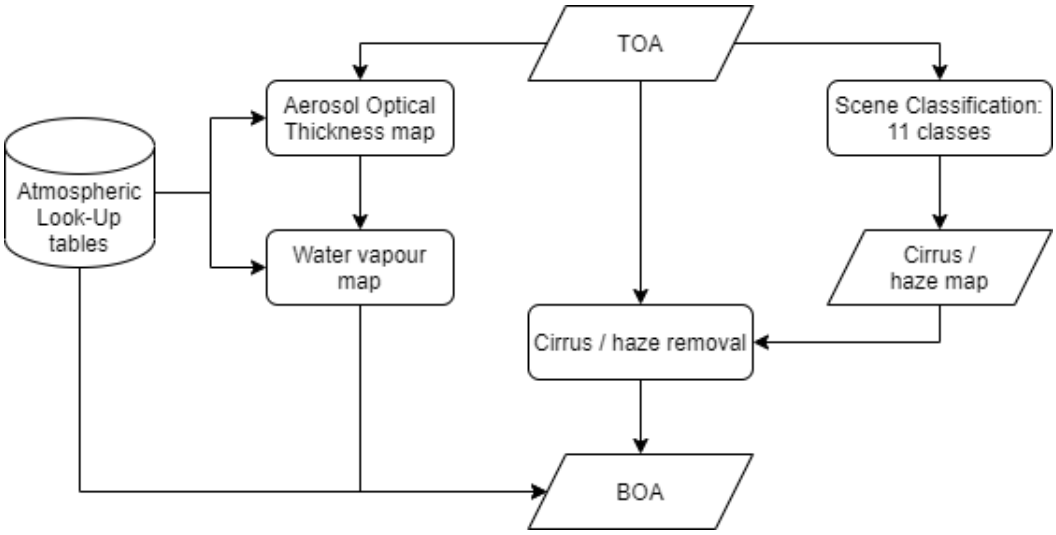


Figure 2.4: A summary of atmospheric correction procedures using Sen2Cor algorithm (modified from Richter et al. [2012] and Main-Knorn et al. [2017]).

- *Level-1A* processing is focused on decompressing the mission source packets. Similar to Level-0, the Level-1A data products are not accessible to users. Output from this level is used for further processing to Level-1B;
- The *Level-1B* processing sequence includes radiometric correction, resampling on the common geometric grid between the Global Reference Image (GRI) and the red band as a reference, tie points collection and filtering for the registration process, and refinement using the viewing model and Ground Control Points (GCPs). The output is a compressed image using the JPEG2000 algorithm that is used in the Level-1C processing phase. The Level-1B products are the lowest product level available to users;
- *Level-1C* data products are radiometric and geometric corrected images. The corrections include ortho-rectification and spatial registration on a global reference system, UTM/WGS84 projection, using a Digital Elevation Model (DEM) to project the image in cartographic coordinates. Top-Of-Atmosphere (TOA) reflectances calculation is also performed at this level. The images are available to users in JPEG2000 format with a GML geographic imagery-encoded header;
- The *Level-2A* processing step performs a scene classification and an atmospheric correction to the TOA Level-1C ortho-image products. The output is a Bottom-Of-Atmosphere (BOA) corrected reflectance image. Besides 13 spectral images, Level-2 also provides several additional outputs: an Aerosol Optical Thickness (AOT) map, a Water Vapour (WV) map, a Scene Classification Map (SCM), and Quality Indicators for cloud and snow probabilities.

2.4.1 Sentinel-2 Level-2A

The Level-2A products have been systematically generated at the ground segment globally since December 2018. Users can also generate BOA products using Level-1C products as input through the Sen2Cor processor as a plugin in the Sentinel-2 Toolbox. Here we describe a general procedure that has been applied to Sentinel-2 images in order to generate the Level-2A products.

The basic principle to retrieve the atmospheric constituents is through the radiative transfer model in the atmosphere. In this case, the relevant atmospheric terms as a set of Look Up (LU) tables are computed using libRadtran [Mayer and Kylling, 2005]. There are five atmospheric constituents in the libRadtran model: Rayleigh scattering, molecular absorption, aerosol, water and ice clouds. More Look-Up tables provide more constituents according to the geographic location and climatology, e.g. midlatitude, rural, urban, and maritime aerosol and ozone content. In order to generate the database, libRadtran shall

be performed using a wide variety of sensor and solar geometries, ground elevations, atmospheric conditions, and calculated with a high spectral resolution, i.e. 0.6nm [Richter et al., 2012].

Before applying the atmospheric correction, a classification of the scene is needed to distinguished 11 categories: medium and high cloud probabilities, cirrus, vegetation, soils or deserts, water, snow, dark area pixels, cloud shadows, saturated or defective pixels, and unclassified pixels [Main-Knorn et al., 2017]. These classes are then used to create a cirrus or haze map that is needed for cirrus correction.

The AOT map provides information regarding the visual transparency of the atmosphere. It is produced based on the dense dark vegetation (DDV) approach [Kaufman and Sendra, 1988], where the correlation between the SWIR2 band and two visible bands, Red and Blue, is used to estimate the visibility. The algorithm requires DDV, dark soils, and water bodies pixels as reference areas of a known reflectance pattern. If no such pixels are found, the algorithm will use a constant AOT, which is specified in the configuration file [Main-Knorn et al., 2017; Richter et al., 2012].

The WV map is retrieved based on the Atmospheric Pre-corrected Differential Absorption algorithm [Schläpfer et al., 1998]. It is applied to the NIRn band as the reference band in an atmospheric window region; and the water vapour band as the measurement band in the absorption region. The absorption depth is estimated by calculating the radiance for an atmosphere with no water vapour. It assumes that the surface reflectance for B9 (water vapour) is the same as B8a (NIRn). The depth of the absorption is a measure of the water vapour column content.

It is noted that the Sen2Cor algorithm is more focused on the land than the ocean surface. The atmospheric correction over the ocean region is performed using the AOT estimated over land pixels in the image. Other processors to perform atmosphere correction exist, e.g. ACOLITE [Vanhellemont and Ruddick, 2014], Regional Coastal Color (C2RCC) [Brockmann et al., 2016; Doerffer and Schiller, 2007], Fast Line-of-Sight Atmospheric Analysis of Hypercubes (FLAASH) [Perkins, 2012; Perkins et al., 2005], and Deep Water Calibration (DCW). ACOLITE uses a set of LU tables generated from 6SV [Vermote et al., 2006] to perform Rayleigh correction. C2RCC uses a large database of simulated water leaving reflectances and the TOA for NN trained to determine the atmospheric correction as well as inherent optical properties of the water body. FLAASH incorporates the MODTRAN radiation transfer code [Matthew et al., 2000]. DCW uses the LU tables based on Antoine and Morel [1999] and Shettle and Fenn [1975].

2.4.2 Sentinel-2 applications for SDB

With medium resolution images and high revisit frequency, the Sentinel-2 observation data contributes to the Copernicus programmes for land monitoring, emergency management, security, and climate change [European Space Agency, 2015]. Although the Sentinel-2 mission focuses more on the land than water, the products have been used several times to extract water depth in various coastal areas, e.g. Persian Gulf [Kabiri, 2017], Eastern Mediterranean [Traganos et al., 2018], South Florida [Caballero and Stumpf, 2019], and Ireland [Casal et al., 2020].

Kabiri [2017] performed linear and ratio transform techniques to retrieve depths in the northern Persian Gulf, namely Nayband Bay. Two Sentinel-2 Level-1C images with relatively low turbidity were selected. The atmospheric correction was performed using the FLAASH module. The study considered all possible combinations of three visible bands: red, green, and blue. In terms of RMSE, the evaluation results produced an accuracy of 2.12 m when all three visible channels were used in the linear transform, and 2.06 m when the ratio between blue and green was performed in the ratio transform method.

Traganos et al. [2018] used multi-temporal Sentinel-2 Level-1C images in 2016 to extract shallow water depths in three different locations around the Aegean Sea, namely Thermaikos Gulf, Apokoronas, and South Crete. The study areas cover mixed bottom types, including rocky reefs, sand, and seagrass meadows. The Thermaikos area has various human activity, such as aquaculture, industry, tourism, and fishing, and thus has an affected coastal system. The atmospheric and water column corrections using deep water pixels, sun-glint correction, and SDB computations were performed using the Google Earth Engine (GEE) platform. The linear transform using the blue and green bands yielded a RMSE of 2.46 m, 1.67 m, and 4.1 m in the South Crete, Apokoronas, and Thermaikos Gulf, respectively. Meanwhile,

the ratio between blue and green yielded less accurate results in the South Crete and Apokoronas, with RMSE roughly 3.49 m, 9.5 m, and better result in one other area, with RMSE approximately 2.88 m.

Caballero and Stumpf [2019] implemented the ratio transform method to generate SDB using Sentinel-2 Level-1C images around South Florida coastal waters, which are West Palm Beach, Key West, and Dry Tortugas. These areas generally have low to medium turbidity with various sea bed covers, e.g. sand, seagrass, coral patches, and reefs. The atmospheric correction was performed using the ACOLITE processor. The results obtained an accuracy, in terms of median absolute error, with a range of 0.45-0.78 m in West Palm Beach, 0.39-0.42 m in Key West, and 0.22 m in Dry Tortugas.

Casal et al. [2020] applied SDB using the model inversion approach on multiple Sentinel-2 Level-1C images to several bays in Ireland. The areas include estuaries and aquaculture, with sandy bottoms, combined with pebbles and rocks, or algae. The depths are relatively shallow, from 0 to 10 m. Satellite images were corrected using two different methods, which are DWC and C2RCC. The results indicated that the correction using the DWC approach obtained a more accurate SDB prediction than C2RCC. The RMSE values of this analytical method varied between 1.60 and 2.43 m, which were similar to the linear transform method from the previous studies in the same study areas.

3 Methodology

This chapter describes the study areas and methodology of this research. Figure 3.1 illustrates a general workflow that is taken in this thesis. A complete set of codes, including sample data, is provided through Github[1]. Before going into the details of each step, Section 3.1 describes the study areas, including an overview of their characteristics. Hereafter, Section 3.2 explains the selection procedure of Sentinel-2 images and in-situ depth data sets. A description of data preprocessing for multispectral images and depths are presented in Section 3.3, followed by an explanation of data preparation for CNN training and SDB model generation in Section 3.4. Furthermore, Section 3.5 provides several parameters and architecture settings used for analysis. Finally, Section 3.6 briefly describes the assessment metrics together with the overview of several experiments that are arranged consecutively to answer the research questions.

3.1 Study area, data, and tools

For this study, several areas were selected considering the availability of data sets which also have mostly clear water conditions. The water conditions were measured visually and by looking for spectral values that differed according to the depth. Figure 3.2 shows the distribution of areas to be studied, and Table 3.1 provides a list of study areas together with the multispectral and bathymetric data source that were used for that area. There are six AOI scattered around the Puerto Rico main island, southern Florida, and Oahu island. Puerto Rico main island is one of the locations that has various characteristics of coastal water. The south part of the island faces the Caribbean Sea, famous for its clear water, while the northern coast faces the North Atlantic Ocean. This situation makes the northern areas generally have bigger waves than the southern. This project used two areas in the south of the island and one area in the north. The following study area is in southern Florida, which has relatively clear water that is shallower than the Puerto Rico areas. Geographically, the area is surrounded by the Bahamas and the Gulf of Mexico. These first four areas were used as study areas to analyze CNN accuracies in different cases, i.e. coastal water conditions. Meanwhile, the last two areas acted as test locations of the previous results. They are located in Oahu island, Hawaii, in between the North and South Pacific Ocean. The areas surrounding Oahu generally have a high clarity level with the same depth range as Puerto Rico coastal waters. Additionally, in order to assess the scalability of this approach, this study also carried out a complete SDB model production of the entire coastal waters of Oahu island.

Table 3.1: An overview of study areas and data sets.

Area of Interest (AOI)	Area name	Location	Sentinel-2 date	Bathymetric data source
1	Ponce	Puerto Rico	3 January 2019	LiDAR 2018, MBES 2018
2	Lajas	Puerto Rico	3 January 2019	LiDAR 2018
3	San Juan	Puerto Rico	2 February 2019	LiDAR 2018, MBES 2018
4	Key West	Florida, USA	28 February 2019	LiDAR 2019
5	Makua	Oahu, Hawaii	10 December 2019	LiDAR 2013
6	Honolulu	Oahu, Hawaii	10 December 2019	LiDAR 2013

This project uses Sentinel-2 Level-2A collection images, which are atmospherically corrected. In-situ bathymetry data from LiDAR and MBES were used. They should meet the IHO Standards for Hydrographic Surveys (S-44) 5th Edition [International Hydrographic Organization, 2008] for Orders special

[1] https://github.com/yustisiardhitasari/sdbcnn.git

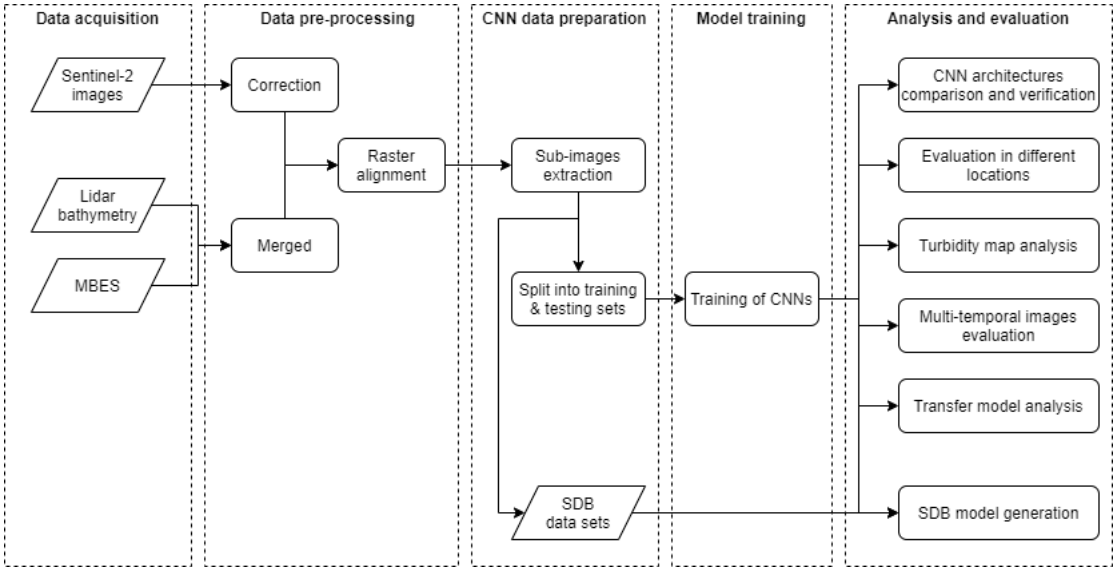


Figure 3.1: Overview of the methodology.

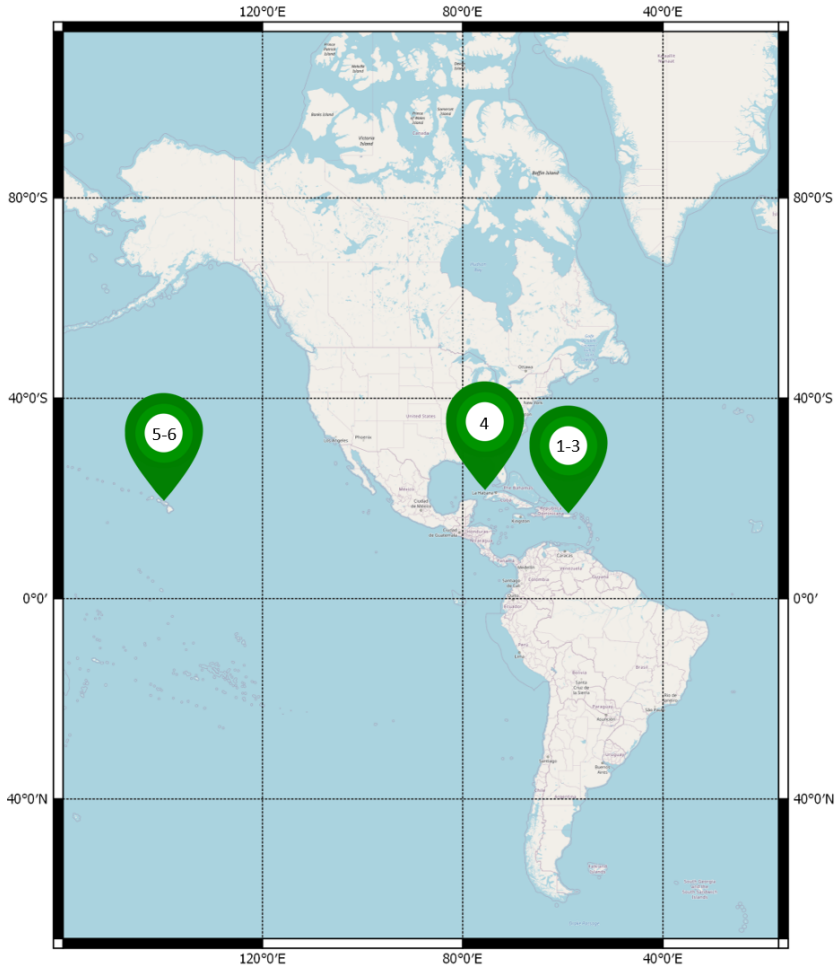


Figure 3.2: Distribution of study areas.

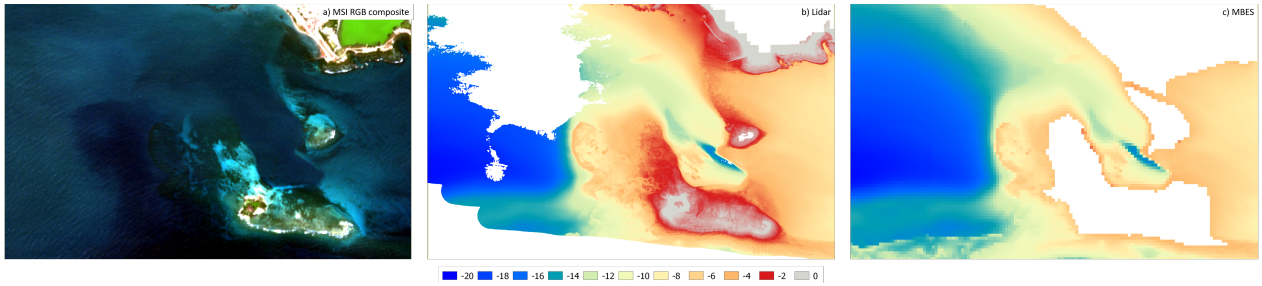


Figure 3.3: RGB composite image and bathymetric depths in AOI-1.

or 1, with the maximum allowable Total Vertical Uncertainty (TVU) at the 95% confidence level for a specific depth is given by:

$$TVU = \pm\sqrt{a^2 + (b * d)^2} \tag{3.1}$$

where a is the portion of the uncertainty that does not vary with depth, b is a constant, and d is the depth. The a and b are defined in the standard regarding the Order. For Order special, $a = 0.25$ m and $b = 0.0075$; while Order 1a, $a = 0.5$ m and $b = 0.013$. The Order special is defined as areas where under-keel clearance, the distance between the lowest point of the ship's hull or keel and the sea floor, is critical, e.g. harbours, berthing areas, and shipping channels. Meanwhile, Order 1a is for areas where under keel clearance is less critical but features that concern to shipping may exist. Order 1a is limited to water shallower than 100 m. Other than these Orders, IHO also specifies Order 1b and Order 2. Order 1b is similar to Order 1a, but the under keel clearance is not an issue here. Meanwhile, Order 2 is intended for areas deeper than 100 m.

In this research, several tools are used to perform each part of the methodology. As an overview, the GEE[2] is used early to select satellite products based on several criteria. The following preprocessing step requires some open source tools, such as QGIS[3], LAStools[4], and PDAL[5] for data handling–point cloud to raster conversion, projection, clip, merge, align, and visualization. Additionally, multiple R[6] scripts are created to perform image stacking and correction. Furthermore, the data preparation for CNN uses a combination of NumPy[7] and Rasterio[8] to generate sub-images in the NumPy array format. After that, training of CNN is conducted using Tensorflow[9] CPU/GPU. The use of a graphics processor is required when training with very large data. Lastly, the evaluation stage uses several modules in Python[10]: NumPy, pandas[11], scikit-learn[12], Matplotlib[13] to compute various metrics, and QGIS to visualize the results.

In order to introduce the character of the coastal waters, the following provides a brief overview of each study area and visualization of the multispectral and depth imagery.

AOI-1: Ponce, Puerto Rico. Ponce was the first study area that was used to build a CNN architecture for SDB. Figure 3.3a visualizes the natural colour composite of the Sentinel-2 Level-2A image on 3 January 2019. The bottom of this study area is a mix of coral rock, patch reefs, algae, and seagrass [Appeldoorn et al., 2016; NOAA, 2017a]. This area is considered to have clear waters in general with a turbidity level that increases at the nearshore. Bathymetric depths from airborne LiDAR (Figure 3.3b) and MBES

[2] https://earthengine.google.com/
[3] https://qgis.org/en/site/
[4] https://rapidlasso.com/lastools/
[5] https://pdal.io/
[6] https://www.r-project.org/
[7] https://numpy.org/
[8] https://rasterio.readthedocs.io/en/latest/
[9] https://www.tensorflow.org/
[10] https://www.python.org/
[11] https://pandas.pydata.org/
[12] https://scikit-learn.org/stable/
[13] https://matplotlib.org/

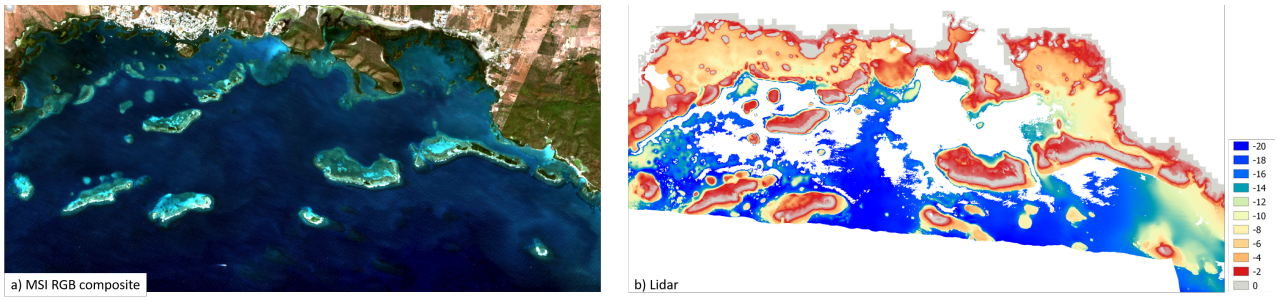


Figure 3.4: RGB composite image and bathymetric depths in AOI-2.

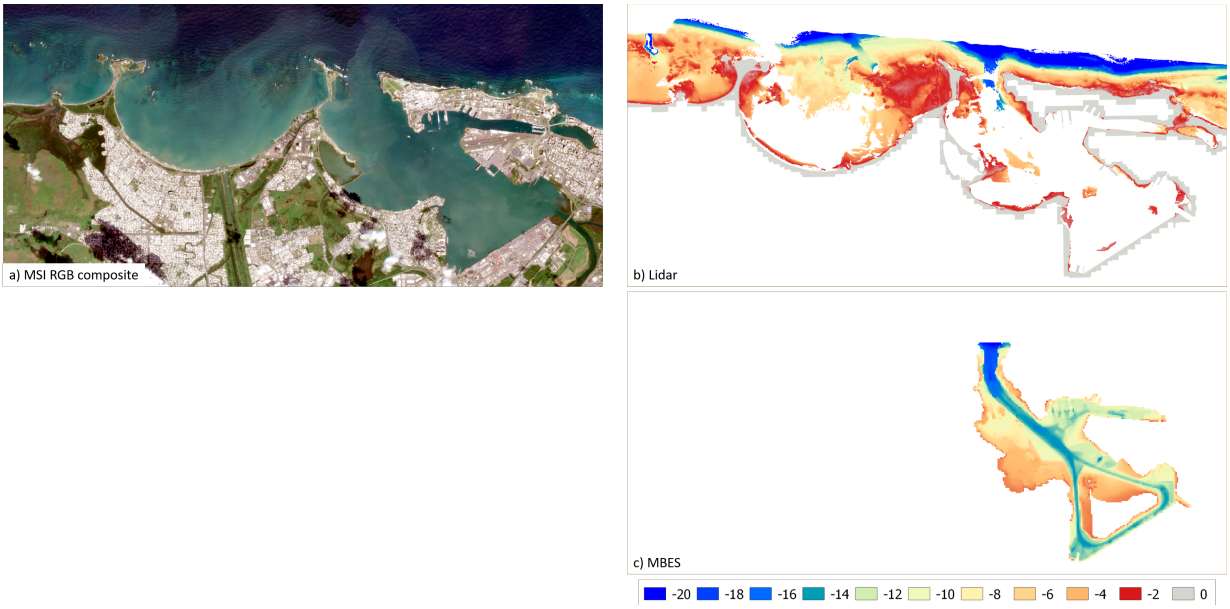


Figure 3.5: RGB composite image and bathymetric depths in AOI-3.

(Figure 3.3c) are available, both of which were acquired in 2018. The LiDAR data is compliant with the IHO Order special and 1a, while the MBES data meet the Hydrographic Surveys Specifications and Deliverables (HSSD) 2018 Edition [National Ocean Service (NOS), 2018] where the TVU calculation refers to IHO Order 1. The LiDAR depth ranges from around 11 m above sea level to around 17 m below sea level, while the MBES data ranges from 5 to 57 m below sea level. As seen in Figure 3.3c, MBES lacks depth in very shallow, 0-5 m; meanwhile, LiDAR is the opposite.

AOI-2: Lajas, Puerto Rico. Lajas is located adjoining and within the same image as Ponce. Figure 3.4a shows the visualization of the Sentinel-2 Level-2A RGB composite on 3 January 2019. The area is filled with diverse bottom types, such as rock or boulders near the shoreline, reefs with high coral diversity and density, and sand around the small islands [Bauer et al., 2012; NOAA, 2017a; Larsen and Webb, 2009]. The sandy bottom with no biological cover appears brighter (turquoise) than the sandy bottom with seagrass and algae. The waters are generally clear in this area, except near the shore; however, higher levels of turbid waters are found sporadically due to river plumes coming from the east part of this area [Bejarano and Appeldoorn, 2013]. Bathymetric depths are available from a LiDAR measurement campaign, which was acquired in 2018. The elevation ranges from 35 to -22 m (Figure 3.4b). The data is also compliant with the IHO Orders special and 1a.

AOI-3: San Juan, Puerto Rico. San Juan is the capital city of Puerto Rico, located in the north of the mainland and has abundant rainfall associated with the North Atlantic Ocean winds. This condition causes more clouds to appear in the northern than in the southern regions of Puerto Rico. The area of study is a port area where ships, boats, and dredging areas are captured by the satellite. Since the

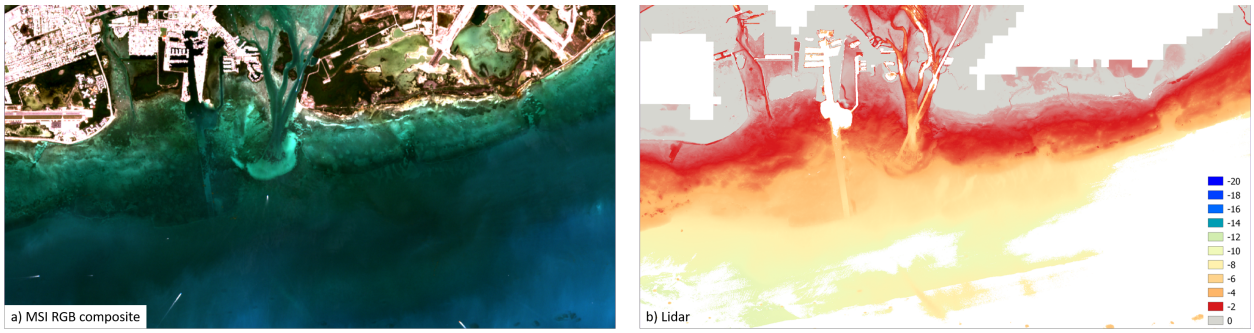


Figure 3.6: RGB composite image and bathymetric depths in AOI-4.

waters are disturbed with port activities, their condition is not as clear as other Puerto Rico areas. The high turbidity level is also seen in Figure 3.5a as green-brown colour pixels. This image was collected on 2 February 2019 and contains some clouds over land areas. Besides seagrass meadows, the area also contains various pollutants, e.g. oil, grease, ammonia, and low dissolved oxygen, as an impact from industrial operations, wastewater treatment system, urban runoff, and sewers [The Puerto Rico Environmental Quality Board (PREQB), 2016]. Outside the port area, reefs and coral rocks are found. The LiDAR and MBES data sets were acquired in 2018, which are compliant with the IHO Standards. As shown in Figure 3.5c, the MBES data is available around the port area with depths up to 20 m below sea level; meanwhile, the LiDAR data (Figure 3.5b) is available outside the port area with depths up to 30 m. The lack of the LiDAR data inside the port area is an indication that the area is not suitable for LiDAR measurement, partly because of its high turbidity level.

AOI-4: Key West, USA. Key West is located in southern Florida, almost at the end of the Florida Keys. The waters have low turbidity with lower nutrient levels in the deeper depth than at the nearshore. Higher nutrient levels and more turbid waters are found at the nearshore waters, partly due to beach wave resuspension strokes underwater sediments [Briceño and Boyer, 2015]. An image on 28 February 2019 (Figure 3.6a) shows that the area of study includes a channel between the northern Florida Keys and the southern, where the stream affects the sediment transport and causes a turbidity level higher than in other areas. Moreover, some locations appear green due to seagrass and algae in the water. Also, there is a dredging area beside the channel which has mixed seabed cover such as sand, mud, algae, and live coral. The seabed cover information is available through the Florida Keys National Marine Sanctuary Digital Atlas[14] The LiDAR data sets acquired in 2019 were used. The data complies with the IHO Orders special and 1a. As visualized in Figure 3.6b, the depths are relatively shallow, with the deepest depth around 10 m. This area has been studied by Caballero and Stumpf [2019] with larger extents, covering the northern and eastern regions. The previous study used the ratio transform method to extract shallow water depths.

AOI-5: Makua, Oahu. Figure 3.7a presents an RGB composite of the Sentinel-2 image on 10 December 2019 around Makua beach. The same area with similar extents has been studied by Sagawa et al. [2019] using another machine learning technique, which is Random Forest, and Landsat-8 images. The area is well-known for its long sandy beach with a coastal slope that increases in steepness. The nearshore area appears white, indicating a high reflective matter, mostly sand with coarse grains and calcareous as a result of the high energy wave in the winter [University of Hawai'i, 2013]. Other bottom types such as coral reef, rock or boulders are also exist. Bottom type classification is provided in the Hawaii Geoportal[15] LiDAR depths acquired in 2013 are available, covering the entire island with elevation ranges from 356 to -58 m. Depths are compliant with the IHO Orders special and 1a. Figure 3.7b presents LiDAR depths in the area of study, demonstrating a smooth transition from shallow to deep water.

AOI-6: Honolulu, Hawaii. As another port area, the Honolulu coastal waters were chosen as a comparison to the San Juan area. Figure 3.8a presents a Sentinel-2 image with the same date of acquisition as Makua, but a different footprint. Besides the port, it also shows an airport runway in the middle part

[14] https://www.arcgis.com/apps/webappviewer/index.html?id=03daf1d686c84ece8172ed394e287c78
[15] https://geoportal.hawaii.gov/datasets/

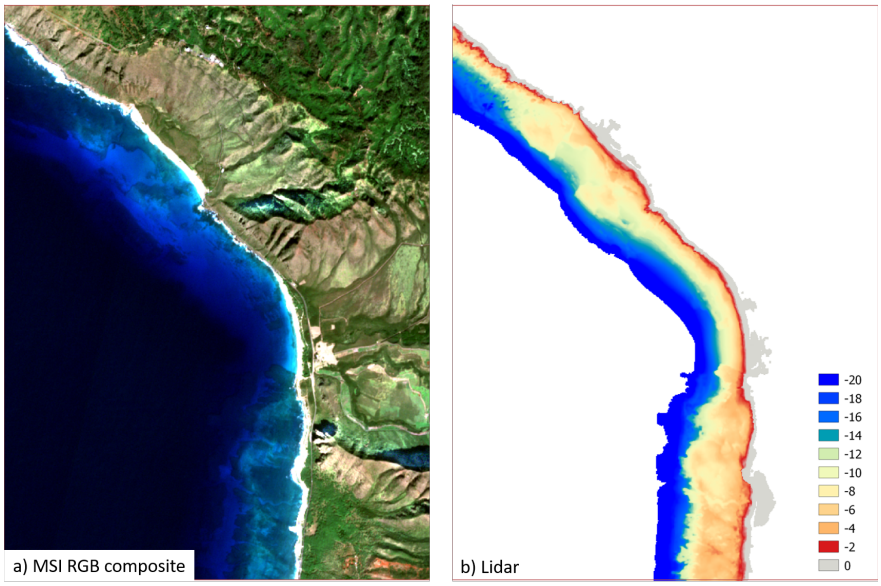


Figure 3.7: RGB composite image and bathymetric depths in AOI-5.

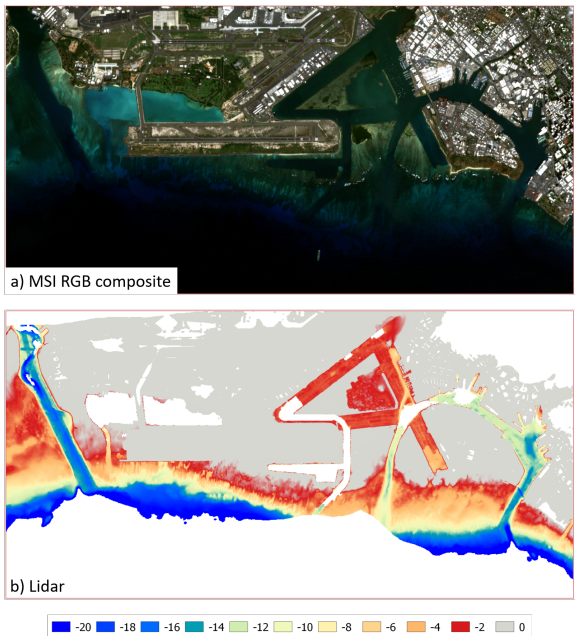


Figure 3.8: RGB composite image and bathymetric depths in AOI-6.

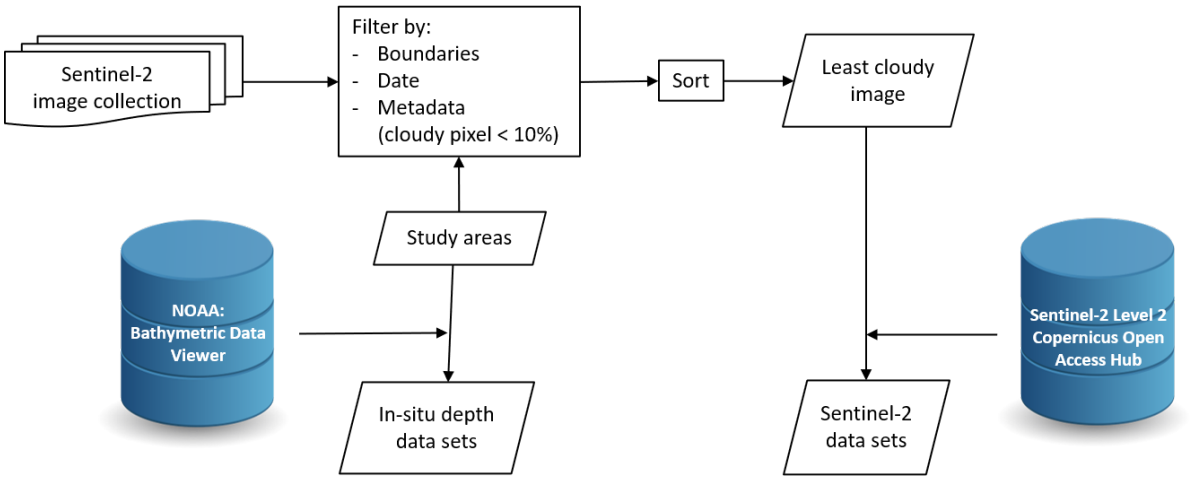


Figure 3.9: Data acuisition procedure.

of the image. Similar to other ports, the Honolulu port is also turbid, contaminated by several pollutants like trash, suspended solids, pathogens, ammonium nitrogen, and phosphorus [The Hawaii State Department of Health, 2018]. Bathymetric depths from LiDAR 2013 are available (Figure 3.8b), showing several channels. The eastern channel is the entrance to the Pearl Harbour area, while other channels head toward the Honolulu harbour areas. All entrance channels to the port are classified as dredged areas, having mud as its bottom type. Besides that, the water bottoms are mostly occupied by reefs, corals, pavements, and algae. In addition, this area has been studied by Lyzenga et al. [2006] using an IKONOS image and a simplified radiative transfer model, where the parameters were empirically derived based on a comparison between multispectral and measured depth values.

3.2 Data acquisition

The Sentinel-2 Level-2A data sets are open and accessible through the Copernicus Open Access Hub[16]. In order to select appropriate images, this study uses the GEE platform to filter Sentinel-2 Level-2A image collection based on the boundary extent of each study area and cloud pixel percentage as shown in Figure 3.9. Additionally, the acquisition date is used to obtain multi-temporal images, i.e. one image per month throughout the year of 2019. The image collection is then sorted based on its clouds percentage, and the least cloudy image is downloaded from the Copernicus database.

The availability of LiDAR and MBES data can be checked through the National Oceanic and Atmospheric Administration (NOAA) Bathymetric Data Viewer[17]. The MBES data sets are downloaded from the same source, while the LiDAR data is collected from the NOAA Data Access Viewer[18]. The LiDAR data is available as a raster or point clouds. This research uses the LiDAR image products where the depths are delivered with 1 m resolution. If the LiDAR raster is not available, then the point clouds are downloaded. The MBES data sets are delivered as raster with a spatial resolution of 32 m. The LiDAR and MBES images are available in AOI-1 and AOI-3, so both are used in these areas. For AOI-2 and AOI-3, only LiDAR raster products are available. Meanwhile, AOI-5 and AOI-6 use LiDAR point clouds instead of an image as the data source for in-situ depths. In addition, NOAA also provides tides predictions[19] that are used for in-situ depth correction. Tidal data of each location is collected following the Sentinel-2 images' sensing date and time.

[16] https://scihub.copernicus.eu/
[17] https://maps.ngdc.noaa.gov/viewers/bathymetry/
[18] https://coast.noaa.gov/dataviewer/#/LiDAR/search/
[19] https://tidesandcurrents.noaa.gov/

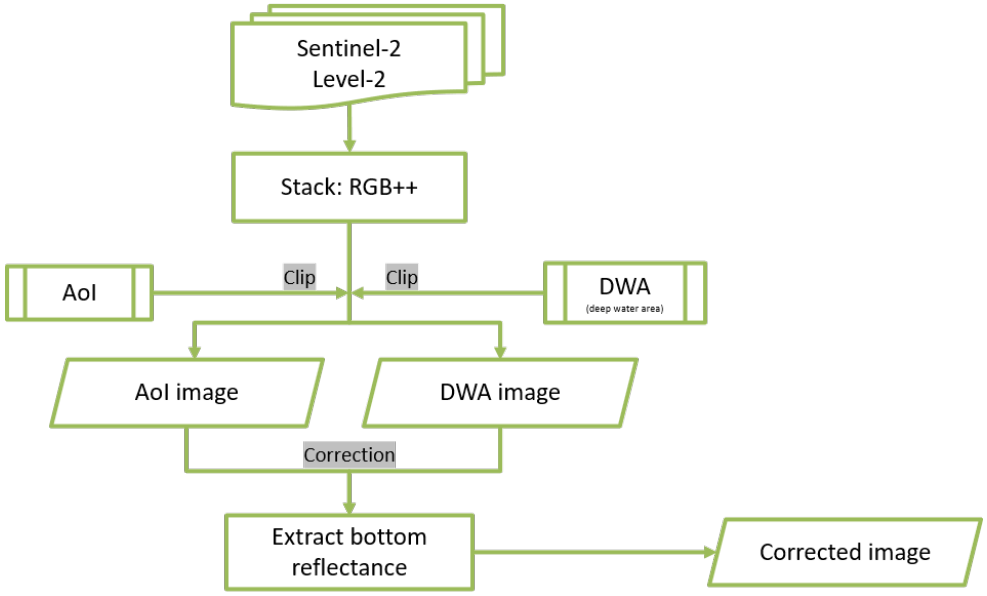


Figure 3.10: Bottom reflectance extraction using deep water pixels.

3.3 Data preprocessing

The Sentinel-2 Level-2A images provide surface reflectance values, meaning that the water column scattering effect needs to be corrected to obtain the bottom reflectance. Considering how light transmits through water, we can assume that the Deep Water Areas (DWA) reflectance values have no reflection effect from the water bottom. That being said, DWA reflectance pixels are beneficial to extract bottom reflectance in shallow water areas. Figure 3.10 illustrates the steps to do such a correction. After stacking multiple bands, the DWA is observed manually for each image based on its natural colour composite, where a darker pixel ideally indicates a deeper depth. The mean of DWA pixels are calculated and then becomes a deduction value to adjust the reflectance values using Equation 2.1. Besides using corrected images, this study examines the SDB results using uncorrected images. Uncorrected images imply that the original digital numbers of Sentinel-2 Level-2A are employed as data input instead of the corrected ones.

Figure 3.11 describes the workflow of in-situ depth data preprocessing. The LiDAR and MBES data have different spatial resolutions, which are 1 m and 32 m, respectively. Meanwhile, the spatial resolution of Sentinel-2 images is 10 m. Thus, both in-situ data sets were resampled to 10 m using the default parameter in QGIS, which is the nearest neighbour. Afterwards, both images were merged using the GDAL merge algorithm available in QGIS. It merges multiple rasters based on input arrangements where the last image overwrites the earlier ones in overlapping areas. In this case, the MBES data was set as the first image and the LiDAR as the last, so LiDAR depths were assigned as output when they overlap, since these are considered to be more accurate. In order to synchronize disparate times between in-situ and satellite images, a tidal correction was applied to the MBES and LiDAR bathymetry data depending on the satellite image sensing time and the vertical datum. The data sets have two kinds of vertical datum: Mean Lower Low Water (MLLW) and Mean Sea Level (MSL). MLLW is the mean of the lowest tide while MSL is the average hourly height; both are recorded during 19 years. For each area, the closest tide station was selected.

A different preprocessing step was applied to AOI-5 and AOI-6 in the Oahu since the LiDAR point clouds are the available data source instead of a LiDAR-derived raster. The original point clouds cover a wide range of elevation. In order to reduce the computational load when converting the data, the point clouds were filtered using LAStools to a specific range, from 4 m to -25 m, since this research focuses up to 20 m depth. The point clouds were converted into a raster using PDAL with 10 m spatial resolution as output. PDAL provides several statistical methods to assign an output pixel. In this study, the mean

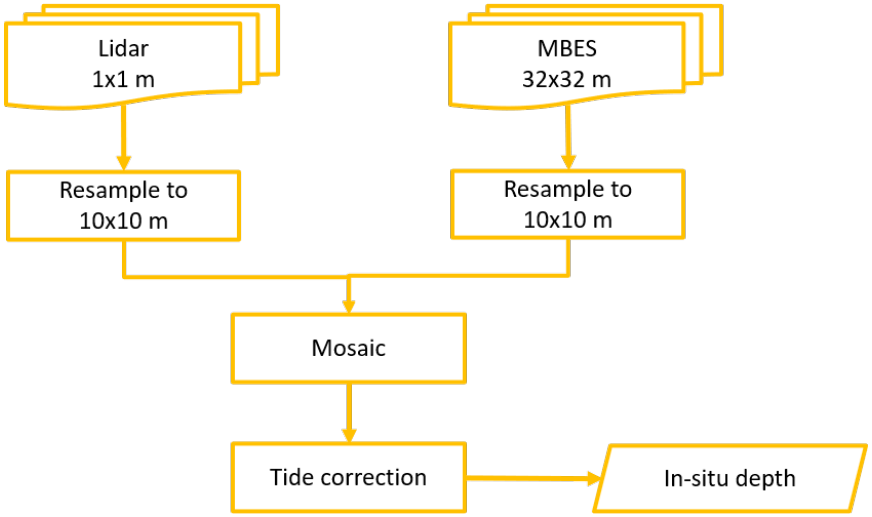


Figure 3.11: LiDAR bathymetry and MBES data fusion and tide correction.

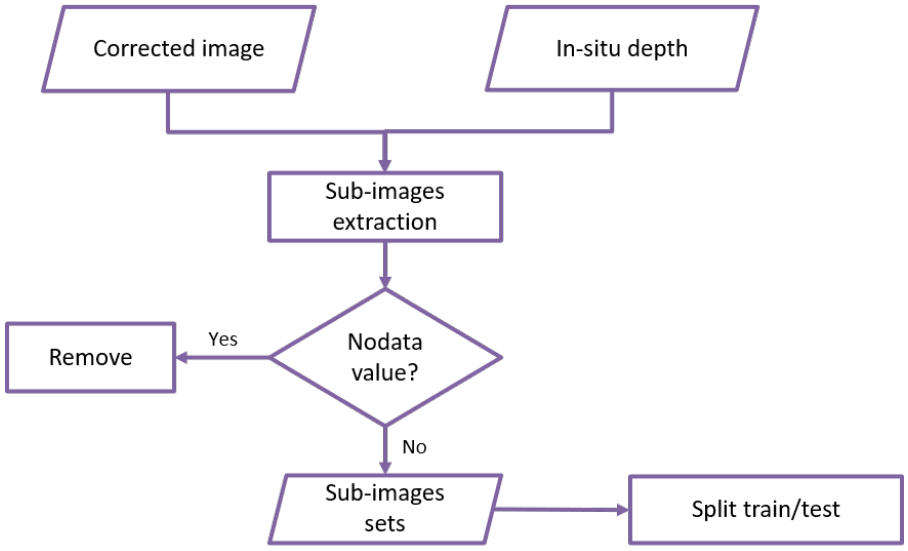


Figure 3.12: Training and testing data generation for CNN training.

value of all points within the default radius, which is $resolution * \sqrt{2}$, is used. The LiDAR point clouds still refer to the ellipsoid; consequently, the output image should be corrected to adjust the elevation refer to the MSL datum. The product metadata provides this correction value, which is -0.601 m.

3.4 CNN data preparation

Subsequently, Figure 3.12 presents the procedure to generate data sets for CNN training. This study applied raster alignment to depths and multispectral images to avoid shifting between pixels. Then, sub-images with a particular window size were extracted from the data. Figure 3.13 provides an example of the sub-image extraction for a window size of five and stride of three. As we can see, the amount of training data was reduced by striding the image pixels. Three different window sizes, i.e. 5×5, 7×7, and 9×9, were implemented during the CNN architecture building process. The window size is limited to 9×9 considering the computational load in the training process since more training data with more neighbours require more memory for the training computation. This project mainly uses the CPU pro-

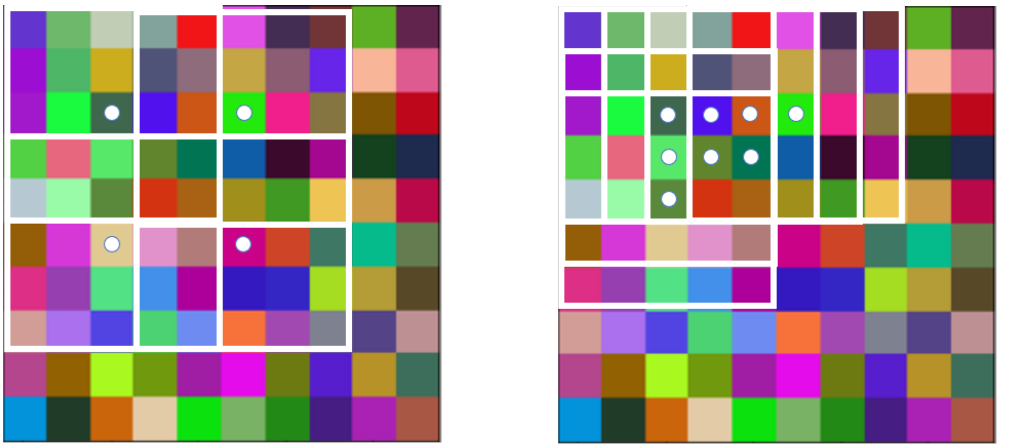


Figure 3.13: An illustration of sub-images extraction of multispectral images and in-situ depth. The boxes are the window size and the dots are the pixel depths.

cessor to run the model and only utilize the GPU processor when the number of training data is very large. Furthermore, the LiDAR bathymetry still captures the elevation on the land, so this project filtered the elevation to include only areas up to 2 m above sea level in the sub-images. At the end of this stage, the sub-images were separated randomly based on the total number, using the *numpy.random.choice* function under the NumPy module. The proportion of the training, including validation, and testing sets were 80% and 20%, respectively.

3.5 CNN training

The main difference in CNN for classification and regression is in the output activation layer function. This study uses a linear activation function in the last layer instead of softmax or other activations used in classification. This study uses ReLU as the activation function for other layers to address the non-linear trend between reflectance and water depth. Batch normalization in the architecture to reduce overfitting is also implemented. As for the hyper-parameters, this project tries several configurations, including learning rate, batch size, and dropout rate.

In order to optimize the model while training, the default loss function, which is the Mean Squared Error (MSE), is used. The function computes the mean of squares of errors between mesured and predicted depths. The squaring results in more error when there is a large deviation between the predicted and ground truth value. The optimizer will try to minimize that value to improve the model in each iteration. After a particular epoch, the MSE will stop improving. The training can be stopped automatically with an early stopping callback, using the MSE value to monitor the loss and a certain threshold used to stop the training. Meanwhile, this project uses 300 epochs to monitor the training. Besides the loss function, the MSE and Mean Absolute Error (MAE) values are used as the metric to evaluate the training results by monitoring the validation accuracy on each epoch. Based on the preliminary experiments, the validation accuracy in the training process was optimal with a learning rate of 0.0001, batch size of 512, and a dropout rate of 0.3. Thus, these parameters were used as the default setting in this research.

Furthermore, to build a complete architecture, a different number of convolutional layers, kernel sizes, and pooling layers is examined. Table 3.2 provides a list of the configurations that are implemented. The preliminary experiments suggested that the training cannot converge with only a single convolutional layer, yielding low validation accuracy. Thus, experiments with a single convolutional layer are excluded from this thesis. This study limits the convolutional layers and kernel size to three since our

maximum window size is 9×9. The use of the pooling layer intends to reduce the processing time without adding parameters during the training process. This project also examines whether it can improve the quality of the result.

Table 3.2: Different CNN architectures used in this study.

Architecture	Conv2D	Kernel size	Polling layer
CNN1	2×	2×2	No
CNN2	2×	2×2	Yes
CNN3	3×	3×3	No
CNN4	3×	3×3	Yes

Up to this stage, AOI-1 with only three channels: red, green, and blue is used. Furthermore, this research tries different band combinations, as shown in Table 3.3, to see whether they can improve our model or not.

Table 3.3: Band combinations.

Channels	Number of channels	Bands
RGB	3	Red, Green, Blue
RGBN	4	Red, Green, Blue, NIR
RGBNSS	6	Red, Green, Blue, NIR, SWIR1, SWIR2
All bands	9	Red, Green, Blue, NIR, SWIR1, SWIR2, VRE1, VRE2, VRE3

3.6 Result assessment

The assessment of the SDB model from CNN is carried out by comparing the predicted and in-situ depth from the test data. We use a standard accuracy assessment by estimating the RMSE (Equation 3.2) to measure the error and the coefficient of determination R^2 (Equation 3.3) to measure the variance of the model and the ground truth values. An additional metric, the Median Absolute Error (MedAE) (Equation 3.4), is used to compare the existing related work. In those equations, d_{ref_i} is measured depth at pixel i, d_{pred_i} is predicted value at the same pixel as d_{ref_i}, and $\overline{d_{ref}}$ is the mean of in-situ data.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(d_{ref_i} - d_{pred_i})^2}{n}} \tag{3.2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(d_{ref_i} - d_{pred_i})^2}{\sum_{i=1}^{n}(d_{ref_i} - \overline{d_{ref}})^2} \tag{3.3}$$

$$MedAE = median|d_{ref_i} - d_{pred_i}| \tag{3.4}$$

4 Experiments and Results

In this chapter, the overview of experiments and their results are presented and discussed. Several experiments are implemented to provide a comprehesive picture of CNN's performance for shallow water depth extraction. As shown in Figure 3.1, five consecutive assessments are elaborated in Section 4.1, with additional experiments in between.The verification of the SDB model using the CNN approach, together with the comparison between different CNN architectures, is provided in Section 4.2. Subsequently, Section 4.3 describes the results in different locations, including other assessment results, which are the CNN sensitivity analysis and the turbidity map. Next, Section 4.4 provides the SDB results using reflectance values directly without correcting the images. Hereafter, Section 4.5 demonstrates the results of SDB using multi-temporal Sentinel-2 Level-2A images. Furthermore, Section 4.6 discusses the SDB results using various pretrained models. Lastly, Section 4.7 attempts to provide a complete SDB model in the entire Oahu island.

4.1 Overview of experiments

SDB comparison and verification. This assessment is done at the first stage of analysis. AOI-1 sub-images with three visible bands as input for training are used. Using the same data sets as CNN, the SDB model is generated through the linear transform method. The results are compared using the accuracy metrics mentioned above. Using the same AOI, several architectures as indicated in Table 3.2 and Table 3.3 are implemented to the sub-image data sets. The overall accuracy between architectures is then compared and discussed. Additionally, RMSE values per depth range are also calculated to evaluate the results more specifically. Based on the analysis, a baseline of CNN architecture is established for further study.

SDB comparison in different locations. The CNN architecture that has been defined is applied to three other study areas, which are the Lajas (AOI-2), San Juan (AOI-3), and Key West (AOI-4). These areas represent three different cases. AOI-2 portrays a neighbour area of AOI-1, which is located in the same image as Ponce. AOI-3 serves as a location in the same coastal waters as AOI-1 and AOI-2, which is on Puerto Rico's main island but has a different water condition and a different image. AOI-4 represents coastal waters that are far away from the previous areas and have different water conditions. In addition, to evaluate SDB results using individual training on each area, a pre-trained CNN model from AOI-1 is reused in the other study areas. This experiment is expected to denote how the pre-trained CNN model works with new data sets with different cases. Next, Gaussian noise is introduced to the test data sets to evaluate the CNN model's sensitivity to noise.

Turbidity map analysis. Concerning turbidity issues in retrieving shallow water depths from multispectral images, this thesis attempts to map the turbidity level using the same image as input for CNN. The training sets are collected in AOI-3 since this area clearly shows variations in turbidity. Then, the trained model is used to generate a turbidity map in AOI-1, AOI-2, and AOI-3. Detail explanation about this experiment, including the CNN architectures that are used, will be provided in Section 4.3.1.

SDB comparison between different image preprocessing schemes. The previous results with corrected images using deep water pixels are compared to SDB results using uncorrected images. It means that the original digital number from Sentinel-2 Level-2A, which represents surface reflectance values, is directly used as data input for CNN training. This experiment intends to analyze whether CNN can rectify the water column disturbance in the training process.

SDB using multi-temporal images. In general, the quality of the machine learning result depends on the quality of the input data sets that are used for training. The quality includes the variation of the data to obtain a generalized model. When the training data represents the entire data, it may produce

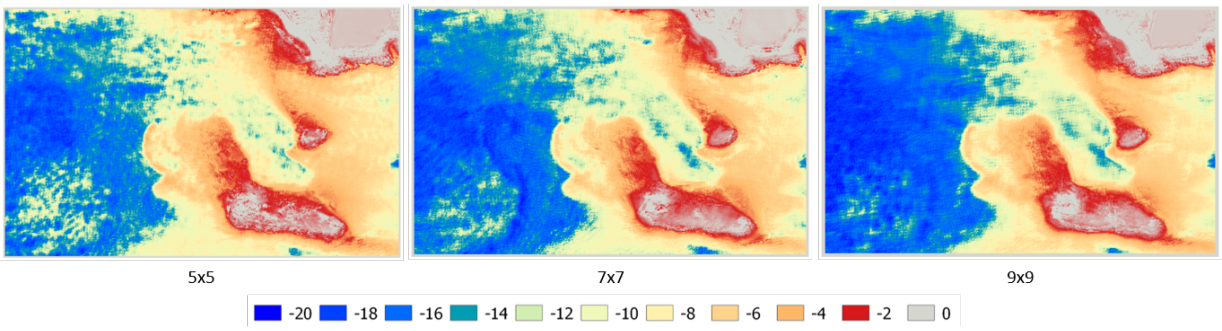


Figure 4.1: SDB results generated using CNN3 architecture with different window sizes in AOI-1.

an accurate and reliable training model. In this case, the project tries to collect the variation of data sets using a series of images in each location. With suitable cloud conditions, this study collects one image per month in 2019 at each location. An additional image is also collected as supplementary test data. All sub-images from each image are combined to train, except the test image. The training model is tested to the multi-temporal images, including the extra test image. The results are compared to the previous results on the same acquisition date.

Analysis of transfer model. Two new data sets are introduced in this experiment, which is Makua (AOI-5) and Honolulu (AOI-6). They are located in Oahu island and set as test locations to implement the pre-trained model. Beforehand, individual training on each area is carried out. Then, each pre-trained model from each AOI is reused in these new locations. Additionally, several combinations of training data in between AOI-1 to AOI-4 are trained and tested in AOI-5 and AOI-6.

SDB model production in a larger area. An implementation of the CNN model to the entire Oahu island as the last experiment is performed. SDB models are generated from three neighbouring images that are acquired in different footprints on the same date. The computations are performed individually per image, and then they are combined at the latest stage to produce a complete SDB in Oahu island.

4.2 CNN architectures comparison and verification

Using 10,820 sub-images of AOI-1 in the training process, Table 4.1 depicts the accuracy of different CNN architectures for 2,705 testing sub-images. The RMSE ranges from 1.48 m to 1.94 m. The pooling layer in CNN2 and CNN4 speeds up the training execution time by approximately 29%-32%, but does not improve the accuracy. A deeper network shows small improvements, but the validation accuracy is smoother for three than for two network layers.

Table 4.1: Accuracy assessment on different CNN architectures in AOI-1 using RGB bands. The RMSE values are in meters.

Architecture	5×5		7×7		9×9	
(Table 3.2)	RMSE	R^2	RMSE	R^2	RMSE	R^2
CNN1	1.59	0.91	1.58	0.92	1.55	0.93
CNN2	1.94	0.89	1.63	0.91	1.55	0.93
CNN3			1.53	0.92	1.48	0.94
CNN4					1.64	0.90

Based on Table 4.1, a larger window size lower the RMSE by approximately 10 cm in average. The comparison between the predicted depth (Figure 4.1) with the in-situ depth (Figure 3.3b and Figure 3.3c) as shown in Figure 4.2 shows that larger window sizes improve the result, especially in deeper depth

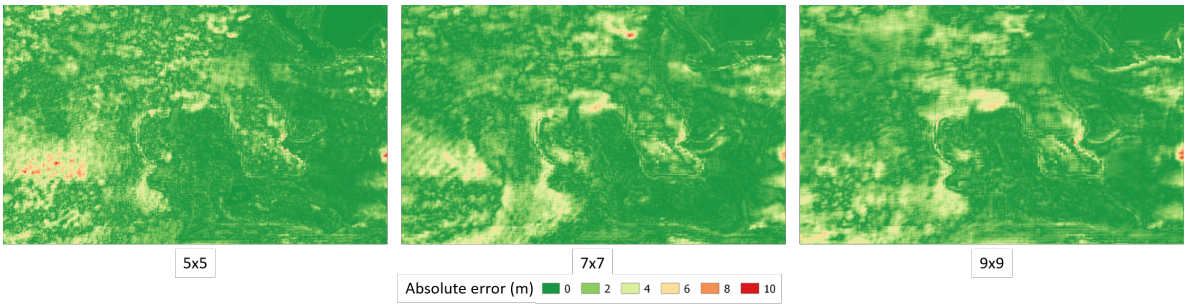


Figure 4.2: SDB error maps in AOI-1 based on computation using CNN3 with different window sizes.

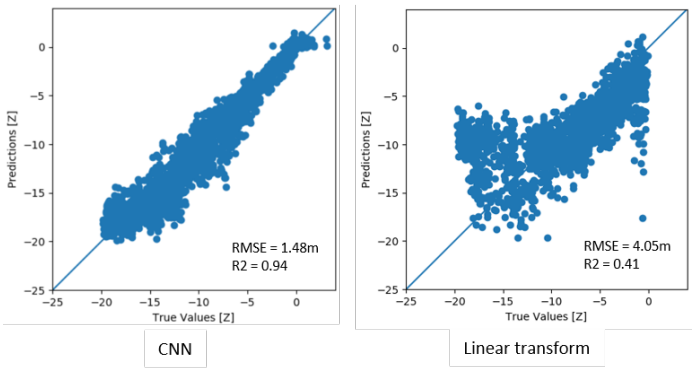


Figure 4.3: Scatter plot of predicted depth vs ground truth in AOI-1. The SDB is computed using CNN3 with a window size of 9×9 and RGB bands.

areas. Figure 4.3 shows the predicted depths with respect to the reference depths at 0-20 m depth range within the same test data sets. It illustrates that the CNN model tends to fit the data. With more training data, the linear transform method yields an overall accuracy of 4.05 m with R^2 of 0.41. For comparison, a linear transform using the same amount of in-situ data as CNN yields a lower accuracy, roughly 5.30 m with R^2 of 0.13. As illustrated in Figure 4.3, the linear transform method cannot fit the values as well, mainly in the very shallow areas, between 0 to 2 m, or in depths deeper than 10 m. Likely, the linear transform method cannot capture the non-linearity due to bottom reflectance variation in this area, since it predicts very shallow depths to be deeper and deep areas to be shallower.

However, some significant errors are still present in Figure 4.1. For example, there is a significant variation in depth in the east part of the image that is not visible in the reference data. The difference between prediction and reference in this area is up to 10 m and 5.67 m on average. However, there are no particular patterns of reflectance shown in the RGB composite image, and the reflectance and depth data pairs in the area and the surrounding area are similar. The errors might reflect the difficulty of CNN in predicting a particular depth range. Table 4.2 points out that the accuracy tends to increase as the depth increases since the bottom reflectance component becomes less accurate with increasing depth.

Table 4.2: SDB accuracy over different depth ranges in AOI-1 using CNN3 with 9×9 window size and RGB bands.

Depth ranges (m)	RMSE (m)
0-5	0.61
5-10	1.53
10-15	2.07
15-20	1.22

Furthermore, Table 4.3 and Figure 4.4 depict the impact of different band combinations on the accuracy

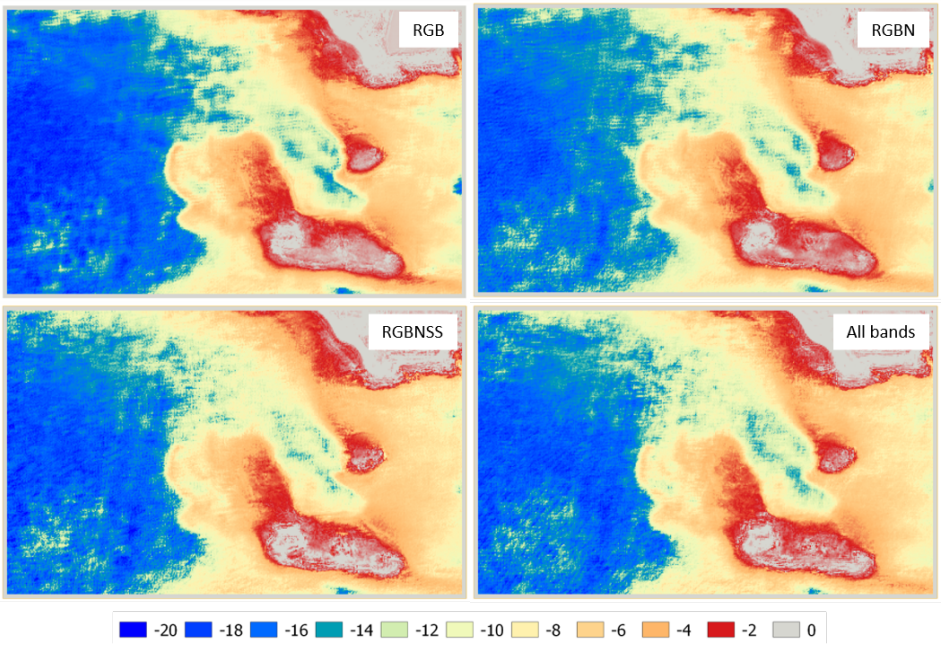


Figure 4.4: SDB model generated in AOI-1 using CNN3 with 9 × 9 window size and different number of channels.

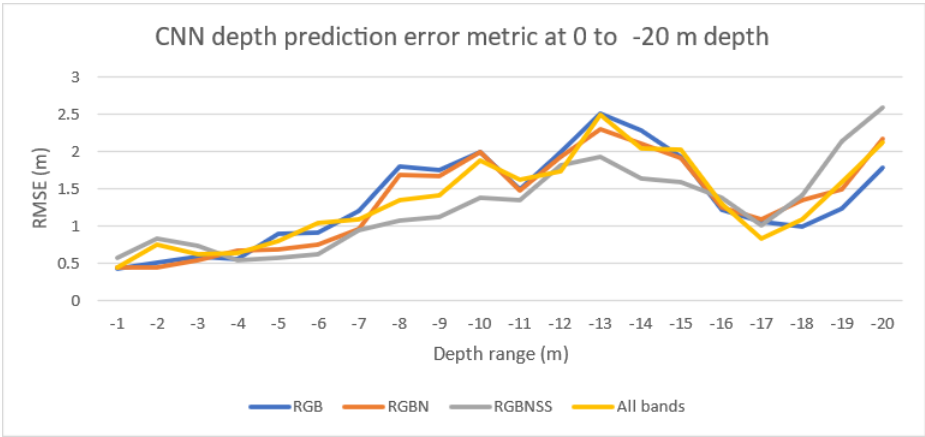


Figure 4.5: CNN3 accuracy metric per depth range 1 m in AOI-1 using 9 × 9 window size and different number of channels .

Table 4.3: Accuracy assessment of SDB using CNN3 with different number of channels and different window sizes in AOI-1.

Channels	5 × 5		7 × 7		9 × 9	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
RGB	1.59	0.91	1.53	0.92	1.48	0.94
RGBN	1.64	0.91	1.40	0.93	1.37	0.94
RGBNSS	1.63	0.91	1.53	0.91	1.31	0.94
All bands	1.68	0.90	1.55	0.91	1.45	0.94

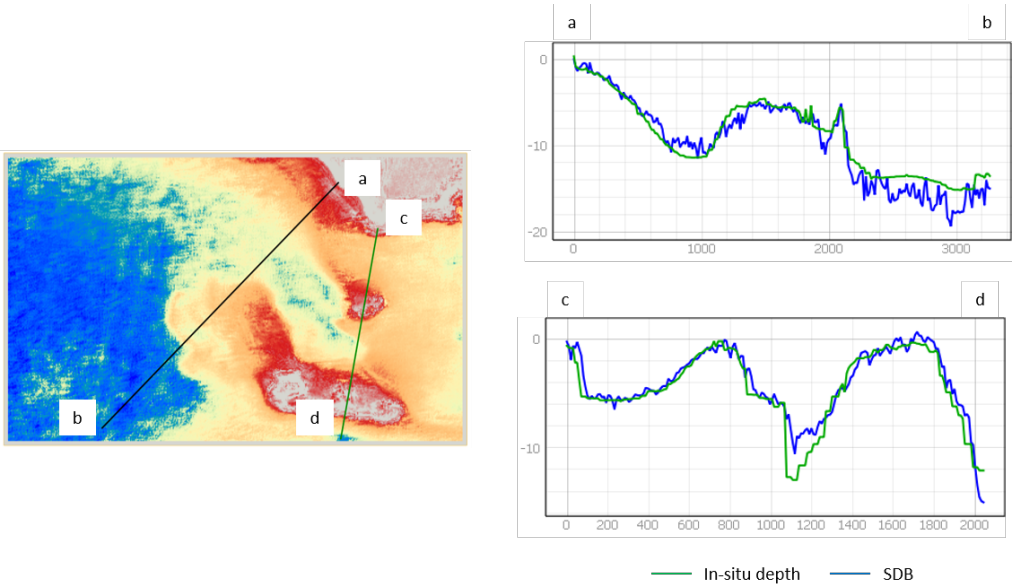


Figure 4.6: Cross profiles at AOI-1, showing comparison between SDB model and in-situ depths. The SDB was generated using CNN3 with RGBNSS bands and 9 × 9 window size.

of depth calculations. The accuracy improves again with larger window sizes, especially for a window size of 9 × 9 with six bands. As shown in Figure 4.4, using a different number of channels affects the SDB model. If we focus on the large error present in the eastern part as mentioned in the previous paragraph, the model can obtain a better prediction as we increase the number of bands used to train. Additionally, Figure 4.5 illustrates the accuracy for depths ranging from 0 m to 20 m. The results demonstrate that the combination of RGBNSS outperforms other configurations, except in the depth range between 1 m and 2 m, and in depths deeper than 17 m. It can be due to a lack of sufficient training data or because too many bands may cause a bias in a certain depth range of the CNN model. In general, the combination of RGB, RGBN, and RGBNSS have similar accuracies up to 7 m. Deeper than that, the SWIR bands contribute to keeping the accuracy lower than others. Meanwhile, the use of all bands does not indicate any particular improvement compared to other combinations.

Two cross profiles were generated in order to assess SDB models over in-situ measurements. As shown in Figure 4.6, the profile section a-b indicates that the SDB produces a good fit to the measurements in depths shallower than 14 m. In the deeper ranges, SDB tends to be more distracted. Meanwhile, the profile section b-c describes the trend over the area shallower than 14 m with land present in between. The cross section illustrates a good fit in the SDB model to detect the coastline area quite precisely.

Based on the experiments on several CNN architectures so far, the CNN3 architecture with a 9 × 9 window size and RGBNSS bands combination as shown in Figure 4.7 can provide the SDB with better accuracy compared to other architectures. Thus, this setting is used as the baseline architecture in the following

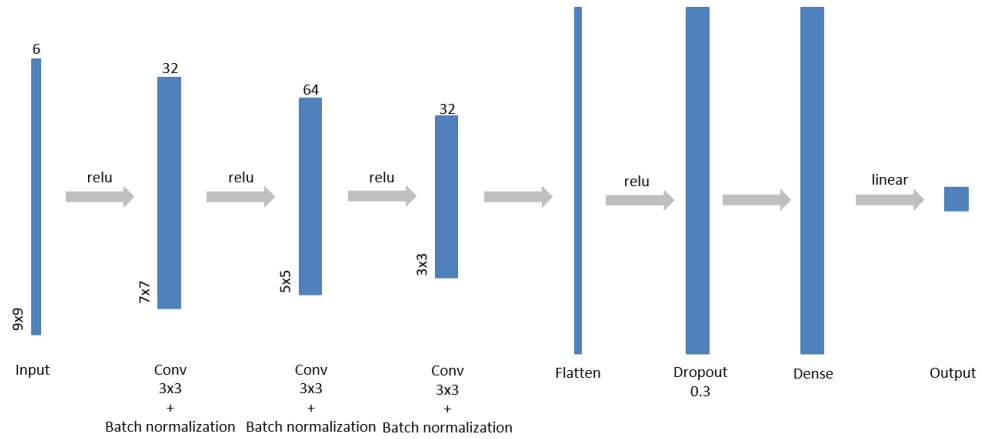


Figure 4.7: The baseline CNN architectures for SDB used in the following experiments.

studies. Hereinafter we call it CNN referring to CNN3.

4.3 SDB comparison in different locations

In order to verify the choosen band combinations, the same CNN architecture was applied to other areas in the same coastal waters, i.e. AOI-2 and AOI-3, and in a different area, i.e. AOI-4. Table 4.4 presents the accuracy of two different channel configurations, which are RGB and RGBNSS. Note that AOI-1 to AOI-3 has the same depth range (0-20 m), while AOI-4 has a shallower range of 0-10 m. Using the RGB combination, the RMSE values within the same depth range (0-10 m) in AOI-1, AOI-2, and AOI-3 are 0.87 m, 1.56 m, and 2.24 m, respectively. Meanwhile, using the RGBNSS combination, the accuracies in AOI-1, AOI-2, and AOI-3 are 0.91 m, 1.61 m, and 2.33 m, respectively. Additionally, the pretrained model from AOI-1 was used to calculate the SDB in other areas. Figure 4.8 to Figure 4.15 show the RMSE graphs for each increment of 1 m depth, a scatter plot showing reflectances with respect to depth, the corrected images, measured depths, the SDB results using the trained and pretrained model, and several cross sections for each AOI.

Table 4.4: Accuracy assessment of SDB using CNN trained in AOI-2, AOI-3, and AOI-4.

Channels / AOI	RGB		RGBNSS	
	RMSE	R^2	RMSE	R^2
AOI-2	2.77	0.83	1.78	0.92
AOI-3	3.15	0.32	3.12	0.46
AOI-4	0.71	0.95	0.88	0.93

In general, the RMSE of the result using the RGBNSS bands is overall lower by 1 m than the RGB combination in AOI-2. However, the result does not show any improvement in AOI-3 and is worse in AOI-4. A more detailed assessment is provided in Figure 4.8 where the RMSE is calculated in increments of 1 m depth. Figure 4.8a illustrates that RGBNSS improves the accuracy in areas deeper than 10 m, while RGB produces a better result at the 2-10 m depth range. Meanwhile, Figure 4.8b shows that the results between both band combinations are alike in AOI-3, except at the 16-19 m depth range where RGBNSS gives a higher accuracy than RGB. Furthermore, Figure 4.8c portrays the result in AOI-4, where the RGB outperform the RGBNSS combination in general, similar to AOI-1 at 0-10 m depth range. Accordingly, the error trend per 1 m depth range between AOI-1, AOI-2, and AOI-3 is not the same. Many factors presumably contribute to these errors, including the light penetration rates with respect to depth, the turbidity levels, and the bottom reflectance extraction method.

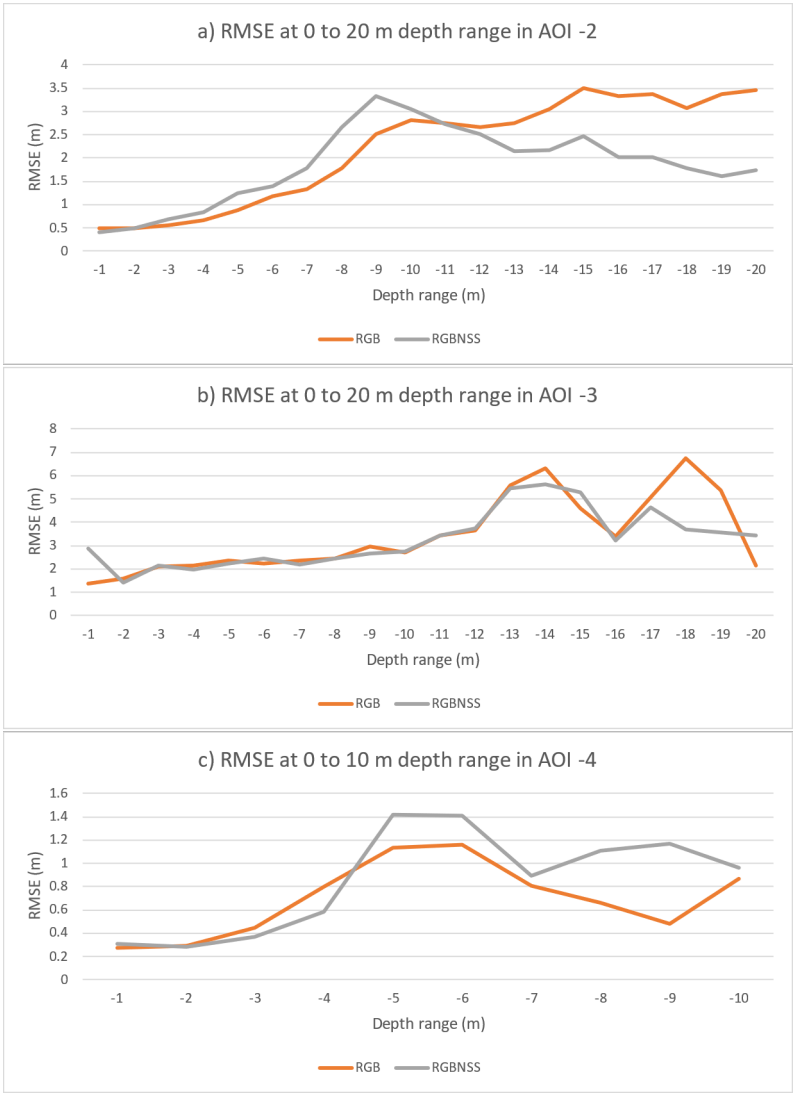


Figure 4.8: SDB accuracy per 1 m depth range using CNN trained with different number of channels in AOI-2 (a), AOI-3 (b), and AOI-4 (c).

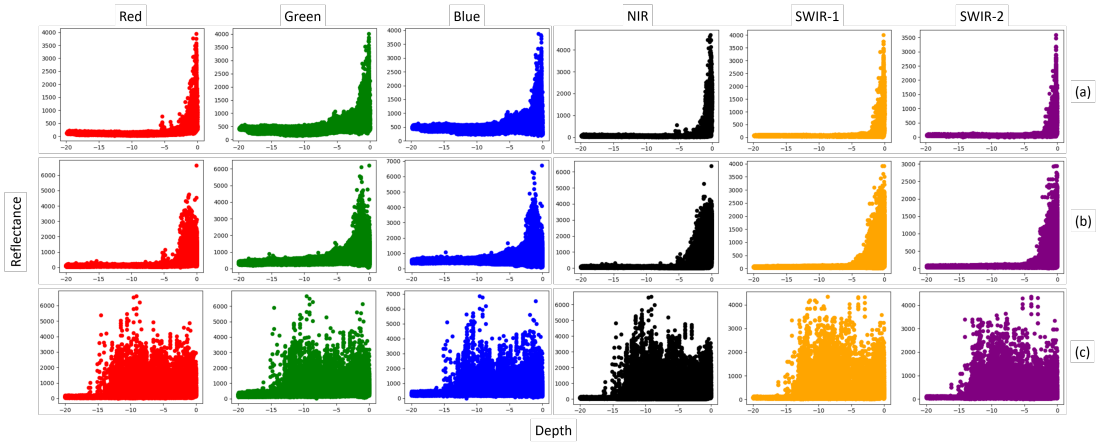


Figure 4.9: Multispectral reflectances scatter plot with respect to depth in AOI-1 (a), AOI-2 (b), and AOI-3 (c), at the 0-20 m depth range.

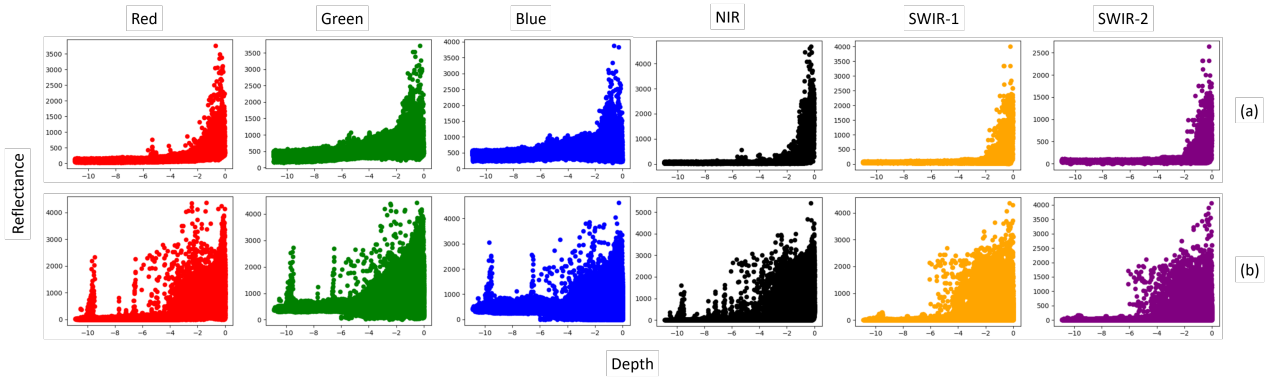


Figure 4.10: Multispectral reflectances scatter plot with respect to depth in AOI-1 (a) and AOI-4 (b), at the 0-10 m depth range.

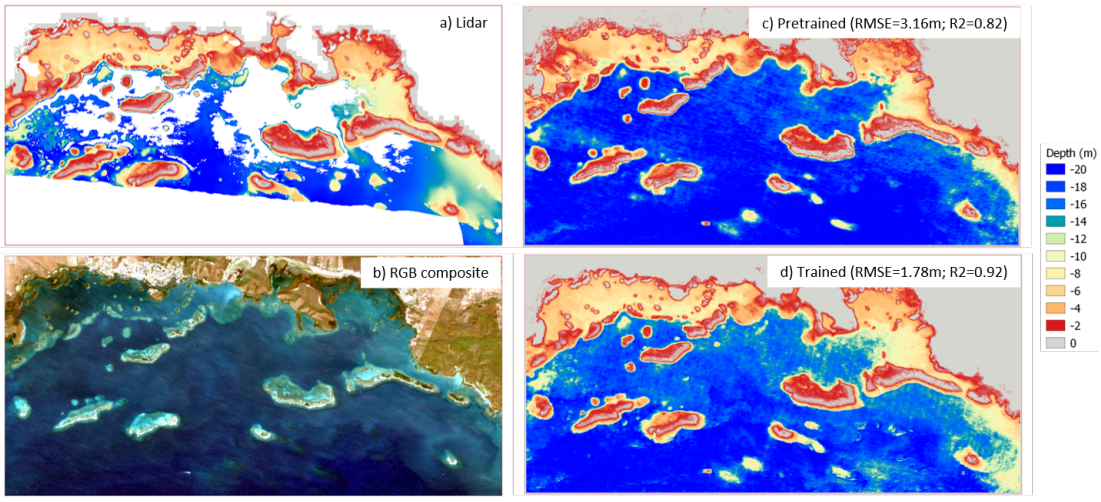


Figure 4.11: Visualization of AOI-2: the in-situ depths (a), the corrected image (b), SDB result using pre-trained CNN model from AOI-1 (c), and SDB result using trained model (d).

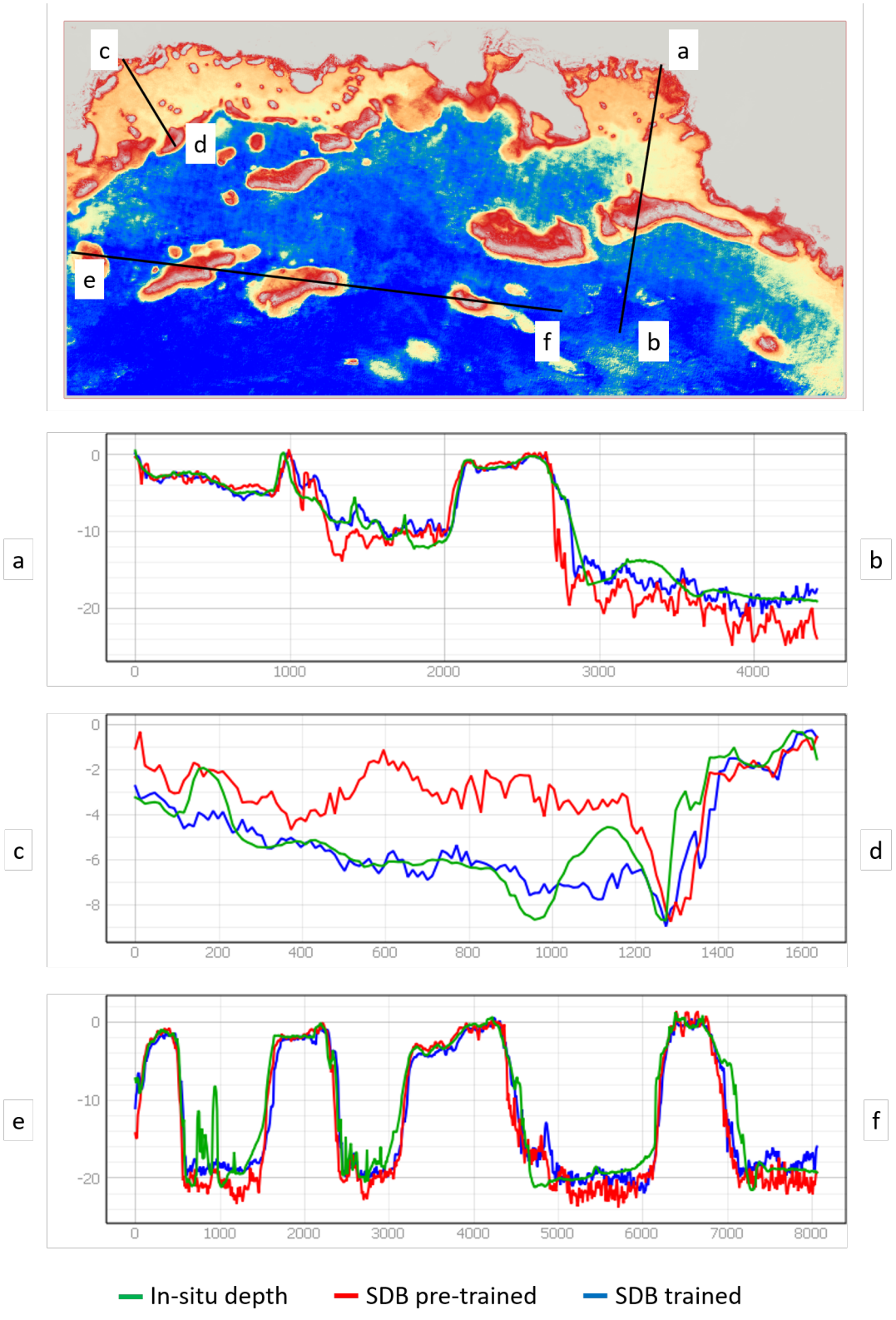


Figure 4.12: Cross profiles of AOI-2, showing comparison between SDB model using trained model, SDB using pretrained model from AOI-1, and in-situ depths.

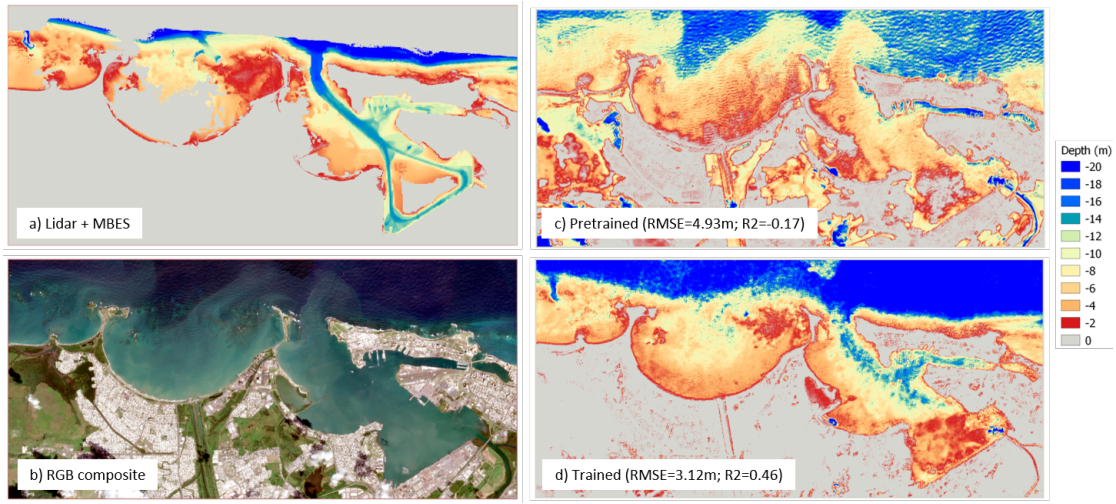


Figure 4.13: Visualization of AOI-3: the in-situ depths (a), the corrected image (b), SDB result using pretrained CNN model from AOI-1 (c), and SDB result using trained model (d).

The light penetration rate in water for each wavelength is different. Red, green, and blue are visible, so it is possible to distinguish different backscatter or absorption rates visually based on their natural colour composite. For example, water colour appears blue when red and green are mostly absorbed or white when they are mostly backscattered. In AOI-3, San Juan has a large port with different and more turbid water conditions than Ponce and Lajas. This condition causes different variations in the reflectance, so the relationship with depth is less clear. The variations are illustrated in Figure 4.9, where the reflectance values in AOI-3 are generally higher and noisier than AOI-1 and AOI-2 because the particles in the water column interfere with the light spectrum in all bands and cause a stronger backscattering effect in the water column. Meanwhile, AOI-4 has a different water condition compared to the study areas in Puerto Rico. Figure 4.10a shows that AOI-4 is more reflective than AOI-1 (Figure 4.10b), especially in the 0-5 m depth range, but more absorbed as well. The reflective pixels occur due to high nutrients, which correlated with more turbid waters than other areas. Some pixels at deeper depths are also more reflective than others due to the ships and their tracks captured in the image (Figure 3.6). The absorbed pixels, which appear darker in Figure 3.6, are due to consolidated substrate and coral. As shown in Figure 3.6a, the nearshore and estuary areas appear greener than other areas, indicating that green light is mostly backscattered while others are absorbed. Away from the nearshore areas, in areas deeper than 4 m, dark pixels appear where most red, NIR, and SWIR bands are completely absorbed, so the reflectance values are almost zero. Based on the spectral reflectance plot, AOI-4 has more extensive reflectance ranges than AOI-1. According to this situation, the satellite image correction process as described in Section 3.3 has problems when correcting zero pixels and produces gaps in the corrected image. Therefore, an adjustment was made before the transformation to avoid producing *nodata* values. Additionally, when bottom reflectance is extracted using DWA from a small area in deep water, their penetration rate in shallow water is adjusted to the DWA value. The correction itself may produce errors because of the generalization of the penetration rate to the entire study area, so the output value does not precisely represent the bottom reflectance.

Regarding the implementation of the pretrained model of AOI-1 to other areas, Figure 4.11c illustrates that by using the pretrained model, the SDB model in AOI-2 has an accuracy of 3.16 m with R^2 of 0.82. The SDB results using the pretrained (Figure 4.11c) and trained (Figure 4.11d) model generally indicate a similar trend with the in-situ depths (Figure 4.11a). The pretrained model is compatible since AOI-2 is located near AOI-1, in the south part of the Puerto Rico main island, so the water condition and the reflectance are similar since they came from the same image. However, some areas in the eastern part near the shoreline are not predicted well; similar things occur in deeper areas. Figure 4.12 provides the comparison between depth values of in-situ, SDB pretrained, and SDB trained in three cross sections: from north to south in the eastern (Figure 4.12a-b) and western (Figure 4.12c-d), and from west to east in the southern AOI-2 (Figure 4.12e-f). Figure 4.12a-b indicates that SDB of the pretrained model encoun-

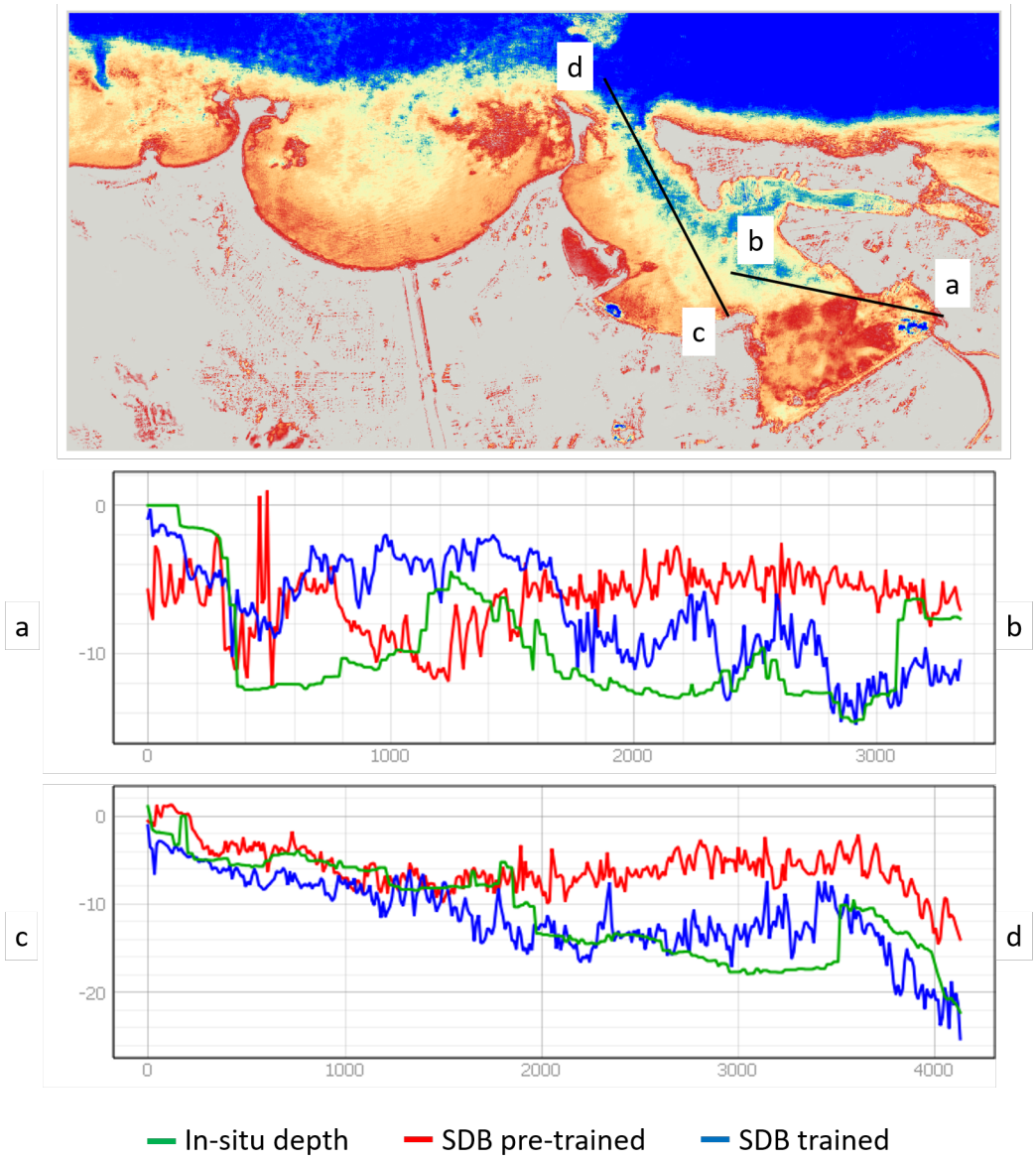


Figure 4.14: Cross profiles of AOI-3, showing comparison between SDB model using trained model, SDB using pretrained model from AOI-1, and in-situ depths.

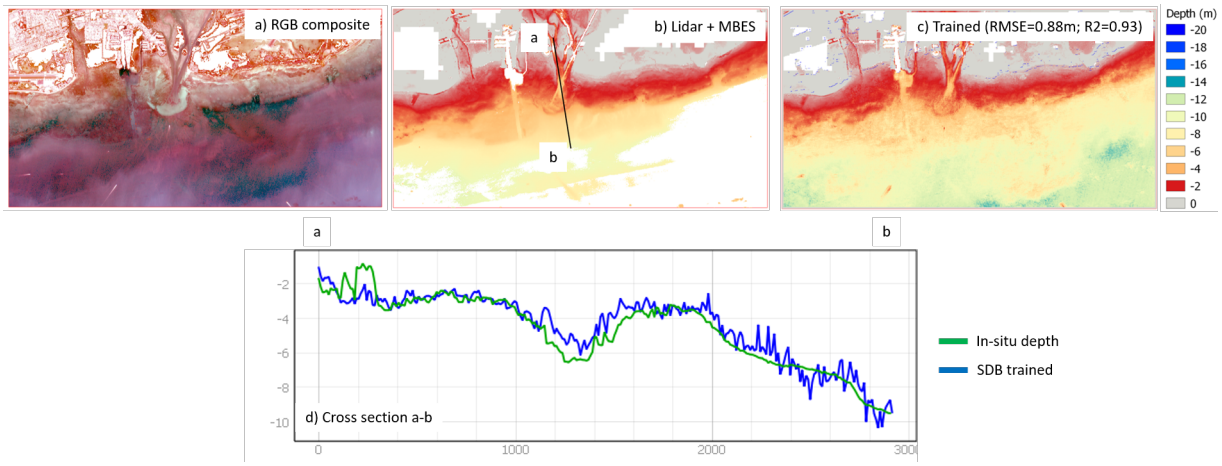


Figure 4.15: Visualization of AOI-4: the corrected image (a), the in-situ depths with cross section line in black (b), SDB result using trained model (c), and the cross profile extends from the inside of the estuary to the sea (d).

ters errors at the 10 m and more than 15 m depth. The SDB result using the pretrained model in the shallower areas in the western generally predicts shallower depths than the in-situ and the SDB trained (Figure 4.12c-d). This result is different with the same depth range as in Figure 4.12a-b. The reason is that the data training in AOI-1 is not enough to represent that area. Eventually, the model gets better when we have training data on that area. Figure 4.12d-e depicts that the result of the SDB pretrained and SDB trained generally tends to fit the in-situ depths.

On the contrary, Figure 4.13c shows that the pretrained model is not suitable for the San Juan area. The coefficient of determination yields a negative value. It means that the SDB model has an opposite trend compared to the in-situ measurements, most likely due to different water condition where AOI-4 is noisier due to turbid waters. Using training data from the San Juan area, the SDB model improved but is still missing some details. For example, there is a clear channel of the entry to the port shown in Figure 4.13a while it is missing in Figure 4.13d. The model is also unable to predict the shallower area nearshore accurately as the error can reach 10 m around the estuary.

Furthermore, Figure 4.14 shows two cross sections in the San Juan harbour area. The first cross profile (Figure 4.14a-b) provides the comparison between different SDB results and in-situ depths from the east around the nearshore area, close to the estuary, to the west. It shows that the errors occur almost at all depth levels, including the very shallow area (0-5 m) where the RMSE of SDB trained only achieve 2.06 m with R^2 of 0.11. Figure 4.14c-d shows a similar trend where the channel is not identified well in the SDB trained. In general, the SDB results predict shallower depths than the references. Besides the noisier reflectance due to sediments or pollutants in the turbid water column, the nearshore area, especially around the estuary, may encounter silting due to suspended solids from the river stream. This result indicates that SDB computation encounters difficulty in more turbid waters since the reflectance values from the bottom are obscured by the particles that cause turbidity.

In the following study area, AOI-4, some gaps were produced in the corrected image due to the correction using the DWA method. As mentioned before, an adjustment for the deduction value was made to avoid *nodata* values. The *nodata* values are produced when the mean DWA value is equal with the reflectance value, so the subtraction result is zero, and the logarithmic of zero is undefined. The adjustment was carried out by reducing the mean DWA value by 0.1 before the subtraction to avoid computing the logarithmic of zero. Figure 4.15a illustrates the corrected image after the adjustment; the darkest areas initially are the gaps. Using the RGBNSS corrected image, SDB in AOI-4 obtains an accuracy of 0.88 m with R^2 of 0.93. Meanwhile, the RGB combination produces an accuracy of 0.71 m with R^2 of 0.95. Figure 4.8c describes the accuracy per 1 m depth increment, in which they generally have comparable accuracy. The most considerable difference occurs at 8-9 m depth. The more significant errors are found in the dark areas near and away from the shoreline, around the dredged zone and estuary, and the area

Figure 4.16: Sample areas taken as data training for CNN to map three different turbidity levels. The levels are specified based on a visual observation of the AOI-3 image. An additional class, whic is land, is assigned to distinguish water and land regions.

where the ships are located.

Figure 4.15c provides the visualization of the SDB results with the default setting, which indicate a good prediction considering the estuary and dredging area are visible in the SDB output. Previous research by Caballero and Stumpf [2019] used the same area but to a larger extent. The study used LiDAR bathymetry acquired in 2016, having the same depth range as this project, and an image from Sentinel-2 Level-1C captured on 8 February 2017. The atmospheric correction was performed using the ACOLITE processor. The SDB was generated based on the ratio transform method, and the accuracy was evaluated in terms of MedAE, which is 0.39 m and 0.42 m. Using the same accuracy metric, the SDB using CNN in this research yields the accuracy of 0.31 m and 0.29 m using the RGBNSS and RGB bands, respectively.

Implementing the pretrained model from AOI-1 to AOI-4 indicates that the pretrained model is not suitable for AOI-4. The accuracy is not plausible for the 0-10 m depth range, where the RMSE is more than 100 m with R^2 of -6.81. Moreover, the pretrained model using the RGB combination still produces unreasonable errors that reach more than 40 m. Most errors occur in the darkest areas as presented in Figure 4.15a. In this case, the adjustment in the correction made the subtraction values lower than zero, and thus the logarithmic transform yielded negative values. Using the pretrained model from AOI-1, the CNN model cannot fit the data since the values are outside its range. With additional bands, more zero pixels exist since NIR and SWIR bands are easily absorbed in the water, so more negative values present in the corrected image and cause worse prediction as a result. Therefore, another approach to avoid this error was carried out and will be discussed in the following section.

4.3.1 Turbidity prediction

Concerning turbidity issues in retrieving shallow water depths from multispectral images, this project undertook an experiment to map the turbidity using CNN by considering different turbidity levels in different locations. This experiment aims to observe how is the connection between turbidity level and SDB accuracy. Several samples were taken from AOI-4, representing three different turbidity levels: low, medium, and high, in different depths. Figure 4.16 presents the areas that were used as training data. These data sets became the input for CNN training. This project used a similar CNN architecture as the baseline but different activation in the output layer where the softmax function was used instead of the linear. The learning rate was also adjusted to 0.001, and the polling layer was activated. The loss

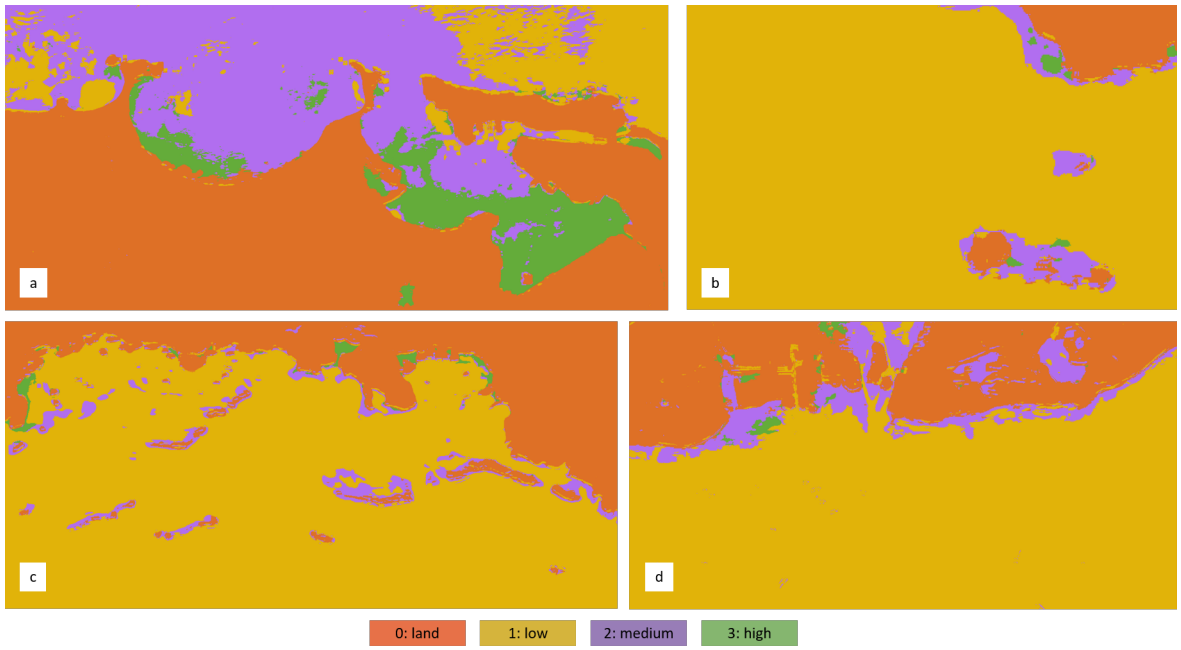


Figure 4.17: Turbidity map of AOI-3 (a), AOI-1 (b), AOI-2 (c), and AOI-4 (d) using CNN trained in AOI-3.

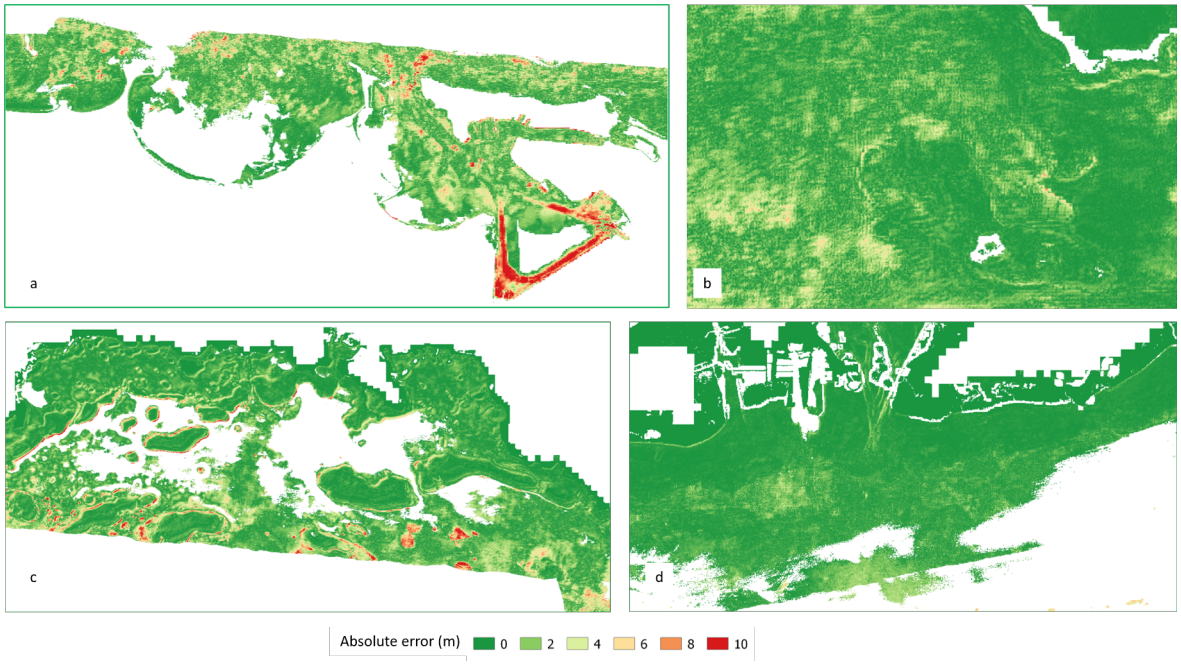


Figure 4.18: SDB error maps of AOI-3 (a), AOI-1 (b), AOI-2 (c), and AOI-4 (d) based on the results using CNN trained model with 9×9 window size and RGBNSS bands.

function changed to the sparse categorical cross entropy. The training validation yielded an accuracy of 0.97. The turbidity map at AOI-3, as illustrated in Figure 4.17a, shows that the nearshore area mostly classified as high turbid waters. Moreover, some of them are classified as land, indicating thick sediments in those areas. Additionally, the trained model is used to map turbidity in AOI-1 (Figure 4.17b), AOI-2 (Figure 4.17c), and AOI-4 (Figure 4.17d) as well. In general, the results indicate a similar trend where higher turbidity are found around nearshore areas. Specifically for the AOI-2, the turbidity map is conformable with several turbidity sampling sites provided in Bejarano and Appeldoorn [2013] where higher turbidity occurs at the edges of the small islands. However, since the training data only covers turbidity, which is mostly caused by sediment in the water, the result does not indicate turbidity due to phytoplankton concentration or algae bloom, where the RGB image appears green as in AOI-4.

These turbidity results are then compared to the error of SDB in each area of interest. Figure 4.18 demonstrates the absolute error per pixel depth in meter. Since the in-situ depths do not cover the entire area, the absolute errors only present at the same location as the measured data. In AOI-3, larger errors are found in high turbid areas such as inside the harbour, while lower errors are located in low turbid waters like in the eastern or western parts. Nevertheless, the correlation between turbidity and accuracy seems not observable at other locations. In AOI-1, AOI-2, and AOI-4, the turbidity over the water areas is low, but there is a variation in the accuracy, indicating that another source such as water column variations contributes more in causing errors in clear water areas.

4.3.2 Sensitivity to noise

To further evaluate the CNN model, this study introduced Gaussian noise to test the sensitivity of the CNN model. The input data values range from approximately -1 to 6. The gaussian noise was implemented within a variance of 1 cm to the Puerto Rico test data. The SDB model was computed 30 times, and then the standard deviation was calculated for each area. Figure 4.19 shows that the standard deviation ranges are varied between the study areas, from 0.3 cm to 92.2 cm. The distribution plots indicate that the predicted depths tend to be close to their mean or expected values, with a maximum average deviation of 0.15 m. It means that the CNN model has a low sensitivity to additional noise.

4.4 SDB results using uncorrected images

In the previous discussion, all images were corrected using deep water pixels to extract the bottom reflectance. Considering the variation of reflectance within the same depth, it is unfair to generalize the water scattering effect based on a specific area only. Moreover, in some cases, this correction method encounters difficulty due to the large absorption of lights by the molecules in the water column cause zero reflectance values. For example, when dark pixels exist in AOI-4 shallow water, the subtraction results were too small for the logarithmic transform and yielded zero or even error (*nodata*) values, which caused gaps in the corrected image as well as the SDB model. While this issue may be resolved by reducing the DWA value by a small number, which is 0.1, before the transformation to avoid the errors, this approach was not suitable if the SDB model is to be built using a pretrained model. Therefore, this study considers another approach. Instead of correcting the satellite images, this thesis attempts to use the Digital Number (DN) of surface reflectance values as input to train CNN, meaning that no correction is applied to the image. This approach considers that CNN is presumably capable of rectifying reflectance values due to the water column effect in training and assumes that the surface reflectance values contain enough information regarding its bottom. This experiment used default parameter settings for CNN training as shown in Figure 4.7. Table 4.5 provides the comparison of SDB model generated using corrected and uncorrected images with RGB and RGBNSS combinations. The accuracy per 1 m depth increment is shown in Figure 4.20. In general, the results indicate that CNN can also do the correction on the fly while training the data. The results in AOI-3 and AOI-4 do not significantly improve, but the accuracy in other areas is improved up to 0.15 m for the RGBNSS band and up to 1.04 m for the RGB channel.

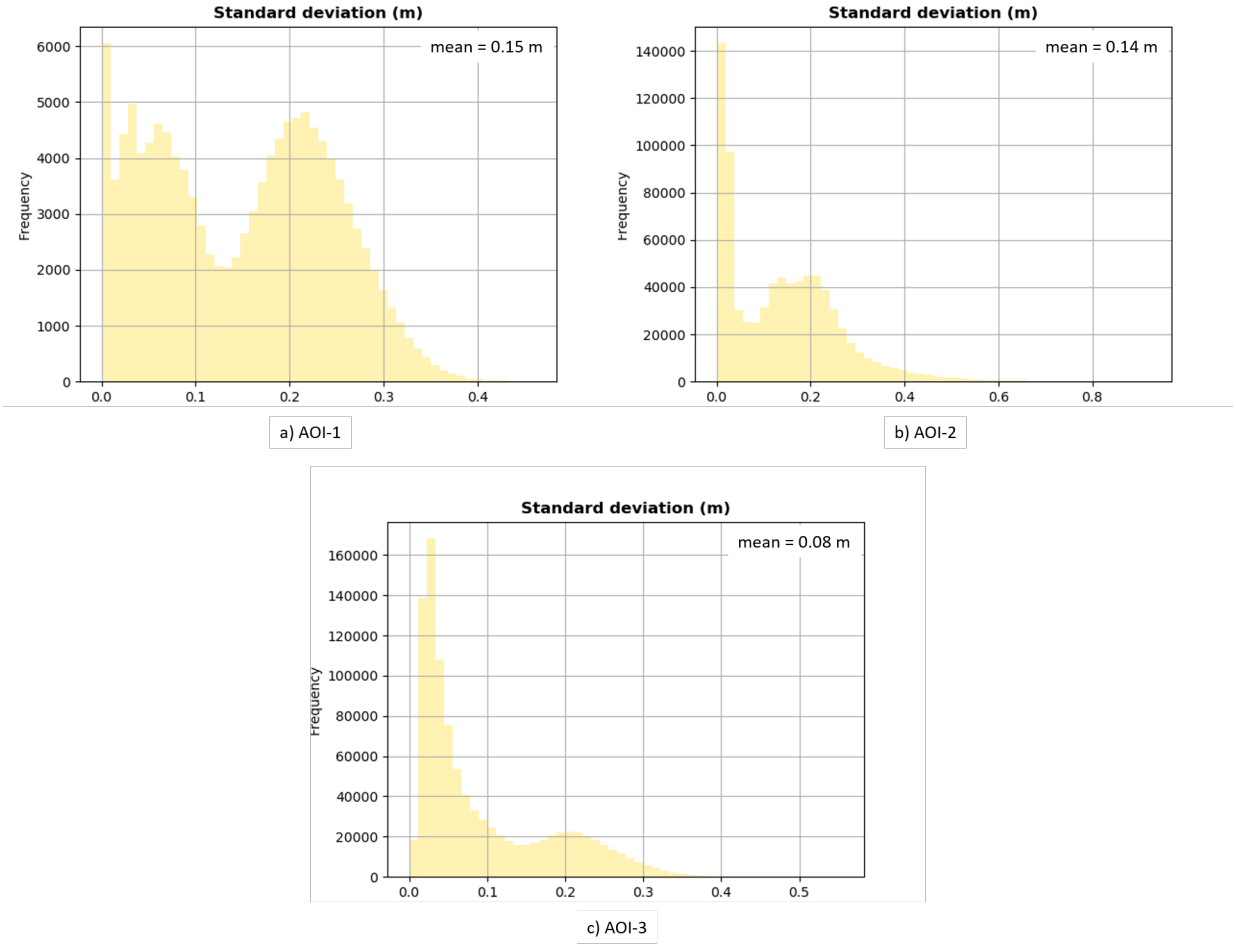


Figure 4.19: Standard deviation of SDB results with the addition of Gaussian noise to the test data in AOI-1 (a), AOI-2 (b), and AOI-3 (c).

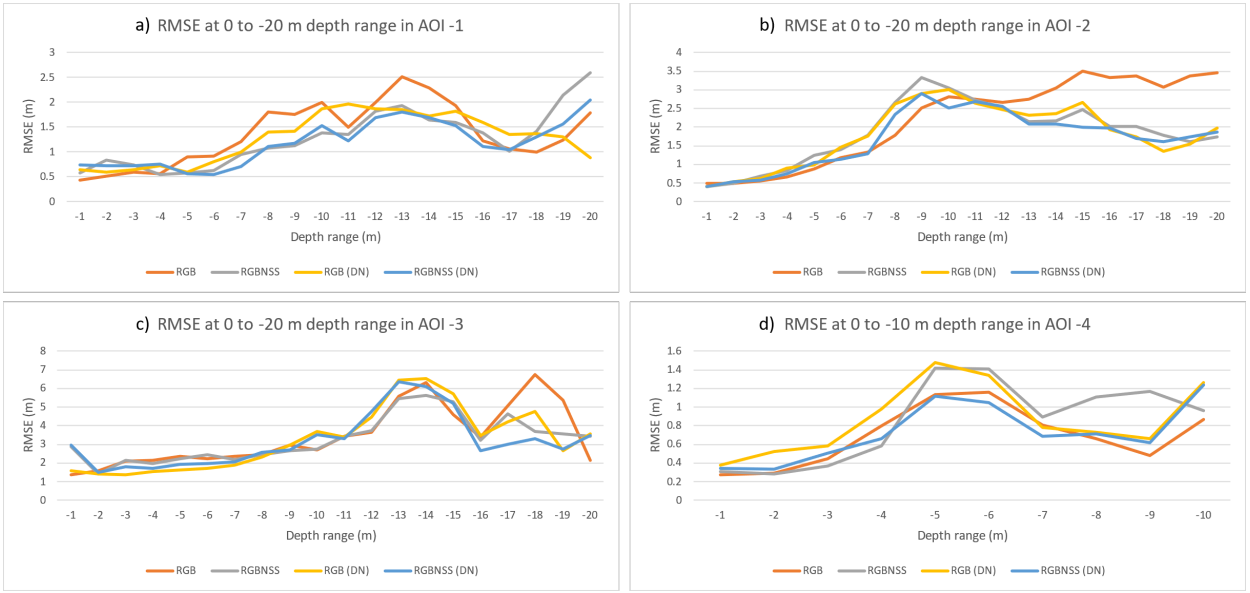


Figure 4.20: SDB accuracies at 0-20 m depth range in AOI-1 (a), AOI-2 (b), AOI-3 (c), and at 0-10 m depth range in AOI-4 (d). The SDB was computed using corrected and uncorrected (using DN) images, with different preprocessing, band combinations, and locations.

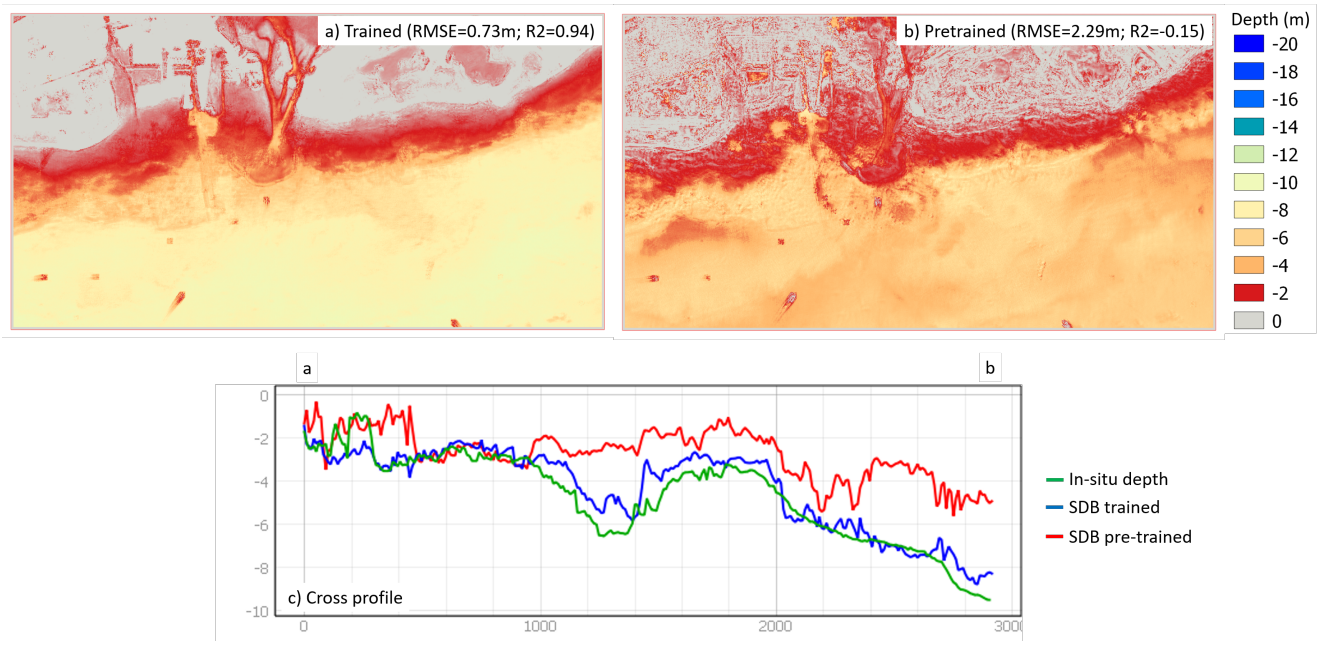


Figure 4.21: SDB results with the RGBNSS channels using uncorrected image (a), AOI-1 pretrained model (b), and the cross profile extends from the inside of the estuary to the sea (c). The cross section line is shown in Figure 4.15b.

Table 4.5: Comparison of SDB accuracy between corrected and uncorrected images as input for CNN training in different study areas and band combinations.

Images / AOI	RGB				RGBNSS			
	Corrected		DN		Corrected		DN	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
AOI-1	1.48	0.94	1.34	0.95	1.31	0.94	1.22	0.95
AOI-2	2.77	0.83	1.73	0.93	1.78	0.92	1.63	0.93
AOI-3	3.15	0.32	3.15	0.47	3.12	0.46	3.03	0.55
AOI-4	0.71	0.95	0.86	0.93	0.88	0.93	0.73	0.94

Back to implementing the pretrained model to AOI-4, using the pretrained model of AOI-1, the SDB yielded an accuracy of 2.29 m with R^2 of -0.15. Even though the result still shows the incompatibility of the pretrained model of AOI-1 to predict AOI-4 depths, the error this time shows a reasonable value. In general, the pretrained model predicts shallower depth than the trained model and the reference. The errors indicate that AOI-1 and AOI-4 have different water conditions so that the pretrained model cannot fit AOI-4 data.

Since uncorrected images result in good accuracy, this study uses uncorrected ones from this point forward. Additionally, Table 4.6 depicts the RMSE values for each increment of 5 m, which indicate that RGBNSS is more accurate than RGB combination in different depth ranges. Thus, the following experiments use the RGBNSS band.

Table 4.6: Comparison of SDB accuracy per 5 m depth increment using uncorrected images in different study areas.

Images / AOI	RGB				RGBNSS			
	0-5 m	5-10 m	10-15 m	15-20 m	0-5 m	5-10 m	10-15 m	15-20 m
AOI-1	0.64	1.24	1.84	1.37	0.71	0.99	1.60	1.35
AOI-2	0.69	2.31	2.48	1.70	0.66	2.03	2.22	1.77
AOI-3	1.52	2.48	5.44	3.92	1.84	2.50	5.31	3.03
AOI-4	0.76	0.95	-	-	0.57	0.89	-	-

4.5 SDB results using multi-temporal images

The number of multi-temporal images for each AOI are different due to diverse cloud percentage levels in every satellite image within one year. Figure A.7, Figure A.8, Figure A.9, and Figure A.10 present the RGB colour composite of the multi-temporal images. AOI-1 and AOI-2 have a complete set of images from January to December 2019. Meanwhile, AOI-3 and AOI-4 do not have suitable images in particular months due to cloud cover. Table 4.7 provides the RMSE and R^2 of SDB results in each AOI and each image. Figure 4.23, Figure 4.24, Figure 4.25, and Figure 4.26 illustrate the SDB results.

Figure 4.22 shows the comparison of RMSE values when using a single image and multi-temporal images in different location. Although there are no significant improvements of accuracy at AOI-1 compared to the results in Section 4.4, the RMSE at AOI-3, AOI-2, and AOI-4 decrease by 0.67 m, 0.19 m, and 0.17 m, respectively. These results indicate that multi-temporal images enhance training data variations and benefit in areas with more noise present in the image. However, the pretrained model of each location delivers different results when used to the test image. A wide range of accuracy is obtained; the lowest RMSE is 4.23 m at AOI-1, while the highest is 0.84 at AOI-4. Some anomalies are also observed, e.g. accuracies in AOI-1 ranges from 0.96 m to 4.81 m with R^2 vary between -1.39 and 0.97. The negative value of R^2 suggests that the prediction depths deviate significantly from its reference values. The

Table 4.7: SDB accuracy using multi-temporal images in different locations. Test images have different dates, except for AOI-1 and AOI-2. AOI-1 and AOI-2 use image on 8 January 2019 as test data. Meanwhile, AOI-3 and AOI-4 use images on 12 and 8 February 2019, respectively. In addition, the MedAE of the SDB in AOI-4 on February is 0.23 m.

AOI / Month	AOI-1		AOI-2		AOI-3		AOI-4	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Overall	2.05	0.85	2.06	0.90	2.93	0.51	1.03	0.89
January	1.11	0.96	1.44	0.95	n/a	n/a	0.82	0.92
February	1.10	0.96	1.22	0.96	2.36	0.70	0.56	0.97
March	0.96	0.97	0.93	0.98	2.52	0.66	0.55	0.97
April	1.28	0.94	1.32	0.96	2.48	0.60	0.54	0.97
May	1.66	0.89	1.32	0.96	n/a	n/a	n/a	n/a
June	1.29	0.93	1.27	0.96	n/a	n/a	0.63	0.96
July	1.22	0.95	1.54	0.95	2.21	0.70	n/a	n/a
August	1.20	0.95	1.35	0.96	2.22	0.67	n/a	n/a
September	1.00	0.96	1.32	0.96	2.03	0.77	0.58	0.97
Ocotber	2.38	0.85	1.72	0.93	2.59	0.63	1.64	0.59
November	4.81	-1.39	5.36	-1.71	4.01	0.25	n/a	n/a
December	2.96	0.75	1.91	0.91	4.76	-1.80	1.84	0.66
Test	4.23	-0.34	2.65	0.76	2.77	0.55	0.84	0.93

variation of the spectral reflectance is not only observed in AOI-1 (Figure A.7), but also in other areas, shown by their multi-temporal RGB composite images (AOI-2: Figure A.8, AOI-3: Figure A.9, and AOI-4: Figure A.10). SDB results in AOI-2 are similar with AOI-1, where the SDB result from the November image indicates an anomaly compared to other months (Figure 4.23 and Figure 4.24). The November output at both AOIs depicts an entirely different depth range where the overall prediction appears shallower than other outputs. The December result also illustrates a difference around the western, deeper depth, and the east near the shore, where the prediction is deeper than other results. In AOI-3, the abnormality is demonstrated in the November and December results as shown in Figure 4.25, where deeper waters inside and outside port appear shallower, especially in December. Within these months, the prediction at the harbour area shows a completely different range compared to others (Figure 4.25). Meanwhile, AOI-4 experienced the irregularity on October and December models (Figure 4.25), when the error is three times higher than the optimum accuracy.

An image in AOI-1 on 24 November 2019 was trained alone to analyze the anomaly further. The result

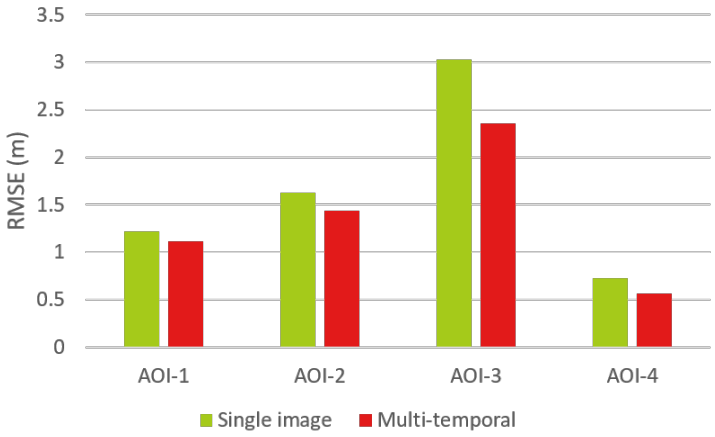


Figure 4.22: SDB accuracy comparison between results with a single image and multi-temporal images.

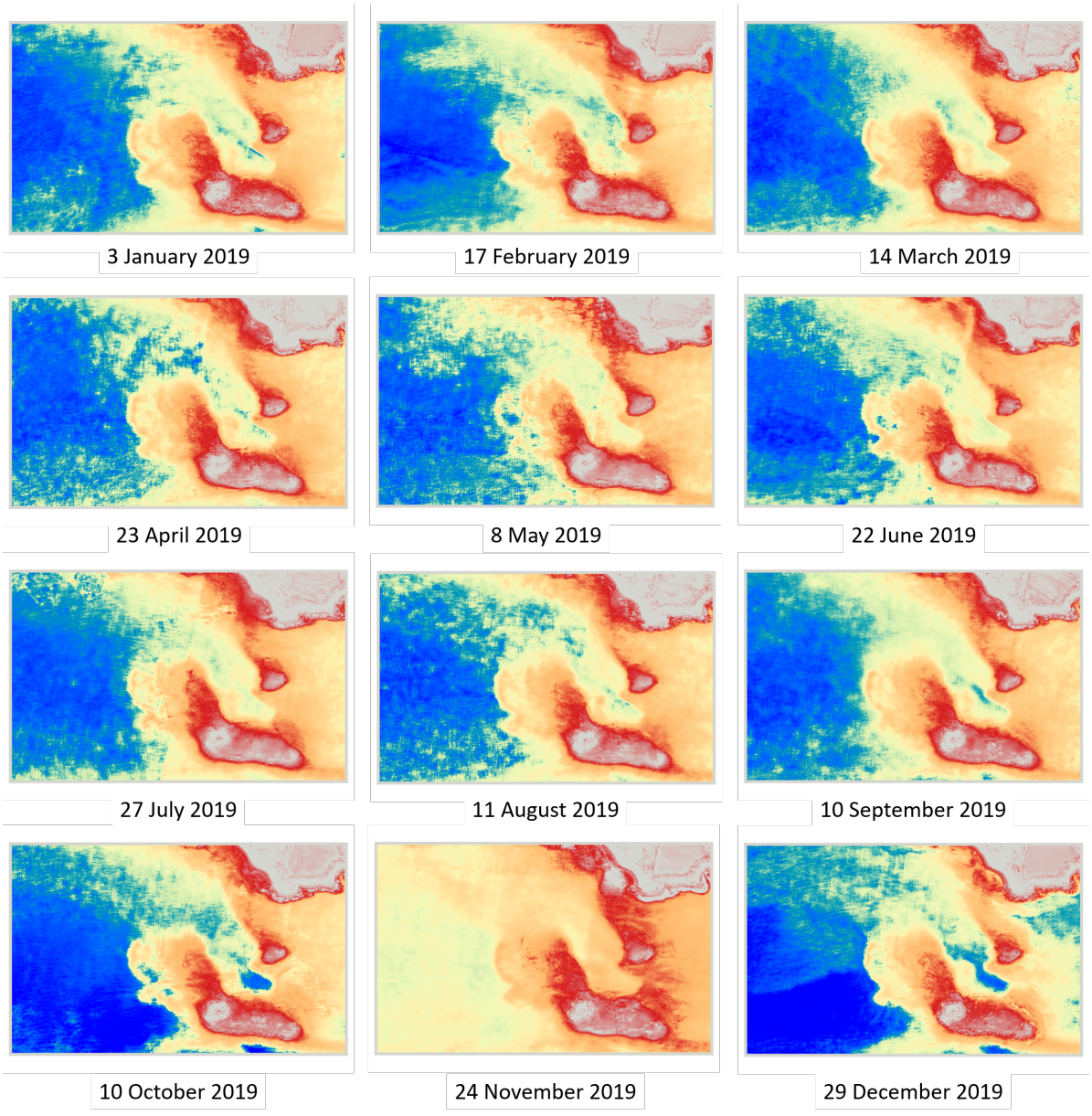


Figure 4.23: SDB results in AOI-1 from January to December 2019.

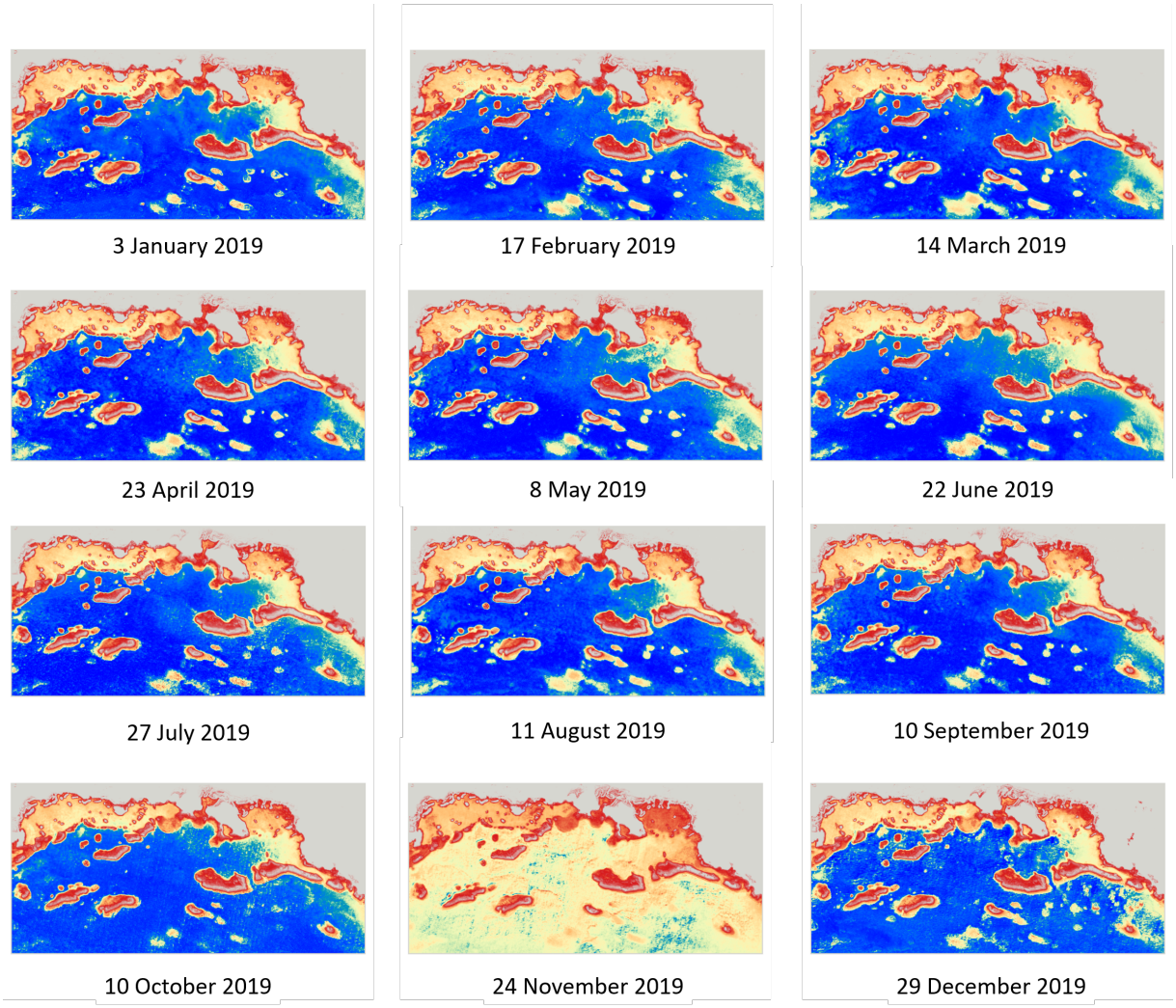


Figure 4.24: SDB results in AOI-2 from January to December 2019.

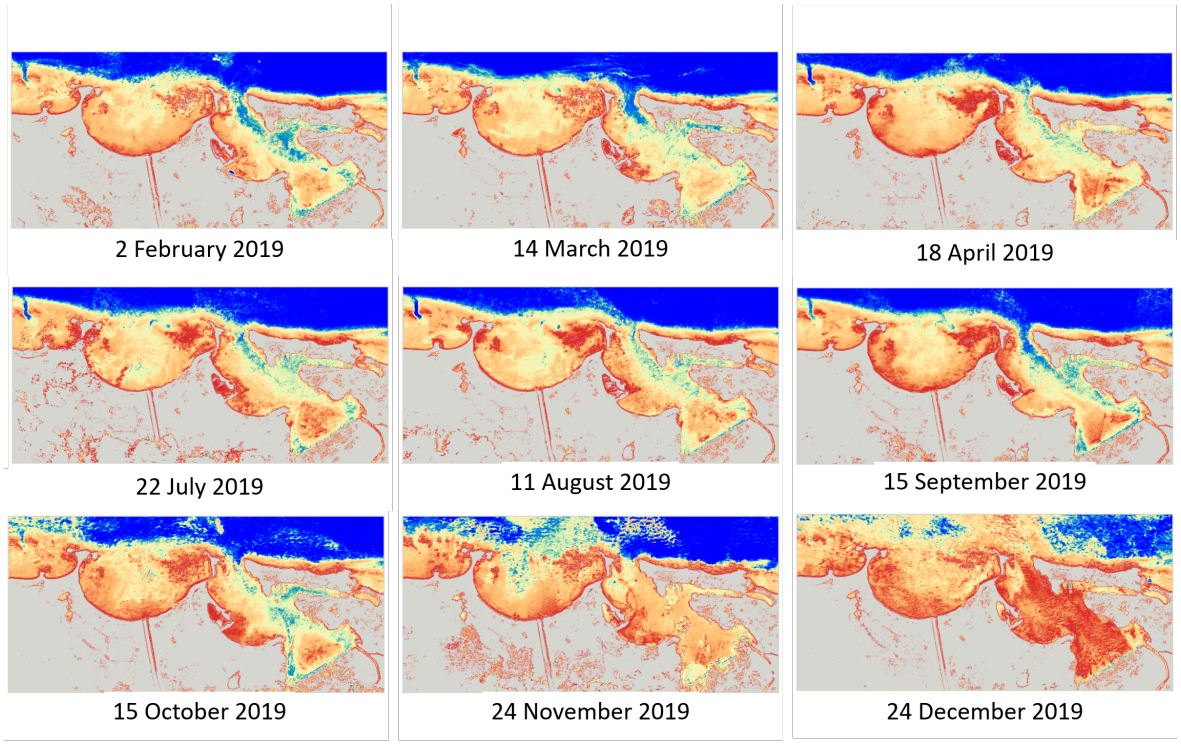


Figure 4.25: SDB results in AOI-3 from February to December 2019.

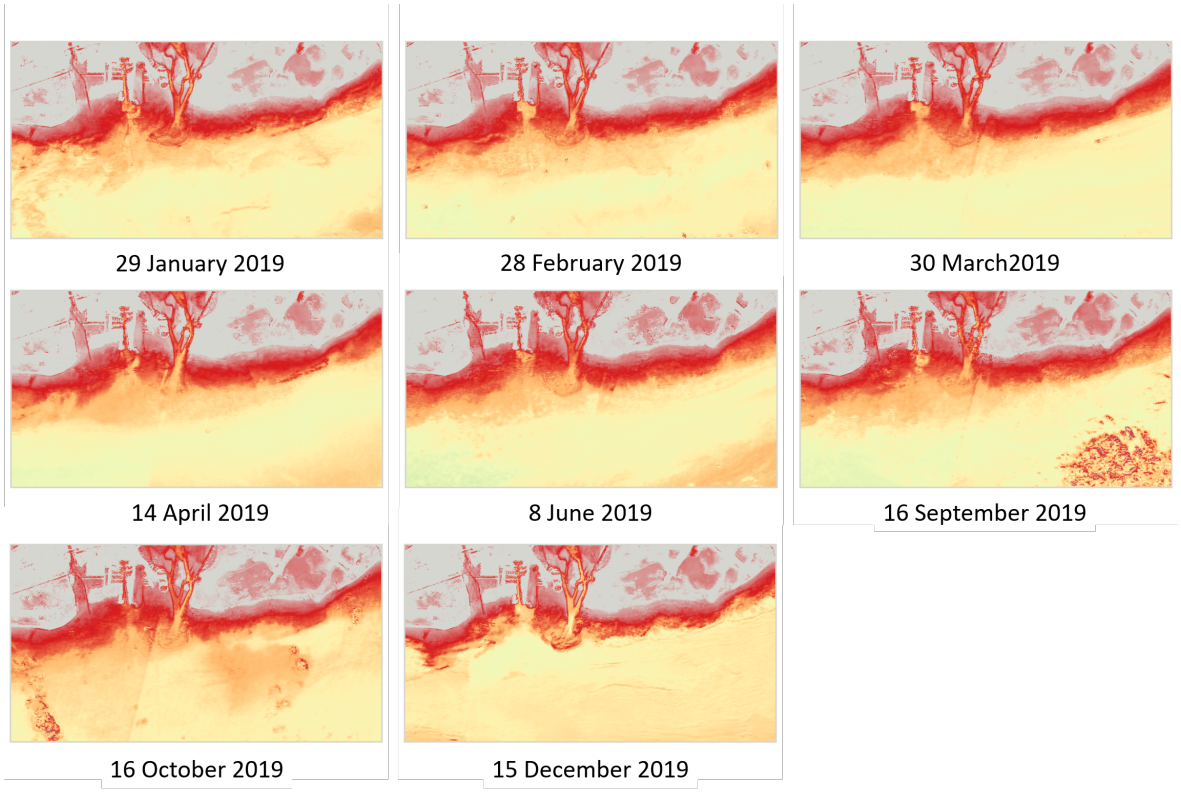


Figure 4.26: SDB results in AOI-4 from January to December 2019.

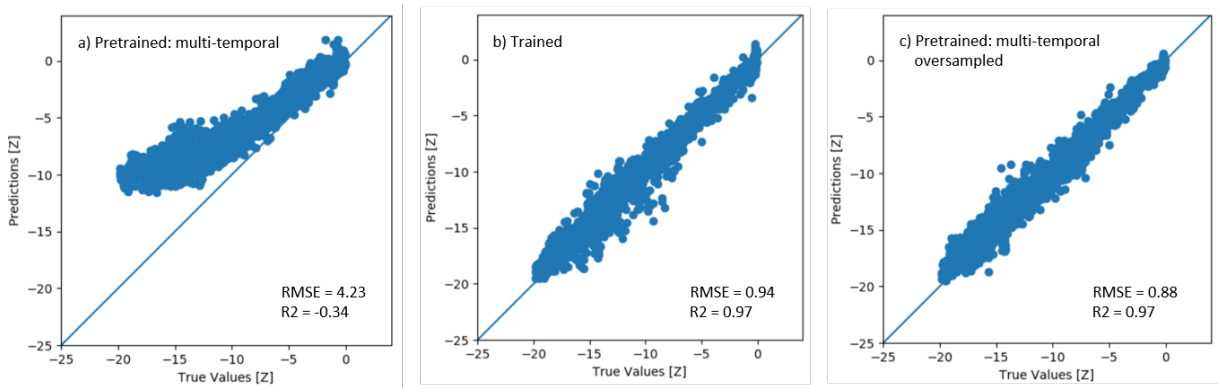


Figure 4.27: Scatter plots of prediction vs in-situ depths at AOI-1 in different cases: using multi-temporal trained model to an image on 24 November 2019 (a); result on 24 November 2019 using individual trained model on the same image (b); result on test image (8 January 2019) using pretrained model on 24 November 2019

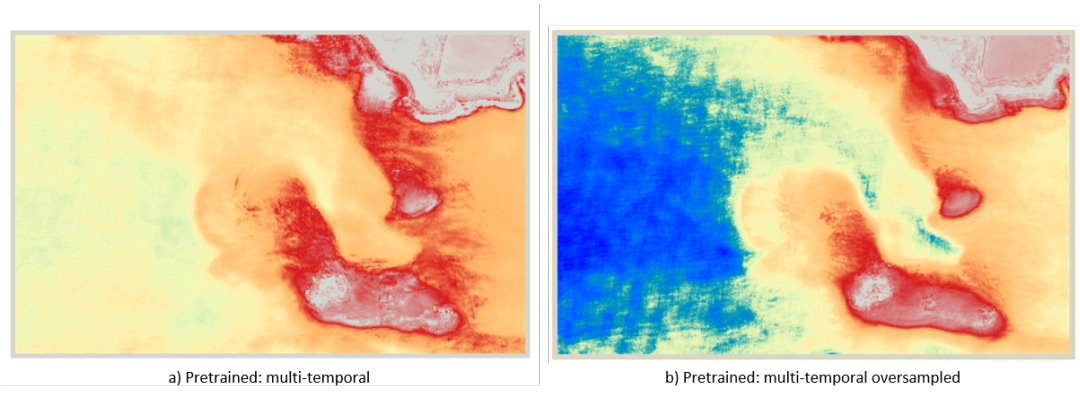


Figure 4.28: Comparison between SDB models on 24 November 2019 using pretrained multi-temporal and pretrained multi-temporal with oversampling.

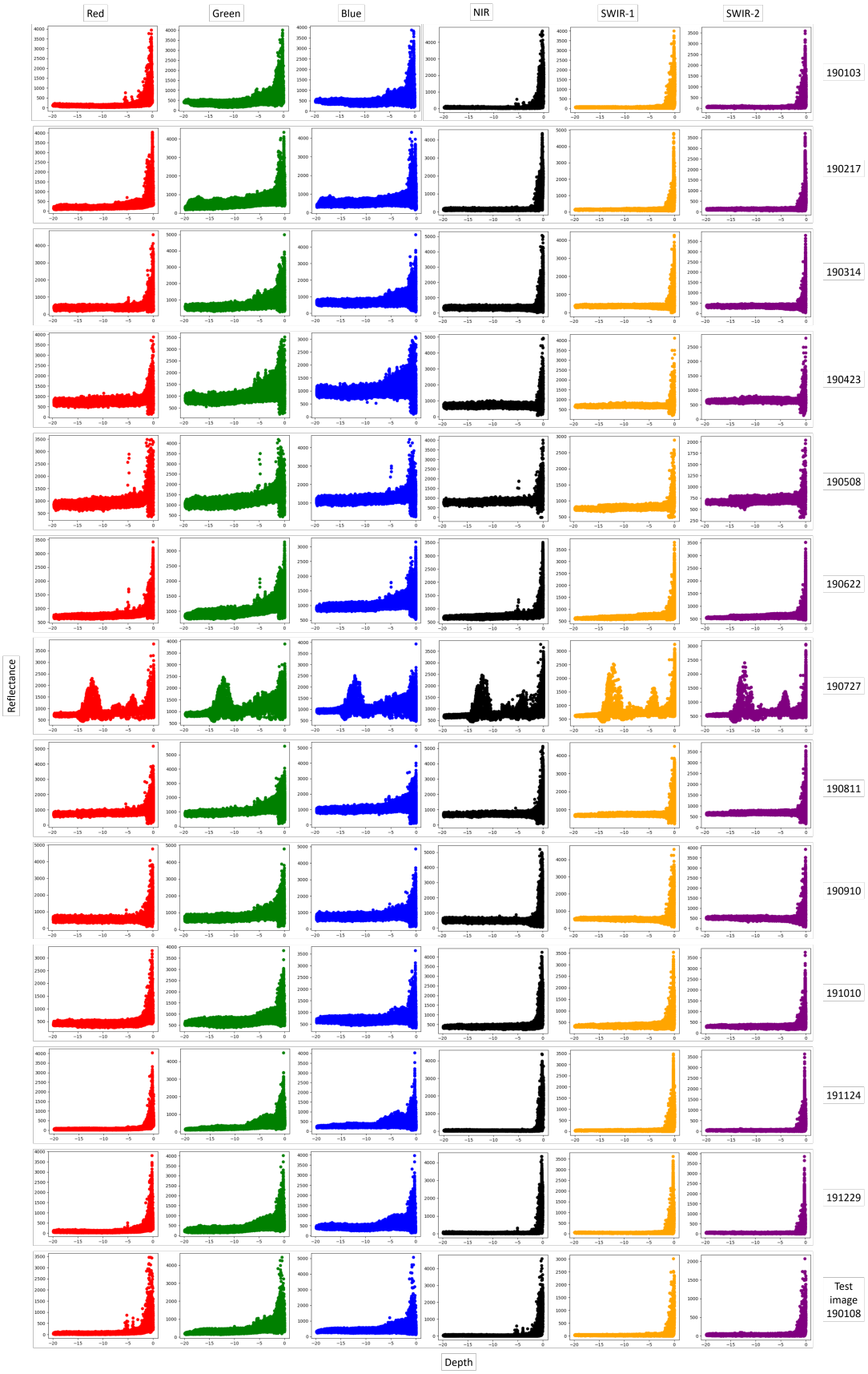


Figure 4.29: Spectral reflectance signature towards depth of multi-temporal images in 2019 at AOI-1.

was then compared with the one from multi-temporal data sets. Figure 4.27a demonstrates the result on 24 November 2019 image using multi-temporal trained model, where the deviation increases as the depth increases. The SDB is generally predicted shallower depths than the reference, with the maximum depth prediction only reaching 12 m. Figure 4.27b depicts that when the image is trained separately, a normal accuracy value of 0.94 m with R^2 of 0.97 is obtained. This result does not indicate any issues regarding the image itself. The errors may occur when the variations are completely different, i.e. the reflectances are way too large or too low compared to the available data sets, so the sample data to represent these variations is insufficient in terms of numbers. It seems that unbalanced data affects the training of CNN. Thus, this study increases the number of training data by simply copying sample data on 24 November 2019 once. As a result, the accuracy to November image was improved significantly as shown in Figure 4.27c and Figure 4.28b. The SDB model, especially in deeper areas, is now consistent with other SDB models. Therefore, the oversampled training data improves the accuracy of SDB in the November image, which is 0.88 m and 0.97 for the RMSE and R^2, while preserving the accuracy in other months.

The multi-temporal images indicate a seasonal variation shown by various image tones where the images appear greener between October and March than in other months, e.g. Figure A.8 illustrates a contrast where the vegetation on land looks green in particular months and darker in the others. Figure 4.29 shows the spectral reflectance plots with respect to depths. This observation pinpoints that each image has a different reflectance signature towards the same depth. If we observe Figure 4.29, the reflectance signatures change according to Puerto Rico's seasonal changes. The dry season in Puerto Rico begins in December and lasts until March, while April brings the wet season, which lasts through November. June and July are usually the driest months, while April and November are the transition period. Similar transition change according to the seasonal change in the spectral signature is also detected, where the reflectances are generally higher around April to September and begin to decrease in September. November image has the lowest spectral reflectance among other months. In July, clouds cover the waters in the northwestern lead into fluctuated reflectances in deeper waters. A more contrast variation occurs in AOI-4 (Figure A.10), where the images can appear dark, greenish, or bluish in different months. Figure A.13 illustrates various reflectance ranges where some light spectrums are higher or lower than others in different months and different depth ranges. Accordingly, when all bands are mostly absorbed, the image appears darker than others. Meanwhile, when green or blue reflect more energy than other bands, the image appears greenish or bluish. Therefore, including several images in different seasons increases the variety of sample data for training, thus increasing the possibility of CNN to generalize the model.

4.6 Transfer model analysis

AOI-5 and AOI-6 represent other locations used to analyze the SDB result when a pretrained model is reused. They are captured in different footprints on the same date acquisition, which is 10 December 2019. The accuracy using the trained model in both areas generally indicates an accurate prediction. Specifically for AOI-5, the same area has been studied by Sagawa et al. [2019] using another machine learning approach, i.e. Random Forest, and Landsat-8 surface reflectance product, which are atmospherically corrected. That research used 26 multi-temporal images from 2013 to 2018, and 9,974 data pixels were used for training. The maximum accuracy, in terms of RMSE, was 1.24 m with R^2 of 0.93. Compared to the same area, with less training data (5,851), this study yields an RMSE of 0.79 m with R^2 of 0.98. Table 4.8 provides RMSE and R^2 values in AOI-5 and AOI-6 using various CNN models that were calculated in Section 4.5.

Figure 4.30 to Figure 4.35 demonstrate prediction vs ground truth plots, SDB models, and absolute error map in AOI-5 and AOI-6, respectively. SDB models in AOI-5 and AOI-6 using the pretrained models indicate discrepancy with the in-situ depths or SDB trained. When the pretrained model of AOI-1 is used to build the SDB model in AOI-5, the prediction results tend to fit the references until 15 m depth (see Figure 4.30b). However, deeper than that, the calculated depths include large errors where the values are predicted to be shallower or deeper than the measured depths. In general, the SDB models become deeper when using other pretrained models, except when using pretrained model at AOI-4 where the result shows a

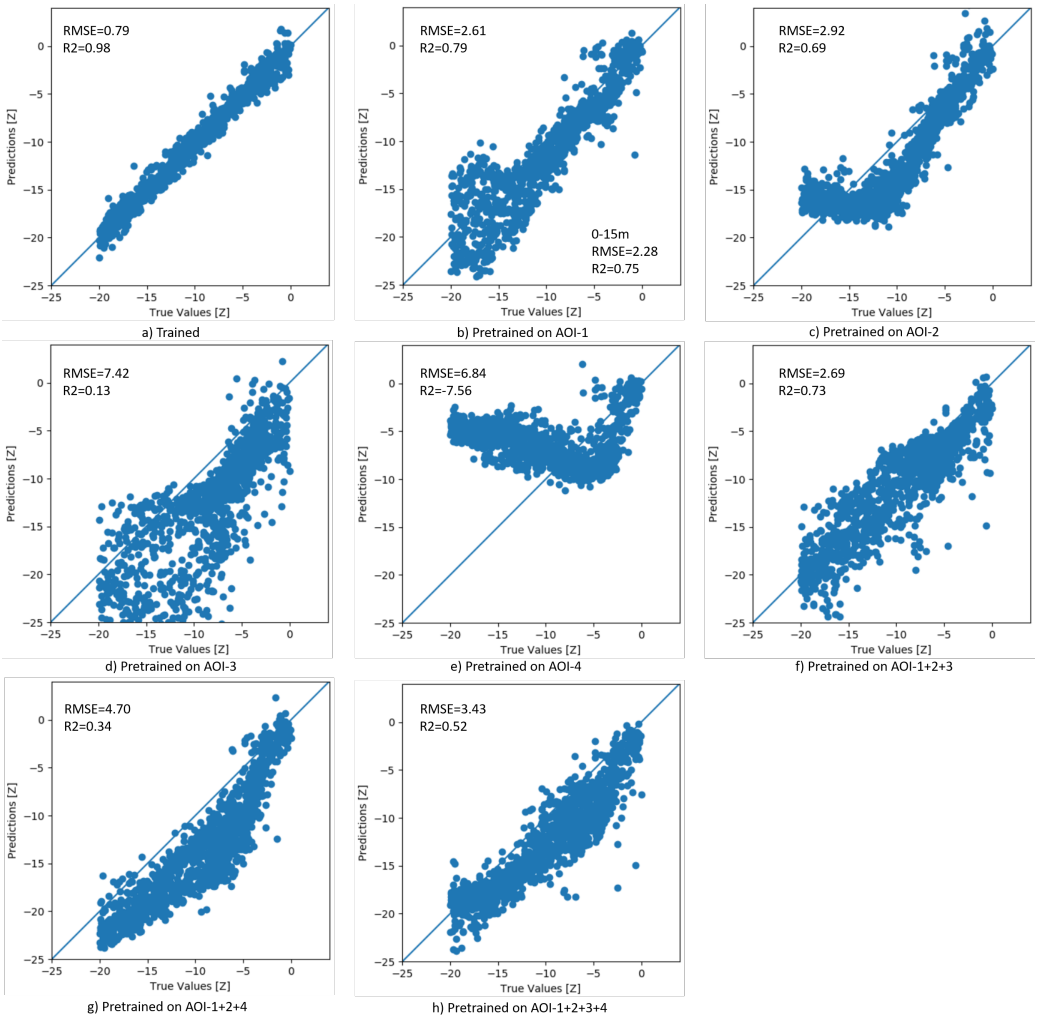


Figure 4.30: Prediction vs reference depth plots in AOI-5 using different CNN models: trained (a); pretrained on AOI-1 (b); pretrained on AOI-2 (c); pretrained on AOI-3 (d); pretrained on AOI-4 (e); pretrained on the combined AOI-1, AOI-2, and AOI-3 (f); pretrained on the combined AOI-1, AOI-2, and AOI-4 (g); pretrained on the combined AOI-1, AOI-2, AOI-3, and AOI-4 (h).

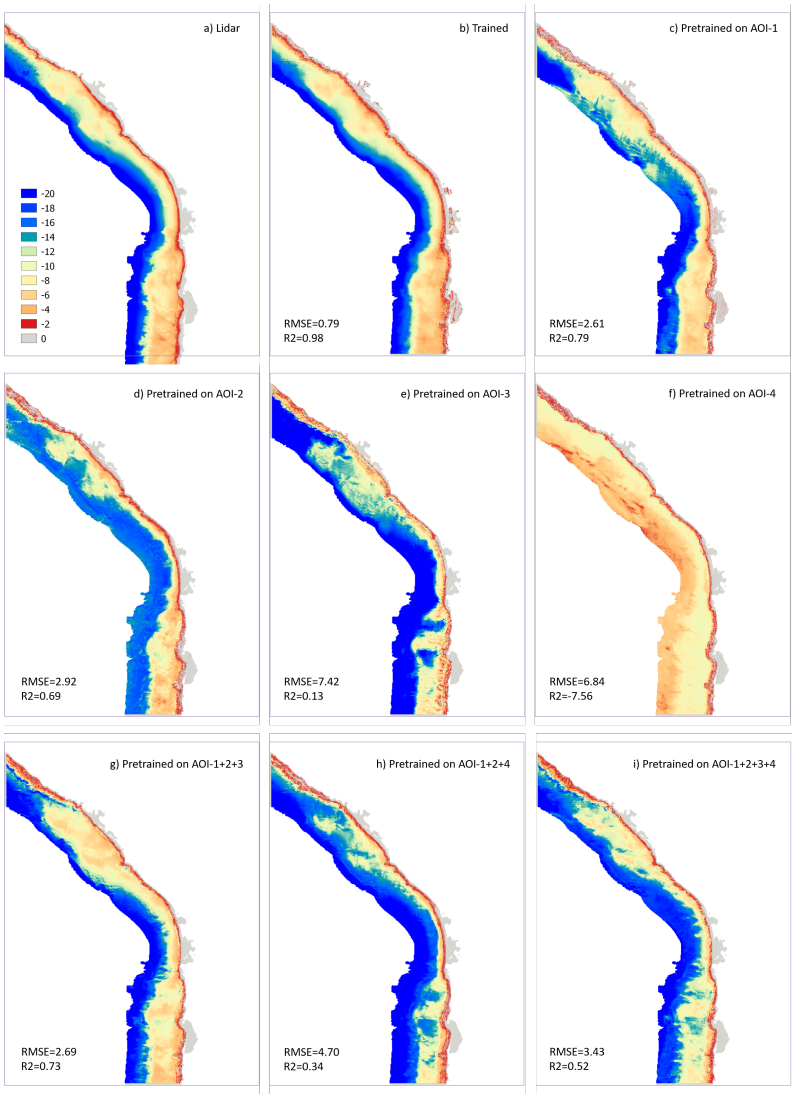


Figure 4.31: LiDAR bathymetric depths (a) and SDB results in AOI-5 using different CNN models: trained (b); pretrained on AOI-1 (c); pretrained on AOI-2 (d); pretrained on AOI-3 (e); pretrained on AOI-4 (f); pretrained on the combined AOI-1, AOI-2, and AOI-3 (g); pretrained on the combined AOI-1, AOI-2, and AOI-4 (h); pretrained on the combined AOI-1, AOI-2, AOI-3, and AOI-4 (i).

Table 4.8: SDB accuracy in AOI-5 and AOI-6 using trained model and pretrained model of different study areas.

AOI / CNN model	AOI-5		AOI-6	
	RMSE	R^2	RMSE	R^2
Trained	0.79	0.98	1.34	0.94
Pretrained on AOI-1	2.61	0.79	3.16	0.73
Pretrained on AOI-2	2.92	0.69	2.98	0.71
Pretrained on AOI-3	7.42	0.13	6.42	0.16
Pretrained on AOI-4	6.84	-7.56	5.56	-4.72

completely different trend as presented in Figure 4.30e with absolute error illustrated in Figure 4.32e. It shows that the pretrained model from AOI-4 is inefficient to predict areas deeper than 10 m. Moreover, errors in Figure 4.31c mostly appear in deeper areas, with a sandy bottom, than shallow waters, with coral reef and hardbottom; especially from the northern to the middle (see Figure 4.32b).

For another comparison, Figure 4.33 and Figure 4.34 portray different SDB results in AOI-6. As mentioned in Section 3.1, Honolulu is another port area, so the disturbance of reflectance due to water turbidity is more substantial than in Makua. More dredged bottoms also appear in this area. Although the result is not as accurate as in AOI-5 within the same depth range, the RMSE is better than in San Juan port. Figure 4.33a points out that several depth values are predicted to be too shallow, and Figure 4.34a shows that it occurs in the western channel, which is a dredge zone. The other dredged areas also appear shallower than the measured depth. This circumstance is similar to the San Juan area. However, even though they share the same characteristics around the harbour, the pretrained model from San Juan is incompatible for Honolulu as shown in Figure 4.33d, Figure 4.34e, and Figure 4.35d. Additionally, the pretrained model of AOI-4 is also inefficient to calculate the SDB in AOI-6. Therefore, the pretrained model of AOI-4 is incompatible for Oahu regions as the shallow water calculated deeper and vice versa. The fact that the pretrained model of AOI-4 resulting in shallower prediction at 10-20 depth range related to the fact that AOI-4 training data only covers 0-10 m depths. Besides different reflectance characteristics, in the case of AOI-4, the pretrained model does not represent the entire depth range.

Another study using the radiative transfer model in AOI-6 has been performed by Lyzenga et al. [2006] using an IKONOS multispectral image. With a combination of empirical and analytical methods, the study yielded an RMSE of 3.01 m. In this study, the accuracy of the SDB using the trained model is 1.34 m, while the optimum accuracy when reusing the pretrained model from another location is 2.98 m. The simplified radiative model delivers a similar accuracy compared to the computed SDB using the pretrained CNN model, but it still requires in-situ depths within the study area to derive the physical parameters empirically. In the case of the transfer model of CNN, no measured depth is needed anymore, and so the SDB can be derived directly. However, the SDB generated using the pretrained model has a wide range of accuracies. Thus, it is not easy to choose which pretrained model should be used. So, extra knowledge about the coastal water characteristics is required to select an appropriate pretrained model. Also, it is better to keep verifying the result with the measured depth.

Furthermore, to produce a more generalized CNN model, this project tries to combine training data from several different locations. The assumption is that the combined training data will enrich the variation of reflectances concerning different water column compositions and different bottom types in different depth ranges. Nevertheless, SDB results using several training data show inaccurate outputs compared to the single data from AOI-1 or AOI-2. Unbalanced training data may become an issue because a particular location provides more information than others. This study tries to balance the data in a combination of AOI-1, AOI-2, AOI-3, and AOI-4 by undersampling and oversampling, taking the number of training data per location into consideration. For the undersampling case, the proportion of training data in AOI-2 and AOI-4 was reduced by 50%, considering the number of training data in AOI-2 is double that of others and AOI-4 only covers 0-10 m depth range. However, the accuracy is lower than the original combination, which is 4.70 m and 0.34 for the RMSE and the R^2, respectively. It occurs possibly because the undersampling method removes some details in the training data. On the other hand,

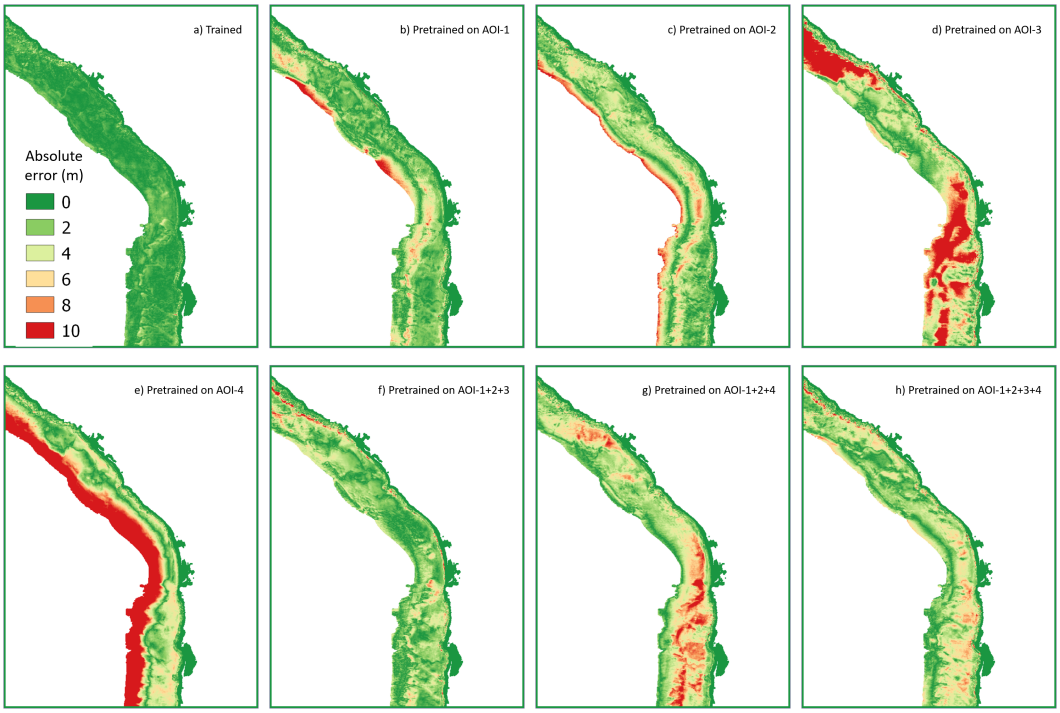


Figure 4.32: Absolute error of SDB per pixel when using different CNN models: trained (a), pretrained on AOI-1 (b); pretrained on AOI-2 (c); pretrained on AOI-3 (d); pretrained on AOI-4 (e); pretrained on the combined AOI-1, AOI-2, and AOI-3 (f); pretrained on the combined AOI-1, AOI-2, and AOI-4 (g); pretrained on the combined AOI-1, AOI-2, AOI-3, and AOI-4 (h).

the oversampling case increases the number of training data in AOI-1 by copying the existing data. The oversampling method also obtains inaccurate results with RMSE and R^2 of 7.4 m and 0.26, respectively.

These results pinpoint that SDB calculation using a pretrained model is likely affected by water characteristics. In this case, AOI-5 and AOI-6 water characteristics are more alike with AOI-1 or AOI-2 than AOI-3 or AOI-4. The combination of pretrained models shows the difficulty of CNN to fit the training data from different locations to a new location, even if the new location has similar characteristics to one of them. A rigorous technique to create balance training data should be considered to study the transfer model comprehensively. Several water characteristics, such as water colour, chlorophyll concentration, turbidity, should be considered while combining training data from different locations, even though they have the same bottom types.

4.7 SDB model production of Oahu island

Using the pretrained models from AOI-5 and AOI-6, the SDB model of the entire Oahu island is produced from three different images that captured different parts of the island. Figure 4.36 to Figure 4.38 present the SDB model, LiDAR bathymetry data, and the absolute error of the SDB model within 0-20 m depth range. Figure 4.36 illustrates that the two pretrained models within the same location are sufficient to provide a decent model. The overall RMSE of SDB Oahu is 2.24 m. In different depth ranges, the accuracy tends to increase as the depth increase. The accuracy of a 5 m depth span is 1.95 m, 2.05 m, 2.59 m, and 3.01 m, for 0-5, 5-10, 10-15 and 15-20 m depths respectively.

Larger errors occur in particular areas where their reflectances may not have been covered in the two pretrained models. For example, Figure 4.38 grid *a* experiences an error almost of up to 20 m. The area shows several channels in AOI-6 but looks like a natural estuary, not a dredged one. Another example is the *b* grid, where the bottom type classification resembles the two sample areas, but their mix is more

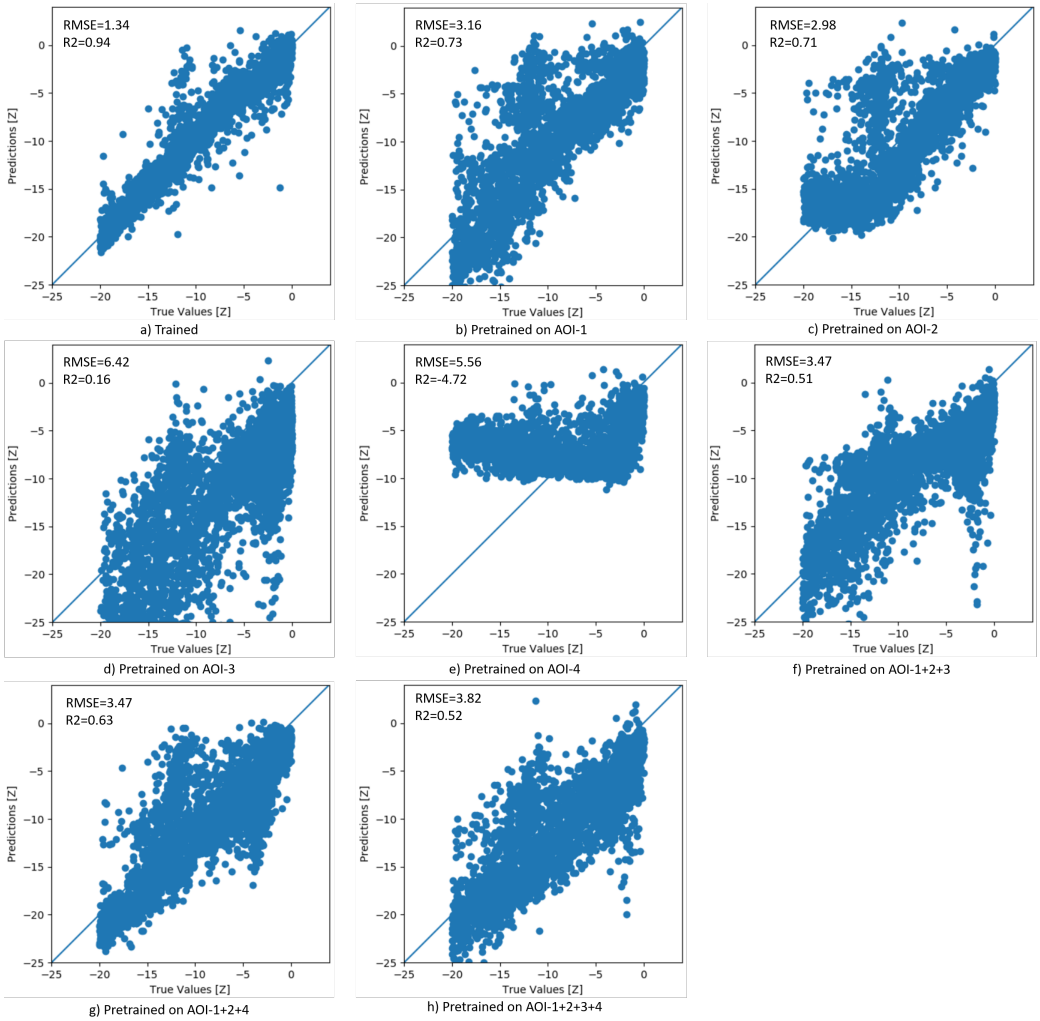


Figure 4.33: Prediction vs reference depth plots in AOI-6 using different CNN models: trained (a); pretrained on AOI-1 (b); pretrained on AOI-2 (c); pretrained on AOI-3 (d); pretrained on AOI-4 (e); pretrained on the combined AOI-1, AOI-2, and AOI-3 (f); pretrained on the combined AOI-1, AOI-2, and AOI-4 (g); pretrained on the combined AOI-1, AOI-2, AOI-3, and AOI-4 (h).

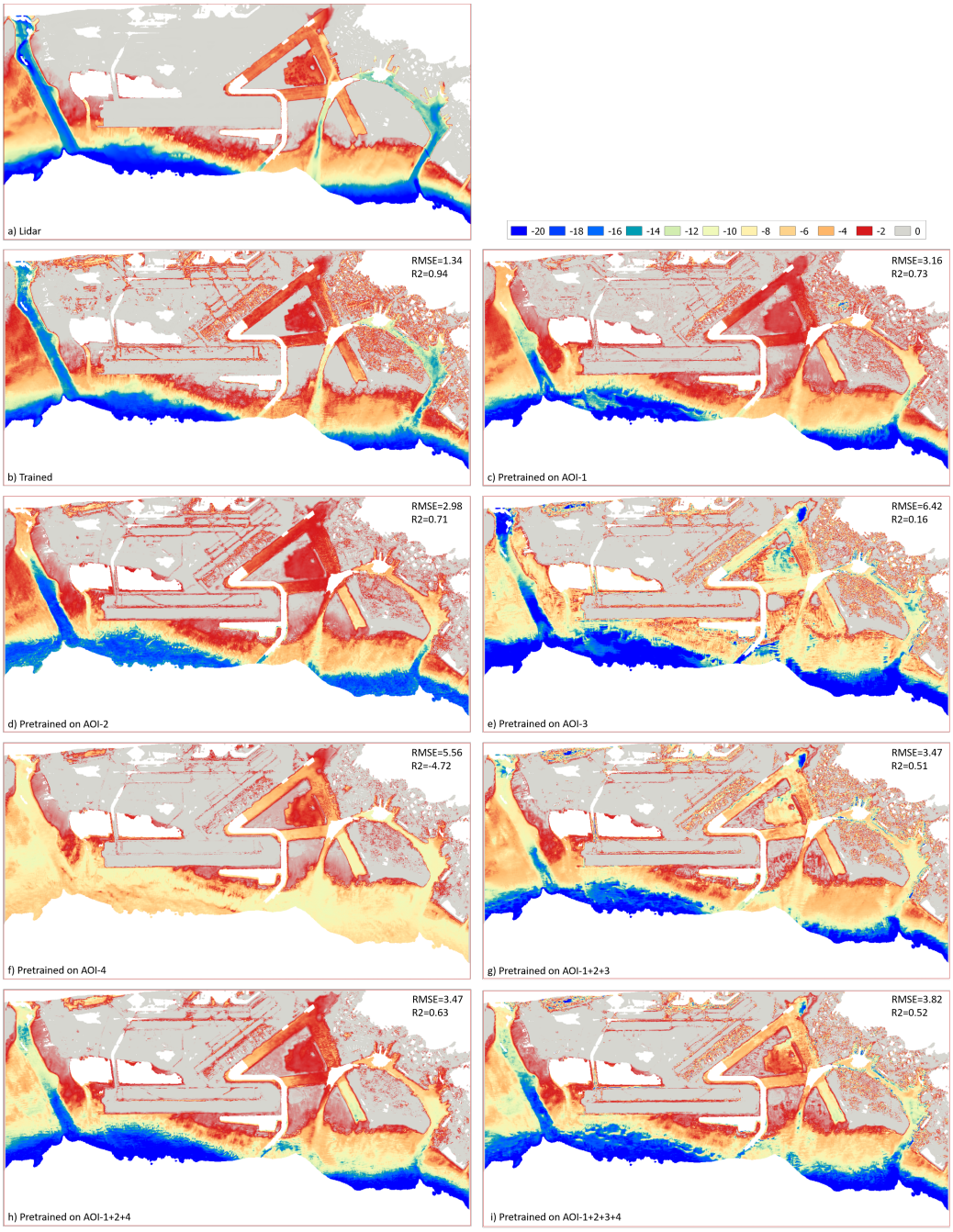


Figure 4.34: LiDAR bathymetric depths (a) and SDB results in AOI-6 using different CNN models: trained (b); pretrained on AOI-1 (c); pretrained on AOI-2 (d); pretrained on AOI-3 (e); pretrained on AOI-4 (f); pretrained on the combined AOI-1, AOI-2, and AOI-4 (g); pretrained on the combined AOI-1, AOI-2, and AOI-4 (h); pretrained on the combined AOI-1, AOI-2, AOI-3, and AOI-4 (i).

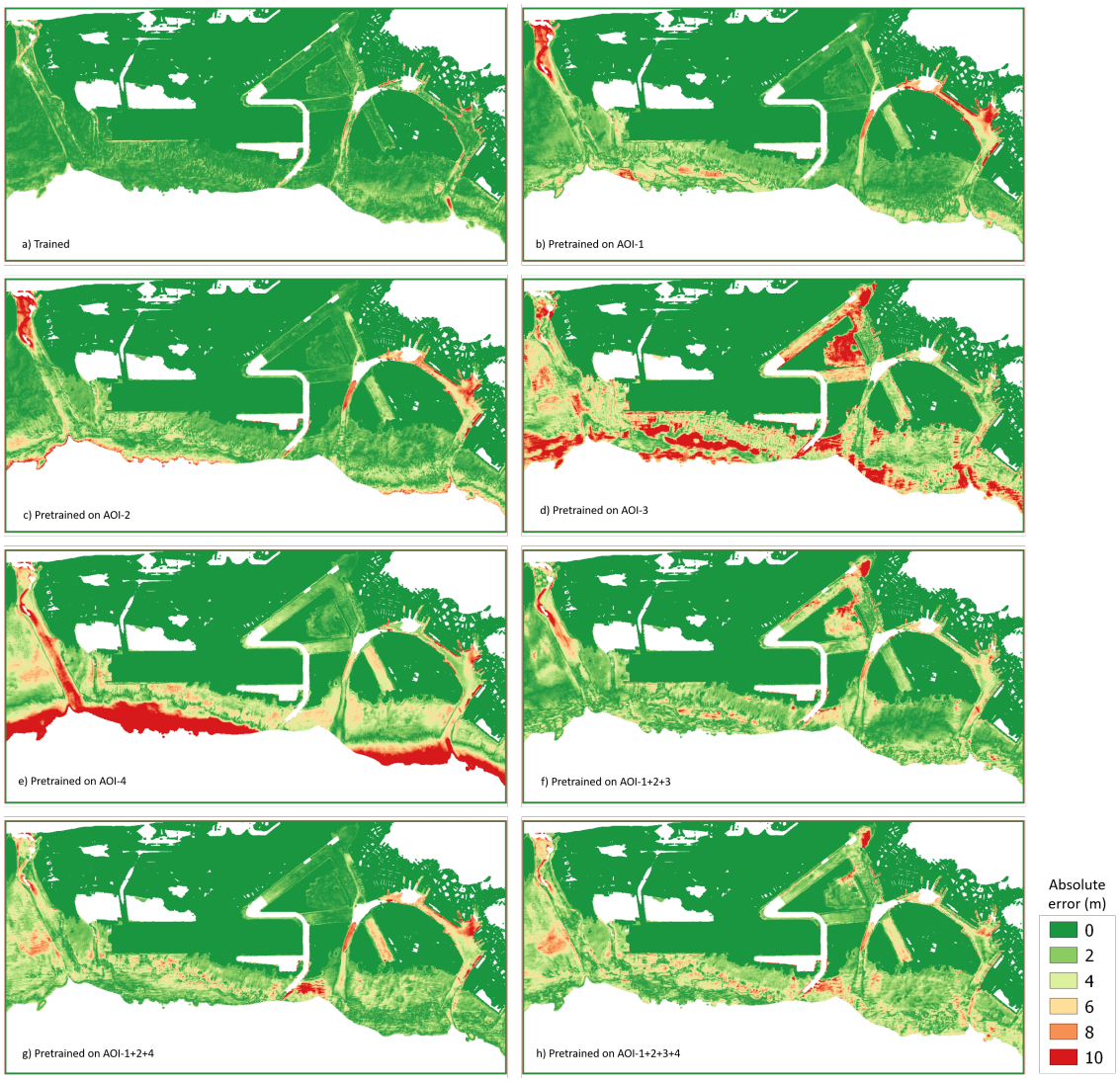


Figure 4.35: Absolute error of SDB per pixel when using different CNN models: trained (a), pretrained on AOI-1 (b); pretrained on AOI-2 (c); pretrained on AOI-3 (d); pretrained on AOI-4 (e); pretrained on the combined AOI-1, AOI-2, and AOI-3 (f); pretrained on the combined AOI-1, AOI-2, and AOI-4 (g); pretrained on the combined AOI-1, AOI-2, AOI-3, and AOI-4 (h).

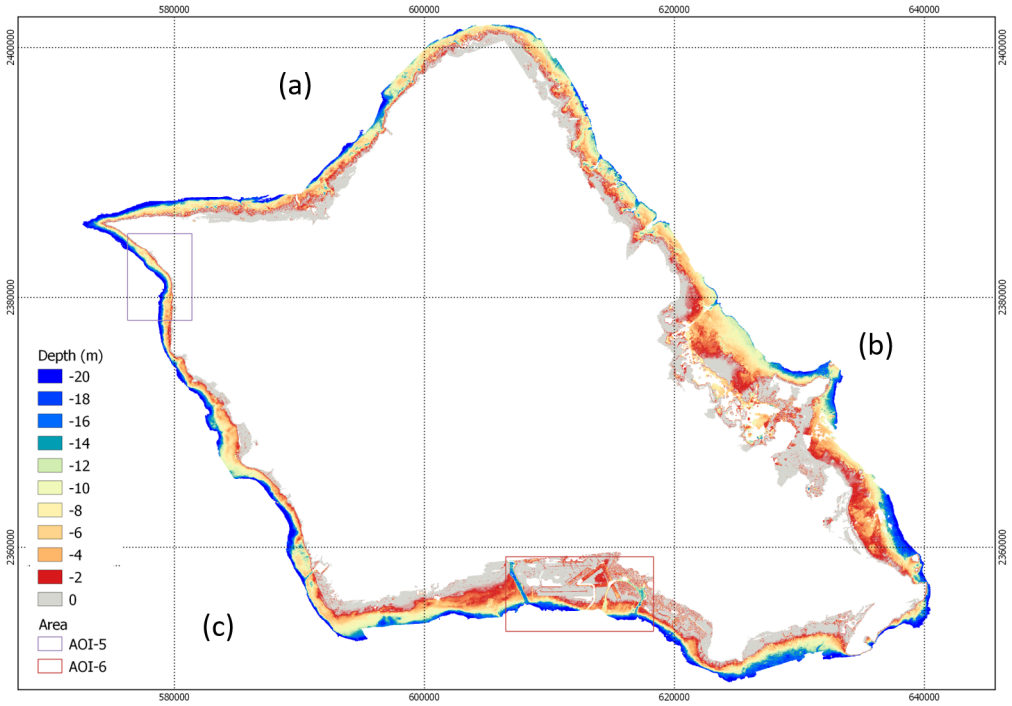


Figure 4.36: SDB models of Oahu island derived using a combination of pretrained model on AOI-5 and AOI-6.

varied than the samples, e.g. small patch reefs mix with mud. Unfortunately, the LiDAR data has gaps, so even though the SDB model covers all waters, the error calculation is limited to LiDAR coverage. Grid *a* and *b* are both predicted to be shallower than the LiDAR depths. Meanwhile, in the grid *c*, the computed SDB predicts deeper than the in-situ data.

The coastal length of Oahu island is about 365 km. The area is captured in three footprints covering the western, eastern, and northern parts. The images are first clipped to reduce the data load, and they are converted into sub-images. Since they have different extents, the size of each data is different, where the northern has the smallest size, followed by the western and the eastern. This project uses a graphics processor to compute the SDB. The total processing time using the GPU for each image is 20 minutes, 17 minutes, and 4 minutes. While it is possible to generate an SDB model in a larger area, such as in Oahu island, note that it requires compatible hardware to do the computation.

4.8 Extending the CNN architecture with regularization

This experiment was conducted as an additional experiment considering the overfitting problem in machine learning. Since we were not sure which regularizer to be used, this study tried three different regularizers: L1 (Lasso), L2 (Ridge), and the combined L1L2 (Elastic Net). The difference between them is in the regularization term as an addition to the loss function. L1 regularization uses the absolute value, while the L2 uses the squared value of the model's weight. Those values are then multiplied by a constant (*lambda*) to control the regularisation term's magnitude. Here we used the default value for *lambda*, which is 0.01. Meanwhile, the L1L2 regularization combines both functions in the regularization term.

The regularizer function was applied to the AOI-1 training data, the single image and multi-temporal images, by activating the *kernel_regularizer* in the Dense layer. Then, the trained model was implemented to other data to extract the SDB model in AOI-5. Figure 4.39 and Figure 4.40 show the performance of CNN during training when different regularization methods were applied. Although there is no significant

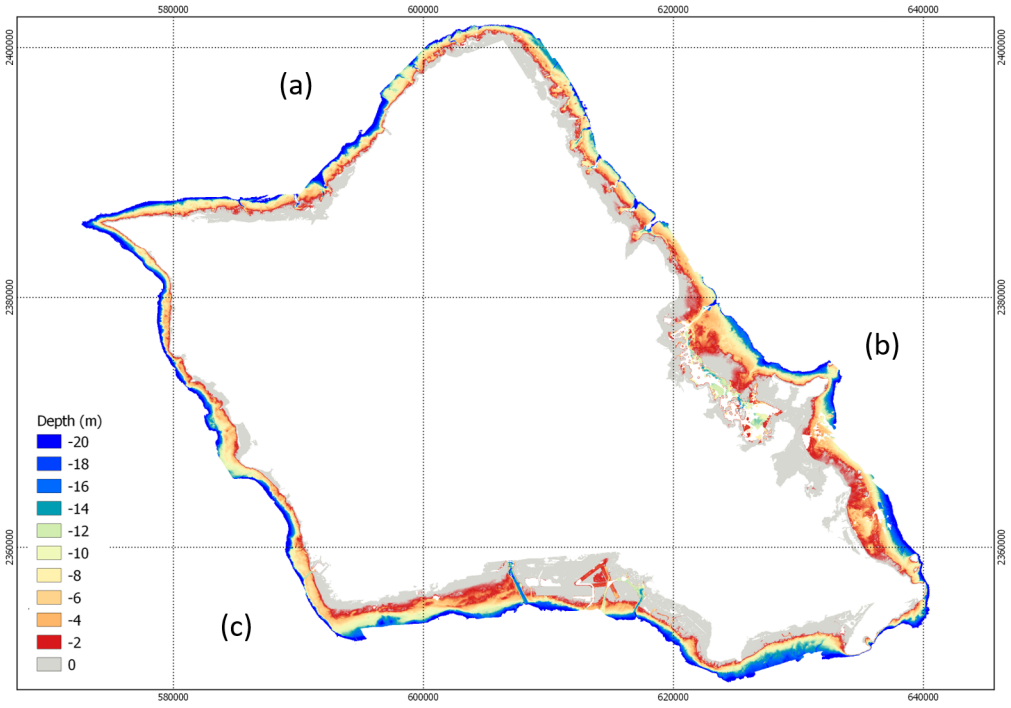


Figure 4.37: LiDAR bathymetry of Oahu island.

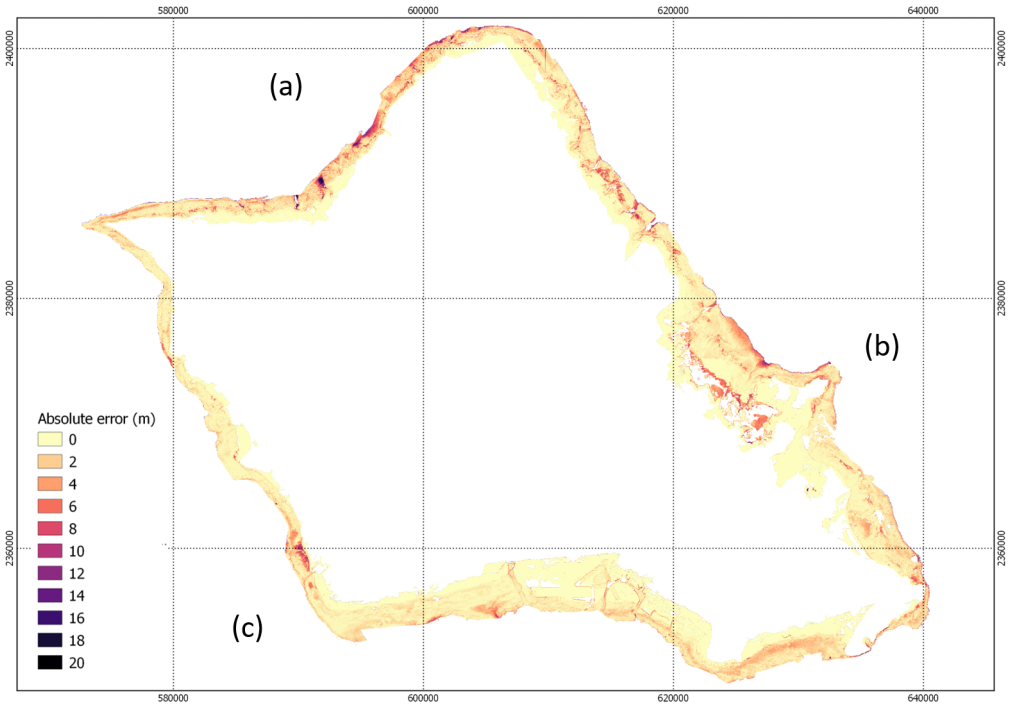


Figure 4.38: Absolute error of Oahu island.

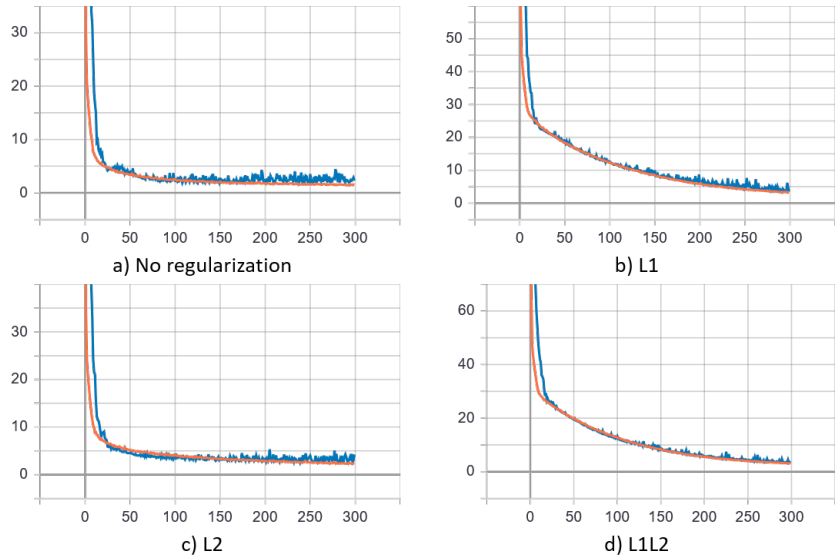


Figure 4.39: The performance on the train (orange) and validation (blue) data during training. No regularization (a), L1 regularization (b), L2 regularization (c), and L1L2 regularization (d) were applied to a single image of AOI-1. The x axis is the epoch, and the y axis is the loss value, which is MSE in meters.

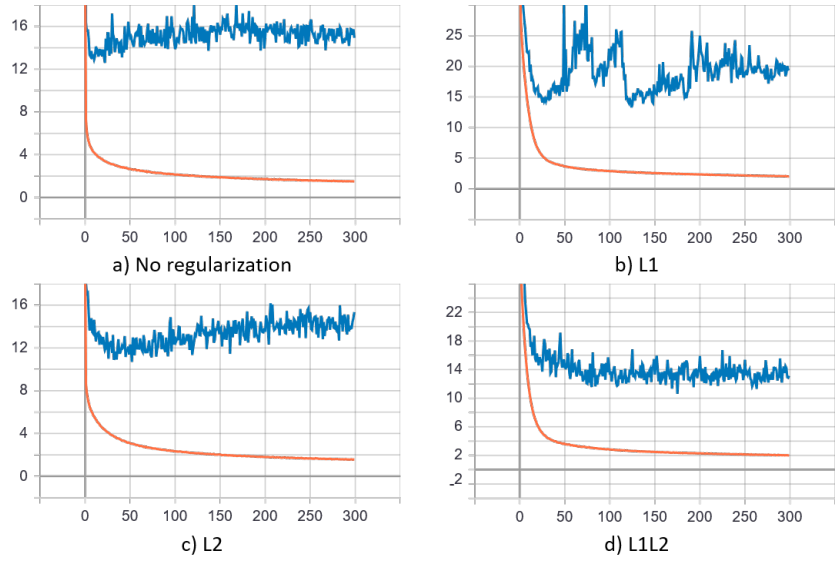


Figure 4.40: The performance on the train (orange) and validation (blue) data during training. No regularization (a), L1 regularization (b), L2 regularization (c), and L1L2 regularization (d) were applied to multi-temporal images of AOI-1. The x axis is the epoch, and the y axis is the loss value, which is MSE in meters.

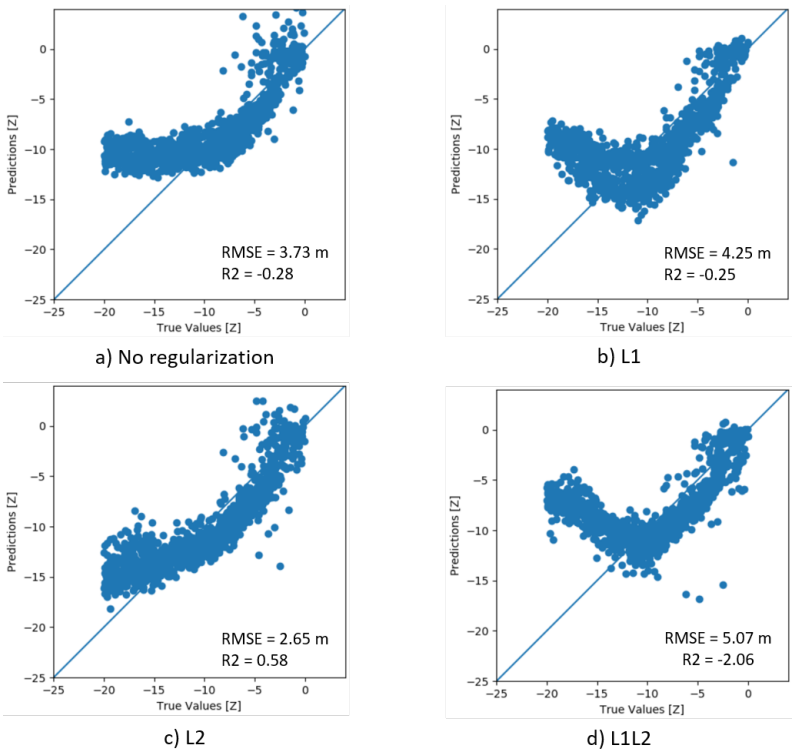


Figure 4.41: The accuracy of SDB in AOI-6 using the pretrained model from a single image of AOI-1 with no regularization (a), L1 regularization (b), L2 regularization (c), and L1L2 regularization (d).

change in the training performance when implementing the L2 regularization, the L1 regularization slows down the converging process of the model in the Figure 4.39b and destabilizing the performance during training in Figure 4.40b. It may occur since the L1 regularizer tries to optimize the model by shrinking some values, eliminating some important features. Another parameter that is needed to be considered is the *lambda*. If *lambda* is too small, then the regularization may not work. Meanwhile, if *lamda* is too large, it will add too much value to the model's weight. The default value used in this experiment is probably too small since the performance of training without regularization (Figure 4.39a, Figure 4.40a) and with the L2 regularization (Figure 4.39c, Figure 4.40c) are not showing any significant change.

Although their training performances are alike, Figure 4.41c clearly shows that the accuracy is improved by 1.08 m when L2 regularization was added to the architecture. Figure 4.42c also shares the same result where the error of SDB decrease by 0.22 m. Figure 4.41 and Figure 4.42 demonstrate that the L1 and L1L2 regularization decrease the accuracy in both cases, while the L2 regularization potential to improve the model by reducing the overfitting. The *lambda* parameter may need to be adjusted to produce a better result.

4.9 Training duration

This thesis utilized the CPU at the beginning of experiments and started to use the GPU processors when the number of training data increases significantly, especially when the training data from multi-temporal images in different locations were combined. As shown in Figure 4.43, execution of CNN training using the CPU processor requires more time than using the GPU. Based on one sample training using the CPU and the GPU, training time using the GPU is 25× faster than the CPU. If we fit both data using the linear trend, they have different slopes where training time using the CPU increases faster than when using the GPU.

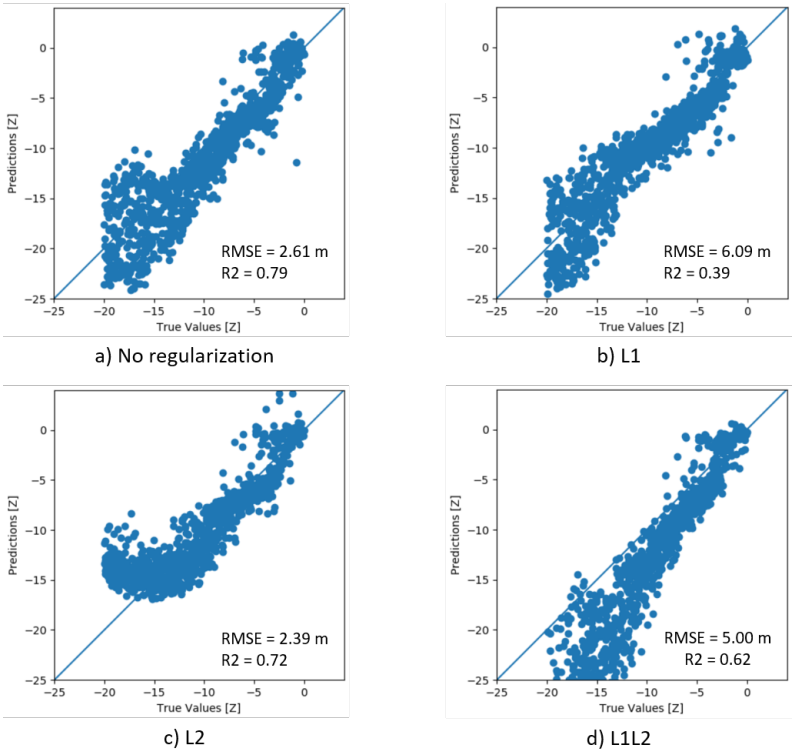


Figure 4.42: The accuracy of SDB in AOI-6 using the pretrained model from multi-temporal images of AOI-1 with no regularization (a), L1 regularization (b), L2 regularization (c), and L1L2 regularization (d).

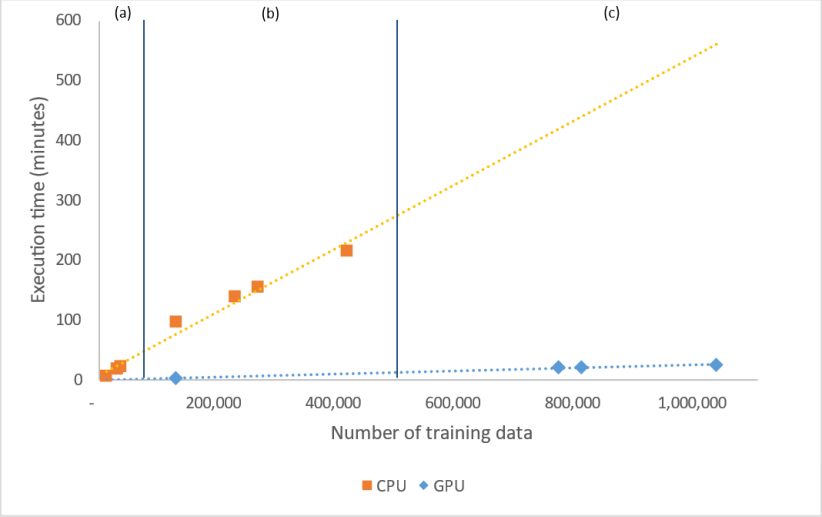


Figure 4.43: Execution time of CNN training with respect to the number of training data n in AOI-1 to AOI-4. The training data was extracted from uncorrected images with 9×9 window size and RGBNSS bands based on different cases: a single image (a), multi-temporal images (b), and a combination of multi-temporal images in several study areas. The training was performed using an HPC cluster with two different processors: CPU for the single and multi-temporal images (a and b); GPU for the combined training data (c) and one sample from multi-temporal data.

5 Conclusions and future work

This chapter covers three parts to narrow down all study cases performed in this thesis. Section 5.1 provides a recapitulation of the main results of each part in the previous chapter, together with a brief discussion. The conclusions, which answer the research questions, are given in Section 5.2. Finally, Section 5.3 provides a general overview of possible studies in the future that were highlighted by this thesis as promising areas of research.

5.1 Summary of main results

This thesis performs various case studies in order to study the potential of CNN to extract shallow water depths from multispectral images. The experiments include a comparison of several CNN architectures, studies in different locations with different water conditions, analysing whether the images need to be corrected before the CNN or not, using multi-temporal images for training, and performing a transfer model analysis. In order to recollect the condition of each study area, Table 5.1 provides an overview of the satellite image conditions, turbidity levels, as well as bottom types. Subsequently, Table 5.2 gives a summary of the SDB accuracy in all study areas and different case studies, followed by a brief discussion of each experiment along with their limitations.

Table 5.1: An overview of coastal water conditions in different study areas.

Area	Satellite image	Turbidity	Bottom type
AOI-1: Ponce	Clear	Low - medium	Sand, reef, seagrass, algae, coral rock
AOI-2: Lajas	Clear	Low - medium	Sand, reef, seagrass, algae, rock/boulder
AOI-3: San Juan	Clear, include boat/ship	High	Sand, mud, seagrass, coral reef, pollutants
AOI-4: Key West	Clear, include boat/ship	Low - high	Sand, mud, seagrass, algae, live coral, coral reef, reef rubble, pavement
AOI-5: Makua	Clear	Low	Sand, coral reef, rock/boulder
AOI-6: Honolulu	Clear	Low - medium	Sand, mud, algae, reef, pavement

The SDB generated using the CNN approach with less training data, compared to the linear-transform method, in AOI-1 obtains a higher accuracy by 67%. It indicates the CNN capability of addressing non-linear trend in the reflectance-depth pairs. Regarding the comparison between different CNN architectures, this thesis points out that the architecture with six channels (RGBNSS) and 9×9 window size outperforms other architectures with less or more channels and a smaller window size. The lowest RMSE is 1.31 m when training in AOI-1, while the highest is 1.94 m. The window size specifies how many neighbours are considered in the training of CNN. CNN training detects more information, which essential for prediction, from the local neighbourhood when more neighbours are connected. Likewise, with the number of bands used, the NIR and SWIR channels add important information in training. NIR and SWIR spectrum have a characteristic of being quickly absorbed by water but still reflective enough in shallow waters, so the reflectance variations benefit SDB computation. In this research, the window size is limited to 9×9 considering the computational load in the training process. More training data with more neighbours require more memory for the training computation. In this case, AOI-2 consumes more training time than other areas. Using different window sizes, the training duration in AOI-2 is around 55 minutes and 22 minutes for the 9×9 and 7×7 windows, respectively. Therefore, the implementation of

Table 5.2: A recapitulation of SDB accuracy using the RGBNSS combination in different study cases performed in this thesis. *n* is the number of training data.

Case		Area	n	RMSE	R^2
1. Single image	1.1. Corrected using DWA	AOI-1	10,626	1.31	0.94
		AOI-2	34,328	1.78	0.92
		AOI-3	29,275	3.12	0.46
		AOI-4	28,165	0.88	0.93
	1.2. Uncorrected image	AOI-1	10,626	1.22	0.96
		AOI-2	34,328	1.63	0.95
		AOI-3	29,275	3.03	0.55
		AOI-4	28,165	0.73	0.94
		AOI-5	5,851	0.79	0.98
		AOI-6	21,712	1.34	0.94
2. Multi-temporal images (uncorrected)	2.1. Accuracy of the same test data as case 1	AOI-1	127,503	1.11	0.96
		AOI-2	413,291	1.44	0.95
		AOI-3	263,890	2.36	0.70
		AOI-4	225,531	0.56	0.97
	2.2. Optimum accuracy	AOI-1	-	0.96	0.97
		AOI-2	-	0.93	0.98
		AOI-3	-	2.03	0.77
		AOI-4	-	0.54	0.97
3. Transfer model	3.1. Pretrained on AOI-1	AOI-5	-	2.61	0.79
		AOI-6	-	3.16	0.73
	3.2. Pretrained on AOI-2	AOI-5	-	2.92	0.69
		AOI-6	-	2.98	0.71
	3.3. Pretrained on AOI-3	AOI-5	-	7.42	0.13
		AOI-6	-	6.42	0.16
	3.4. Pretrained on AOI-4	AOI-5	-	6.84	-7.56
		AOI-6		5.56	-4.72
	3.5. Pretrained on the combined AOI-1, AOI-2, and AOI-3	AOI-5	804,684	2.69	0.73
		AOI-6		3.47	0.51
	3.6. Pretrained on the combined AOI-1, AOI-2, and AOI-4	AOI-5	766,325	4.70	0.34
		AOI-6		3.47	0.63
	3.7. Pretrained on the combined AOI-1, AOI-2, AOI-3, and AOI-4	AOI-5	1,030,215	3.43	0.52
		AOI-6		3.82	0.52

larger window sizes should consider the hardware capacity. It may be advantageous to use more neighbours to identify how many neighbours are effective to be considered, especially when a high resolution multispectral image is available.

Furthermore, using corrected images in different locations, the SDB accuracy ranges from 0.88 m to 3.12 m. The lowest error represents a shallower depth range, 0-10 m, than other study areas. Meanwhile, the highest error occurs in AOI-3, which represents more turbid waters than other locations. Although the DWA method for bottom extraction can be simply carried out, its implementation has drawbacks when the reflectance is disturbed by turbidity, e.g. in AOI-3, or when the absorption rate of Red/NIR/SWIR bands reaches 100%, e.g. in AOI-4, where the corrected images possibly contain gaps or negative values which affecting the SDB computation. It turns out that the DWA method is not efficient to extract bottom reflectance considering different water column concentrations within the study area are not supposed to be represented by a small area in deep water. Another experiment using uncorrected images depicts that SDB can be generated even with a better accuracy without bottom extraction. So, the convolution process can rectify the water column effect that occurs in the reflectance value.

Afterwards, a better SDB model is obtained using multi-temporal images. The accuracy improves up to 0.67 m, where the highest upgrade occurs in AOI-3. The reflectance values of water regions relate to the

water column and the bottom type. They are correlated to the seasonal change in the local area. The addition of multi-temporal images enhances reflectance-depth data variations to increase the possibility of generalizing the model. Based on the results, a trained model using multi-temporal data generally delivers an SDB model with similar accuracy in each month. However, the trained model still does not represent the overall data since inaccurate results still arise in some images in which the reflectance values are outside the typical range. For example, the November/December case in Puerto Rico regions and October/December case in AOI-4, where the accuracy is significantly lower than other months. In those cases, adding the training data using oversampling method helps to balance the variations. As for the results, the inaccurate results become more accurate, following the RMSE range of other results within the same area.

Next, using the pretrained model of different locations, SDB models of the new locations yield various accuracies. Although some pretrained models can predict the new study areas quite well, some others are not suitable to be used at all. It occurs because the pretrained model still does not represent the entire variation in the new data, or as the result of different corrections applied to different images. Moreover, when a pretrained model of different water characteristic is used, the prediction tends not to fit the data within the trained model and thus results in a different SDB model. This phenomenon occurs when the pretrained model of AOI-4 is used in the other data. On the other hand, a pretrained model from southern Puerto Rico regions can fit the other data up to 15 m depth. Deeper than that, the SDB model becomes inaccurate. It indicates that the sample data from Puerto Rico is not representative of the new data, especially at the 15-20 m depth range.

In order to build a more general CNN model, several sample data sets in different locations are combined and trained. By doing this, the sample data includes many reflectance variations due to the bottom types, turbidity, water column effects, seasonal changes, and locations. The CNN was expected to learn these variations to derive a general model so that the trained model can generate the SDB model with fair accuracy in different regions. However, the results cannot achieve better accuracy than when a pretrained model from a single location is used, which is not as expected. Likely, the sample data from turbid waters should not be mixed with others since the output gets a lower accuracy. In addition, the sample data from AOI-4 only includes 0-10 m depth variations, while a poorly result occurs in deeper depths. So, the addition of AOI-4 sample data cannot resolve the problem because it does not cover the overall depth range. This result describes a limitation for the CNN method to extract shallow water depths. When combining the data, the proportion of each input characteristic should be taken into account, while it is not the case in this project. All sample data from several locations are currently merged directly without considering the distribution of bottom types and turbidity in detail.

In general, water column effects and bottom type variations affect model training. As mentioned several times, turbidity causes disturbances in the reflectance value; this also affects CNN training, where the sample data is more challenging to learn when there is more noise present in the image. It is possible to classify turbidity level using the CNN method. The sample data can be acquired manually based on the water colour. The turbidity map derived from the CNN method denotes that turbidity in nearshore waters is generally higher than an area away from the shoreline. In addition, turbidity is also higher in certain areas such as dredged zones, estuaries, and polluted areas. As a consequence, the SDB result becomes inaccurate as well in these environments. In this thesis, the sample data of turbidity is collected from the AOI-3 area. So, the sample data only covers turbidity due to sediments or pollutants, excluding the influence of algae/phytoplankton concentration.

In addition to turbidity, bottom type variations also cause variations in the reflectance value. That is why it is essential to have a complete and balanced set of sample data representing different types of water bottoms. This project currently uses a random sampling method to select the sub-images to generate the training data. Although the results of SDB within the training data coverage are reasonably accurate, the transfer model's accuracy still needs to be improved. It could be argued that the reflectances complexity may not be represented entirely in a balanced manner when combining the sample data from different locations. A thorough identification of training data distribution to balance the variation must be analyzed more comprehensively. It could be done by first classifying the pixels based on their bottom types and then using that information to collect the same number of samples in different bottom types and different depths. Besides that, the accuracy is also affected by the depth itself, where the relation-

ship between reflectance and depth generally become less clear as the depth increase due to the light penetration characteristic in water.

5.2 Conclusions

This thesis addressed the main question: *to what extent can convolutional neural networks be used for accurate shallow water depth extraction using Sentinel-2 satellite images?*. In order to answer this question, this study specified four sub-questions which will be answered based on the results and analyses that have been discussed previously. The following paragraphs provide brief answers to each sub-question.

- **Q: What kind of preprocessing is needed for the data sets?**

  A: Surface reflectance values are sufficient to be used as input data for CNN training, so no additional preprocessing is needed.

  In this thesis, bottom reflectance values were extracted using some deep water pixels. This method has a limitation when encountering reflectance pixels with high absorption rates or too much noise. Those pixels cause negative values or *nodata* that become gaps in the output. *Nodata* pixels are undesirable since one of the primary purposes of SDB is to fill the gaps that are occurred from in-situ measurements using SBES, MBES, or even LiDAR. Besides, even though the negative values did not affect the SDB result locally, they are not suitable for input when using the pretrained model from the outside since the corrected values are out of range. Instead, the surface reflectance values sufficient to produce the SDB model with the same or even better accuracy. Based on the results, the improvement ranges from 0.09 m to 1.04 m from the prior RMSE. The lowest improvement occurred in AOI-3, indicating that the training data using corrected or uncorrected approaches are still noisy due to turbid waters. The highest improvement found in AOI-2 when using the RGB combination indicates that the corrected image may alter the information contained in the original values, resulting in a more inaccurate output. Additionally, it is more efficient to use DN values in training duration as the training time can be up to 50% faster using the same architecture. Therefore, it can be concluded that SDB can be produced using the CNN approach without additional image preprocessing and features specification to define the predictors.

- **Q: What kind of CNN architecture can be used for SDB?**

  A: These architectures are considered suitable for the SDB task: triple convolutional layers with 3x3 kernel size and batch normalization, a window size of 9×9*, RGBNSS channel, ReLU activation at the convolution layers, linear activation at the output layer, and dropout layer. The hyperparameters baseline can be started with 512 batch size, a learning rate of 0.0001, and 300 epochs.*

  These architectures were selected considering some existing CNN architectures, mainly for the classification task, and several trials. Based on several preliminary experiments, specific hyperparameters were defined to be used in this thesis. They include learning rate, batch size, and epoch. These hyperparameters became the default when building a baseline architecture for SDB. Several options in the architectures were examined, which are the number of convolutional layers, the number of channels, and window sizes. The use of multiple convolutional layers yielded similar RMSE, but the validation accuracy is more stable when using three rather than two network layers. The use of larger window sizes improved the accuracy, especially at depths deeper than 10 m. The regular bands that are usually used in SDB are red, green, and blue. Adding NIR and SWIR bands improve the accuracy up to 0.99 m. Additionally, the L2 regularizer can be included in the architecture to reduce the overfitting and achieve better generalization.

- **Q: What is the accuracy of the method (compared to the linear transform method, in different depth ranges, in different locations)?**

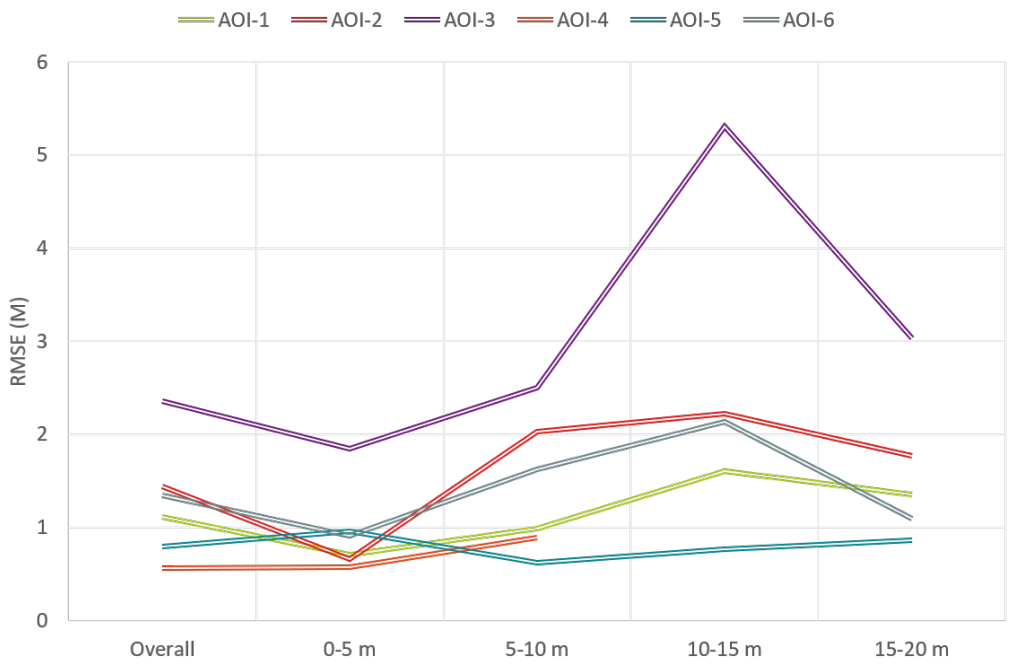


Figure 5.1: The RMSE values of SDB results in different locations and in different depth ranges.

A: Shallow water depth extraction using CNN has better accuracy than the linear transform method. Additionally, CNN also produces more accurate outputs than previous studies within the same areas (Table 5.3). In different depth ranges, the accuracy tends to decrease as the depth increase (Figure 5.1). Different locations with different water characteristics provide a diverse accuracy (Figure 5.1), where the accuracy is lower in turbid waters due to sediments and higher in clear waters.

Table 5.3: SDB accuracy comparison between different methods in different locations. AOI-1, AOI-5, and AOI-6 were assessed based on the RMSE values, meanwhile AOI-4 was evaluated using the Median Absolute Error. All values are in meters.

AOI	CNN	Other methods	
AOI-1	1.11	4.05	Linear transform
AOI-4	0.23	0.39	Ratio transform
AOI-5	0.79	1.24	Random Forest
AOI-6	1.34	3.01	Radiative transfer

The accuracy of the SDB model generated using CNN with less training data is 67% better than using the linear transform. In addition, several study areas have been studied using other methods, such as the ratio transform (AOI-4), Random Forest (AOI-5), and the simplified radiative transfer model (AOI-6). The accuracy of SDB in these areas using the CNN approach is higher than other methods. The increases are by 0.16 m (MedAE), 0.45 (RMSE), and 1.67 (RMSE) in AOI-4, AOI-5, and AOI-6, respectively. Besides, CNN was able to produce better accuracy with less amount training data.

In different locations, the optimum RMSE ranges between 0.54 m and 2.03 m. The highest accuracy occurred in AOI-3, where the backscattering dominates the reflectances from particles in the water column that cause turbidity. Meanwhile, the lowest RMSE is obtained in AOI-4, where the waters are relatively clear, so the surface reflectance contains much backscattering from the bottom. Within

the same depth range, SDB in AOI-1 is also produced similar accuracy. These optimum RMSE values were acquired when using the multi-temporal images. This thesis implemented several experiments consecutively, starting from the architectures comparison to the multi-temporal cases. In general, the RMSE of each study area has continually improved in each experiment. The improvement rates are different for each location. Using the same test data, the accuracy improved by 0.37 m (from 1.48 m to 1.11 m) in AOI-1, 0.34 m (from 1.78 m to 1.44 m) in AOI-2, 0.76 m (from 3.12 m to 2.36 m) in AOI-3, and 0.32 m (from 0.88 m to 0.56 m) in AOI-4.

In different depth ranges, the error of deeper areas relatively higher than shallower depths. The relationship between reflectance and depth becomes less clear as the depth increase because less energy is backscattered from the bottom due to the light absorption and scattering effect in the water column. Based on the RMSE in different depths per 5 m increment, all study areas showed a similar trend of the highest RMSE at 10-15 m depths. It indicates that the most variation in the reflectances, due to turbidity, phytoplankton density, or different bottom types, is likely occurred at 10-15 m depths.

- **Q: To what extent can the pretrained model be reused?**

  A: The SDB results using different pretrained models yielded various accuracies. Some pretrained models are suitable to be reused for particular depths. Although the pretrained model can be used for the new data, it is challenging to decide which pretrained model should be used. Thus, it is advisable to verify the results.

  Implementing the pretrained models of different study areas, or their combination, to the new areas yields a wide range of accuracy. Some pretrained models are more compatible to be reused than others, but the output is suitable only at particular depth ranges. Meanwhile, other pretrained models are not suitable to be reused due to a lack of data at particular depths, or the reflectances are too noisy. The pretrained model of turbid waters at the harbour area tends to be local, meaning that the pretrained model is unsuitable for reuse even if the new area is also a harbour because the turbidity rate may differ, so the level of noise in the reflectance is not the same. The combination of several training data from different study areas also yields a lower accuracy. It can be argued that the combination did not represent balanced training data, resulting in an unreliable trained model. Therefore, supplementary information concerning the bottom types and water column properties is helpful to preserve the variety of data. The quantitative data of water properties is relatively hard to acquire, so that the absorption rates may be strived qualitatively based on the colour composite image. Since the accuracies diverge, it is hard to predict which pretrained model compatible with which data. Therefore, it is recommended to verify the SDB model using measured depths considering the risk when computing the SDB using a pretrained model.

5.3 Future work

Several experiments have been performed in this thesis. The limitations of the method have been discussed. From that, several ideas that are beneficial for the SDB knowledge are recommended for future work. They include the procedure of balancing training data, alternative in-situ depths data source, and examination in a different water temperature or salinity.

Balancing the training data. Noticeably, the transfer model as discussed in Section 4.6 encounters difficulties to fit the pretrained model to a new data. Although some of them are more suitable than others, some others are entirely incompatible with being reused. In order to improve the CNN pretrained model, a thorough identification of the water bottom and column properties is needed to balance the data. The proportion of sample data in different bottom types, different water column properties, and different depths is expected to be equal and cover all variations. Several additional data sets, such as benthic habitat classification map and water properties measurements, will help understand the study area's characteristic. They can be rasterized and become the basis to generate a uniform sample. Additionally,

CNN can probably be used to perform bottom type classification when the data is not available or is incomplete by using training data from the available sites.

Verification using another data source. As mentioned before that, it is hard to choose which pretrained model is compatible with the new data. Consequently, the computed SDB still needs to be verified using measured depths considering the risk of computing the SDB using a pretrained model. However, it becomes another disadvantage when no measured depths are available since most bathymetric measurement data is not open. As an alternative, Ice, Cloud and land Elevation Sattelite (ICESat-2) launched by NASA can be used to substitute in-situ data from the traditional survey. ICESat-2 is designed to measure the elevation of the ice sheet and land, but it also can measure water depth in clear coastal areas. The bathymetric depths can be derived after several corrections, including datum conversion, noise filtering, segmentation of surface and bottom signals, and refraction correction. The output can be used to verify the computed SDB. Moreover, if the accuracy of ICESat-2 depths is accurate enough compared to the traditional survey, it is also possible to use ICESat-2 depths as in-situ data for training.

SDB in the colder water and freshwater. The study areas in this thesis are primarily located in tropical latitudes with typical coastal waters, considering that different bottom types and turbidity represent different locations. It is recommended to explore other areas outside tropical regions, such as in the northern latitude where the water has a lower temperature, or in the river/lake/freshwater where the salinity is lower than seawater. Temperature and salinity affect water column properties and may also affect the backscattering level from the water bottom. Therefore, the SDB calculated in those areas may have a different result, and the pretrained model from the coastal water may not fit. However, this assumption has not yet been studied in this thesis.

A Additional figures

A.1 Training and testing depth data distribution

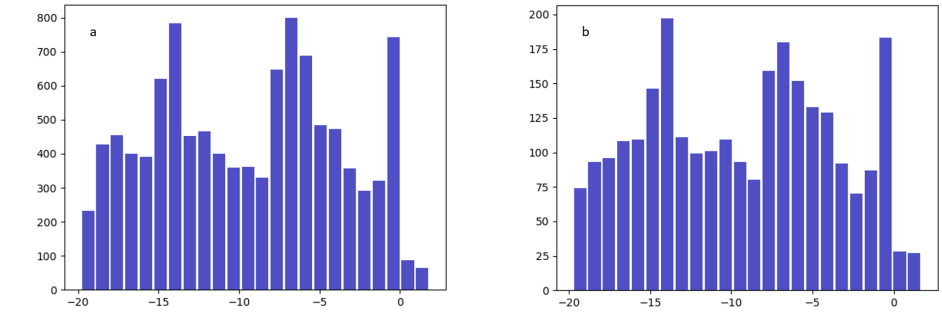


Figure A.1: Depth data distribution for training (a) and testing (b) in AOI-1.

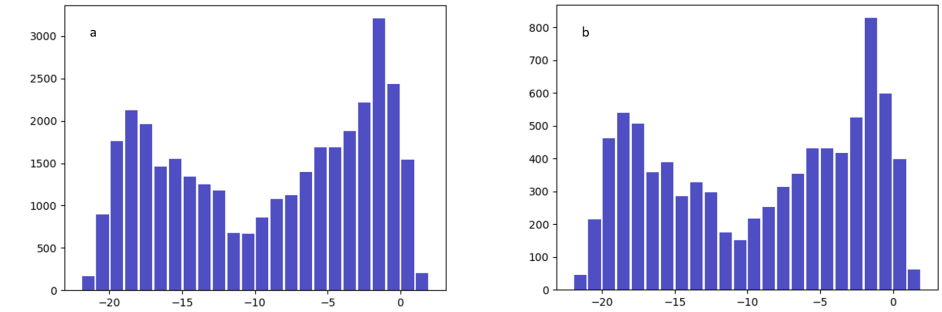


Figure A.2: Depth data distribution for training (a) and testing (b) in AOI-2.

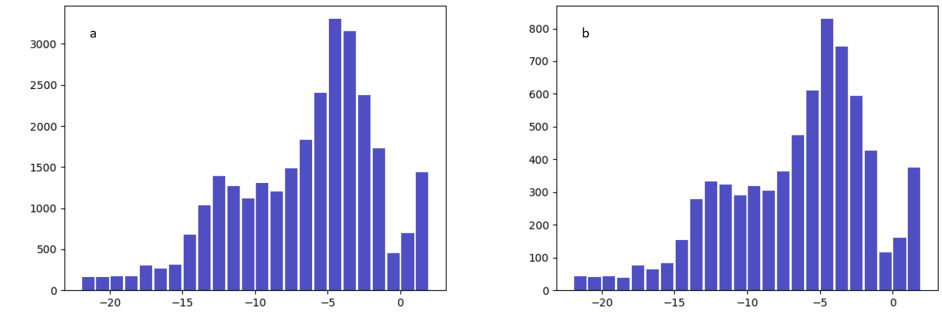


Figure A.3: Depth data distribution for training (a) and testing (b) in AOI-3.

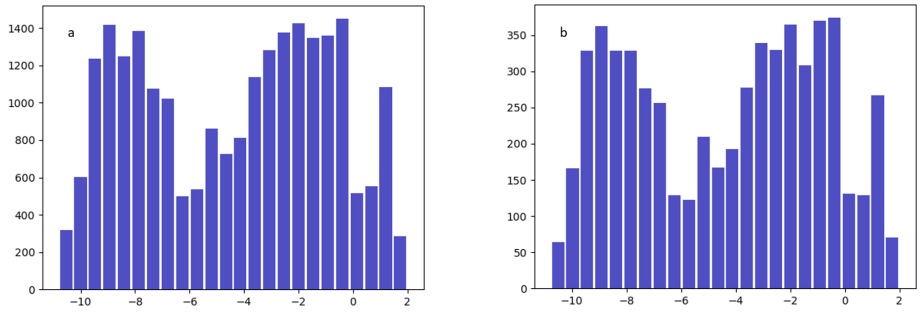


Figure A.4: Depth data distribution for training (a) and testing (b) in AOI-4.

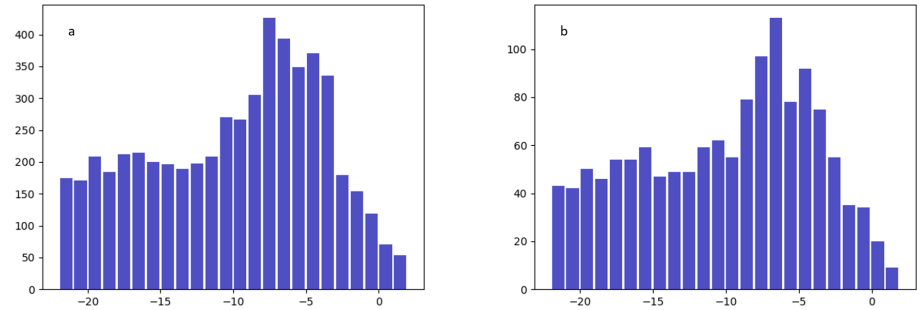


Figure A.5: Depth data distribution for training (a) and testing (b) in AOI-5.

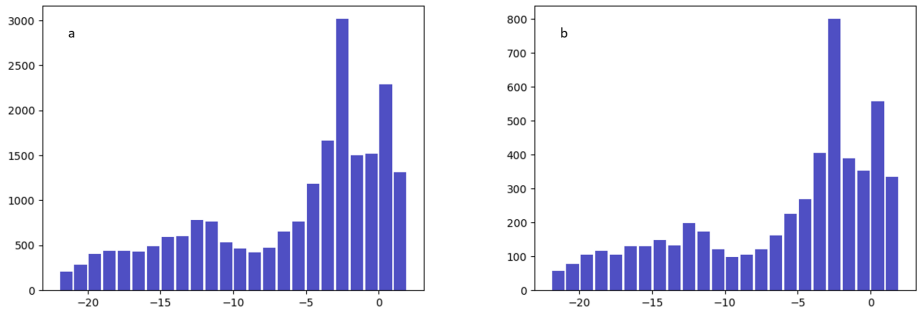


Figure A.6: Depth data distribution for training (a) and testing (b) in AOI-6.

A.2 Multi-temporal images

A.3 Spectral signatures

A.4 Tensorboard

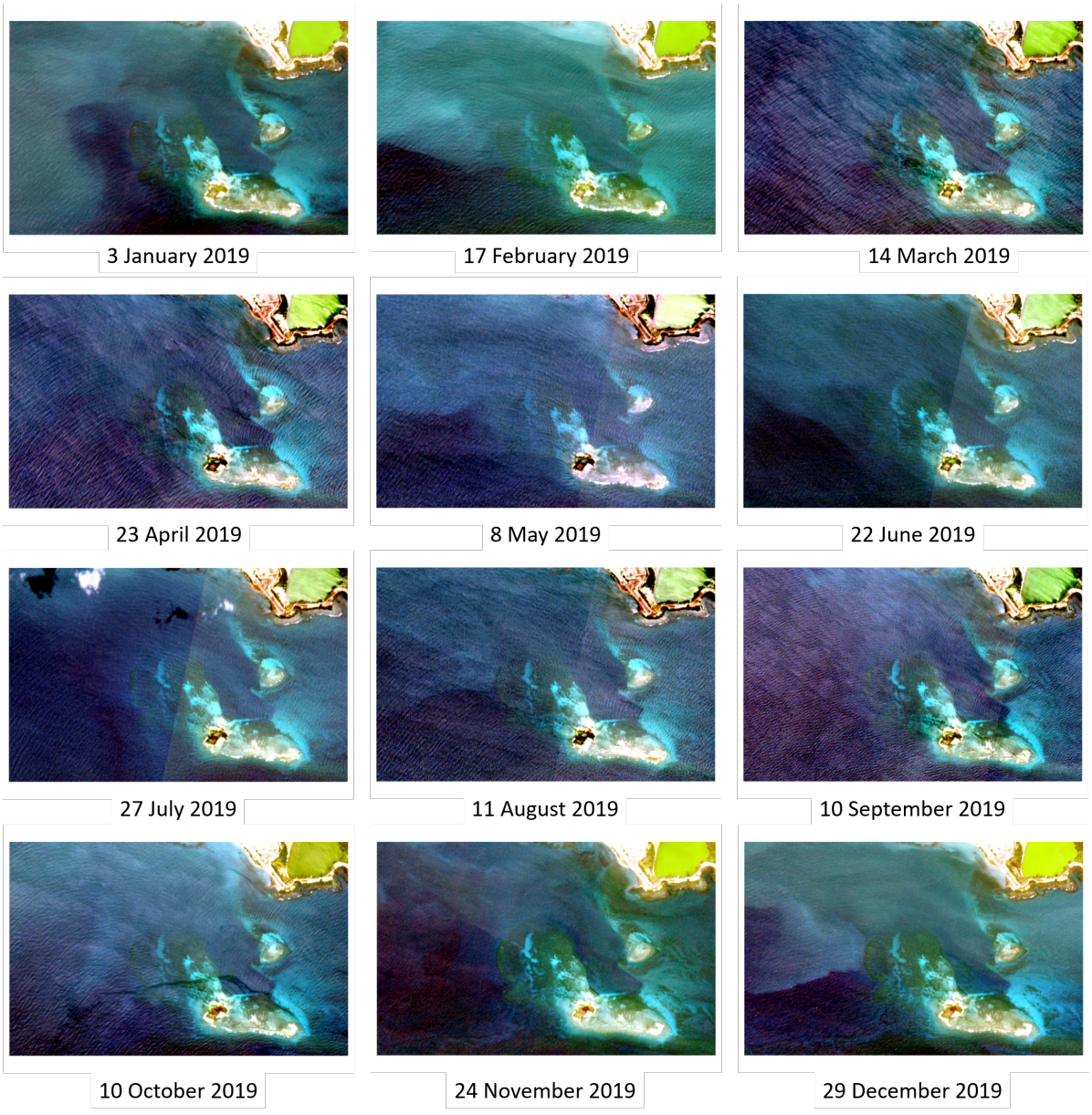


Figure A.7: RGB composite images of Sentinel-2 Level-2A in AOI-1 from January to December 2019.

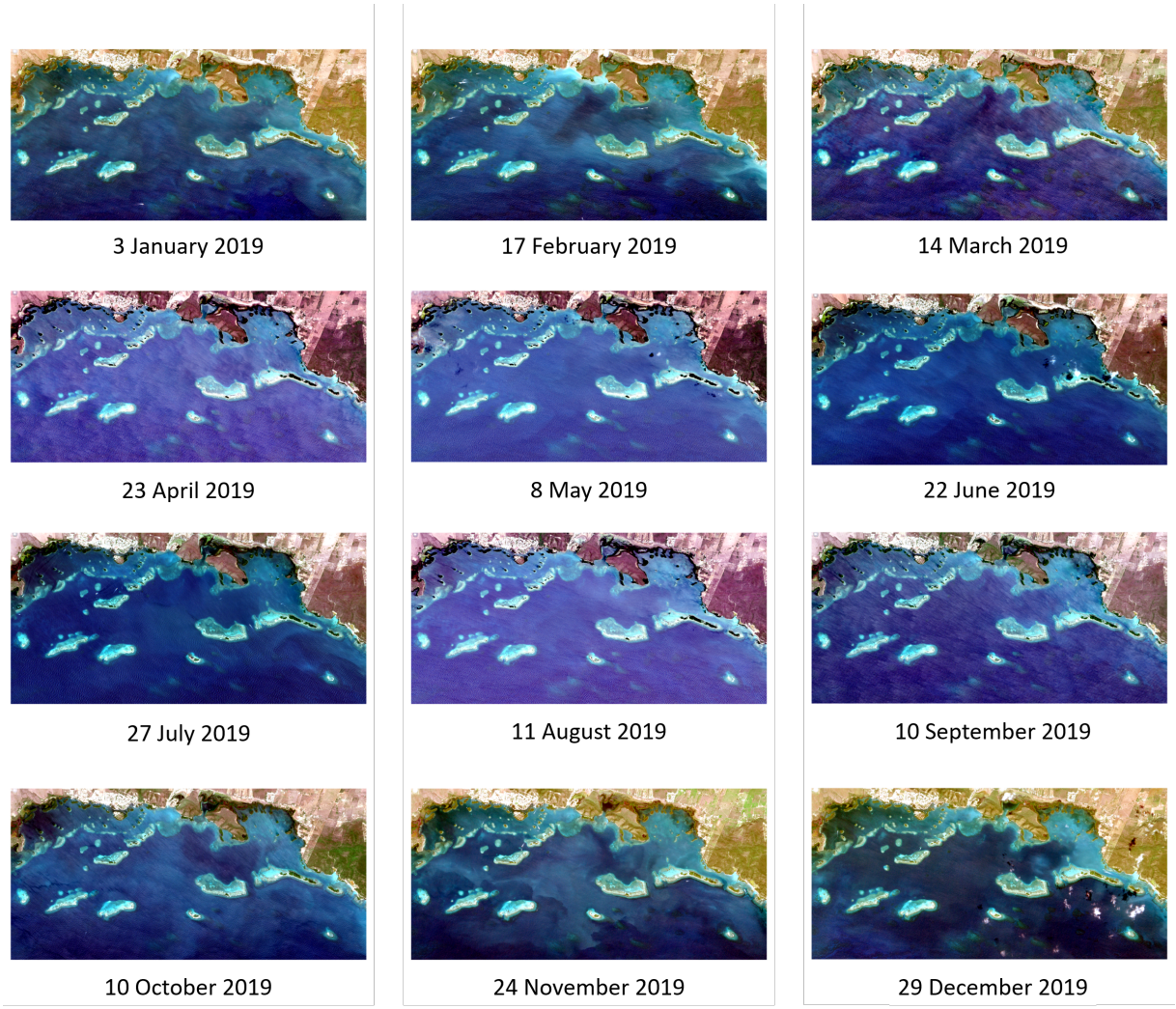


Figure A.8: RGB composite images of Sentinel-2 Level-2A in AOI-2 from January to December 2019.

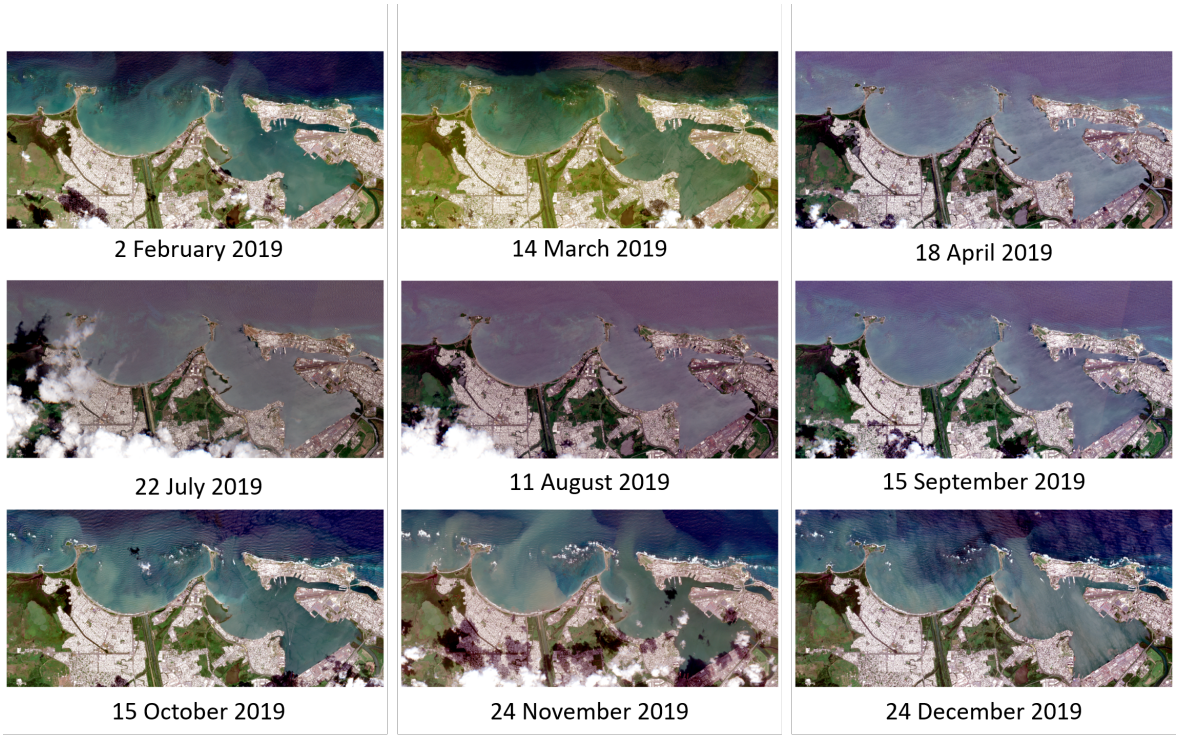


Figure A.9: RGB composite images of Sentinel-2 Level-2A in AOI-3 from January to December 2019.

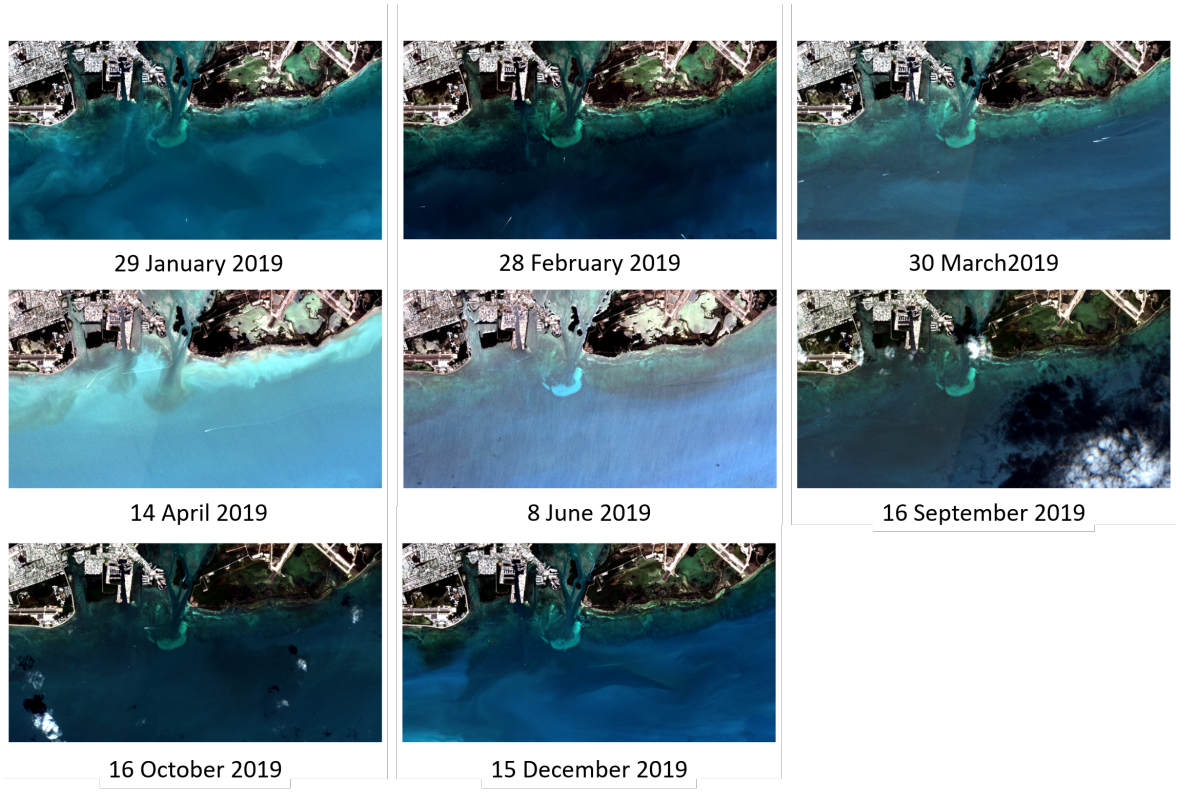


Figure A.10: RGB composite images of Sentinel-2 Level-2A in AOI-4 from January to December 2019.

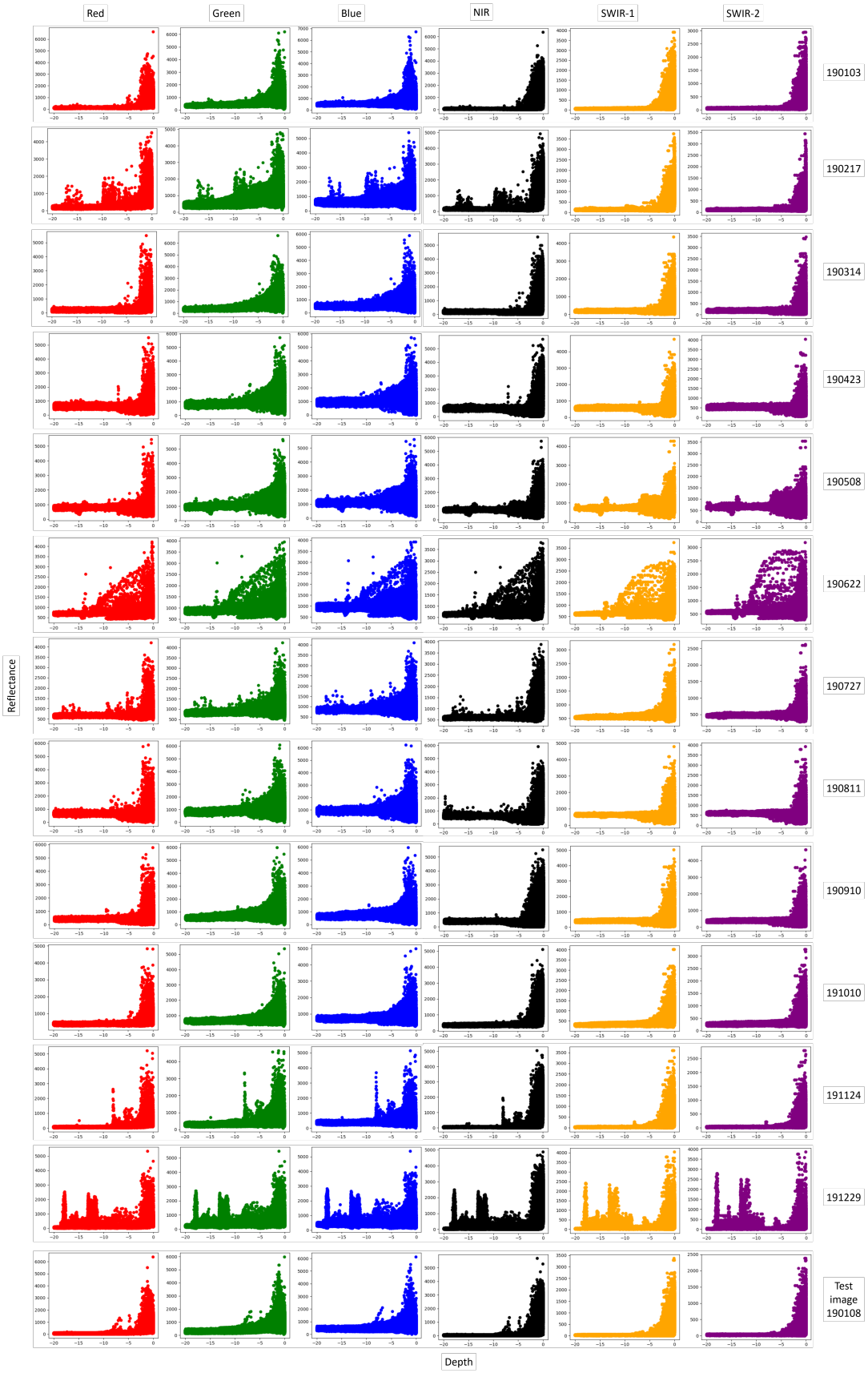


Figure A.11: Spectral reflectance signature towards depth of multi-temporal images in 2019 at AOI-2.

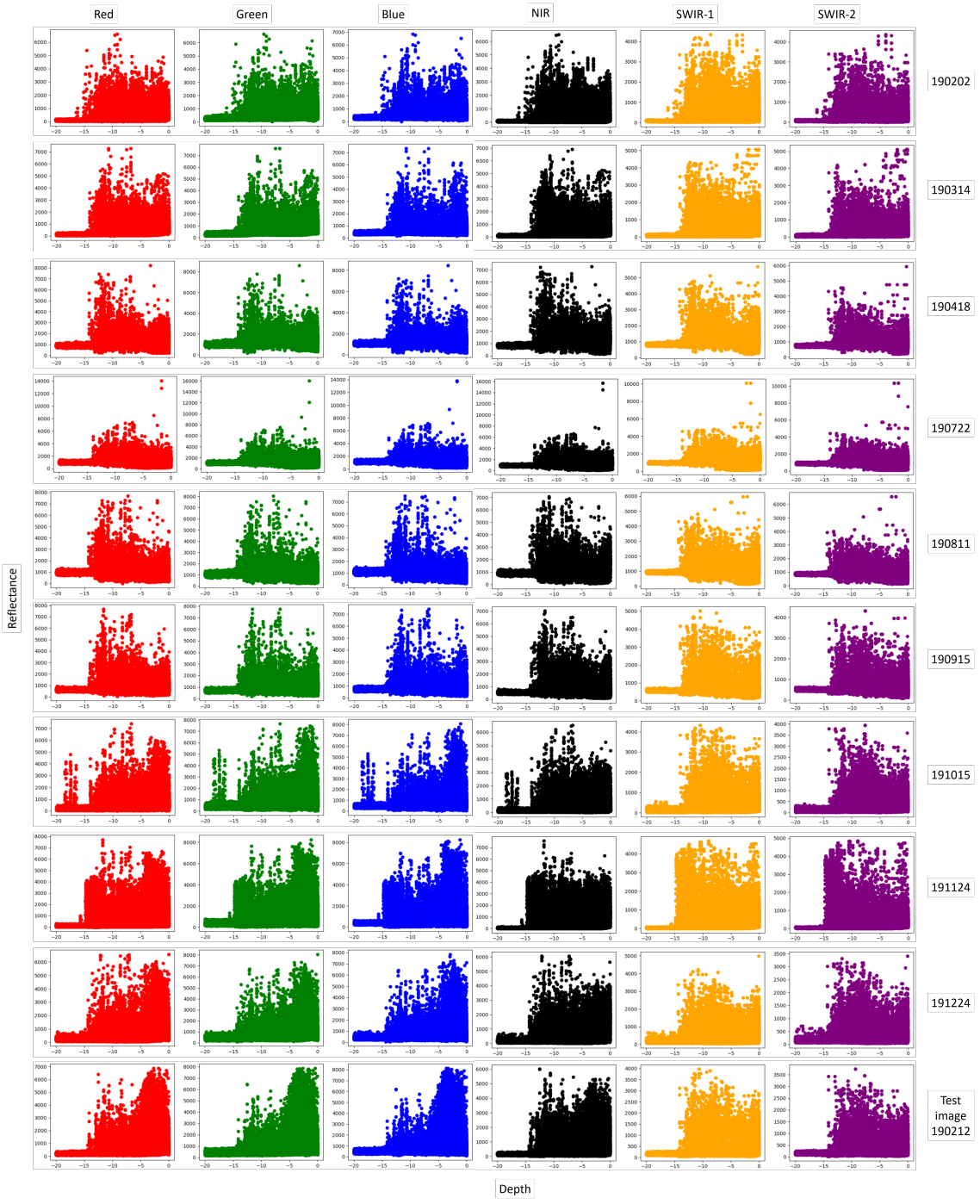


Figure A.12: Spectral reflectance signature towards depth of multi-temporal images in 2019 at AOI-3.

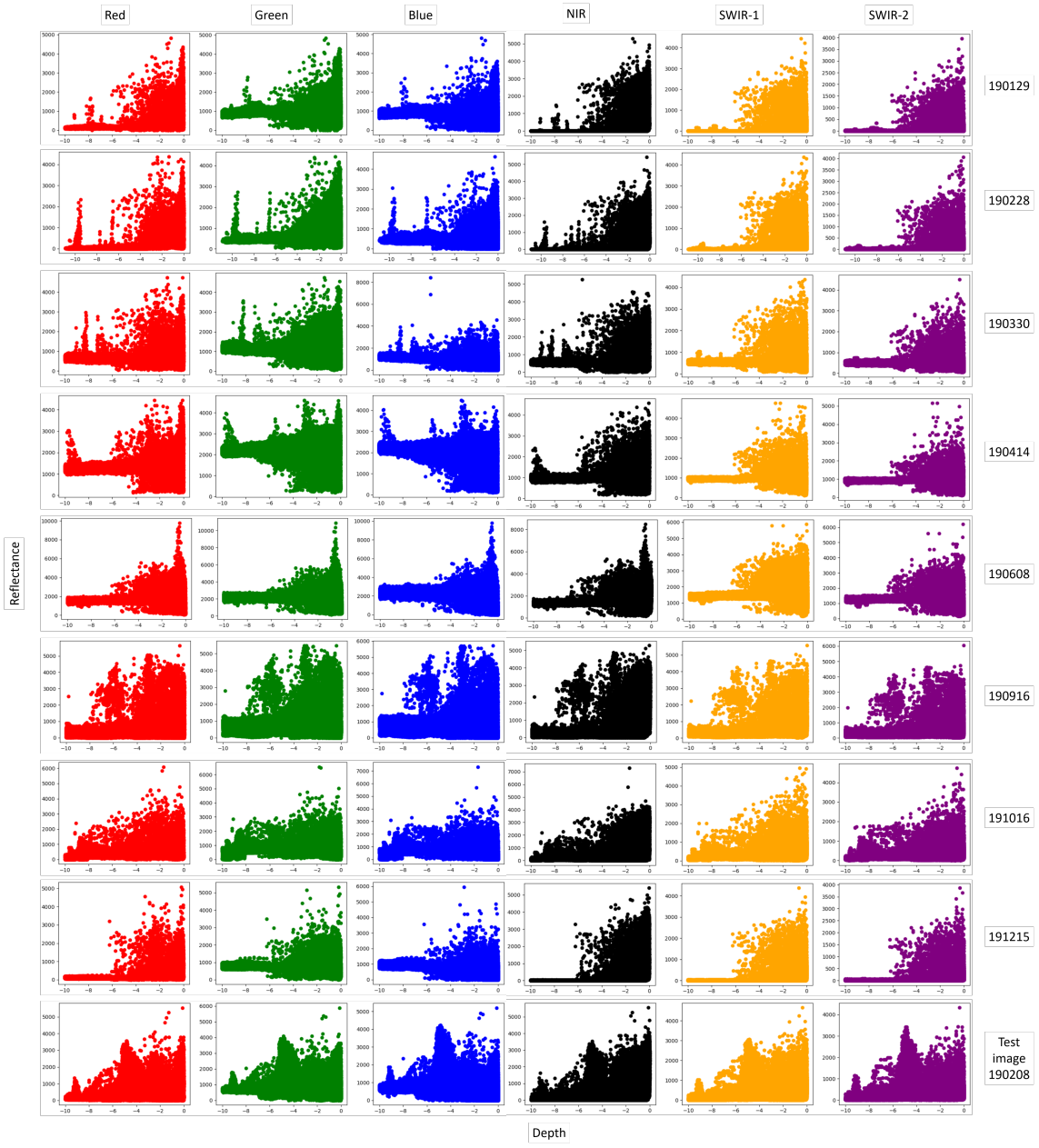


Figure A.13: Spectral reflectance signature towards depth of multi-temporal images in 2019 at AOI-4.

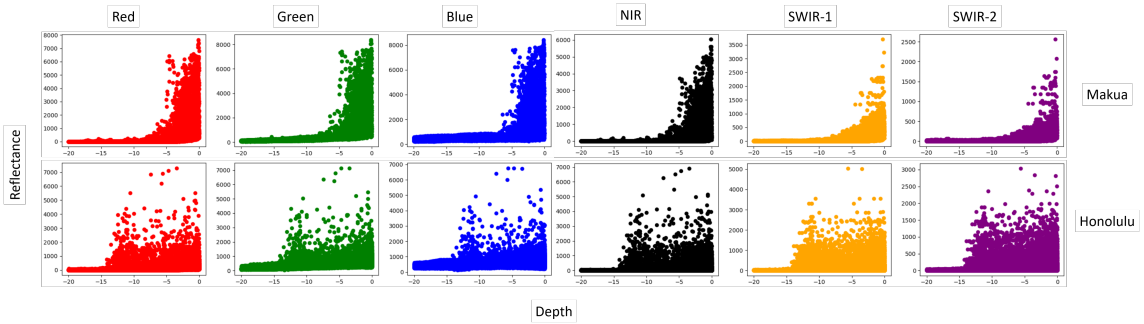


Figure A.14: Spectral reflectance signature towards depth of Oahu images at AOI-5 and AOI-6. Images are taken on 10 December 2019 in different footprints.

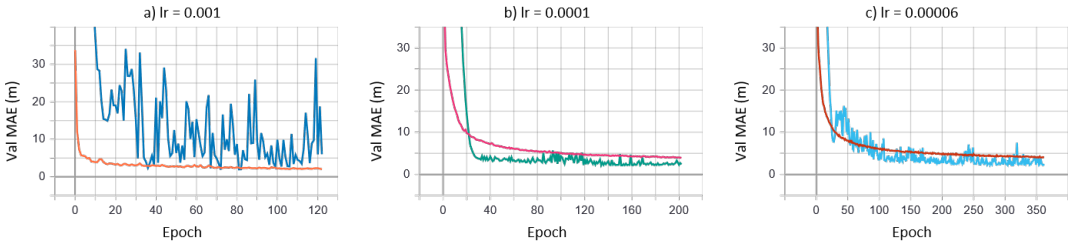


Figure A.15: Preliminary experiment for tuning the learning rate.

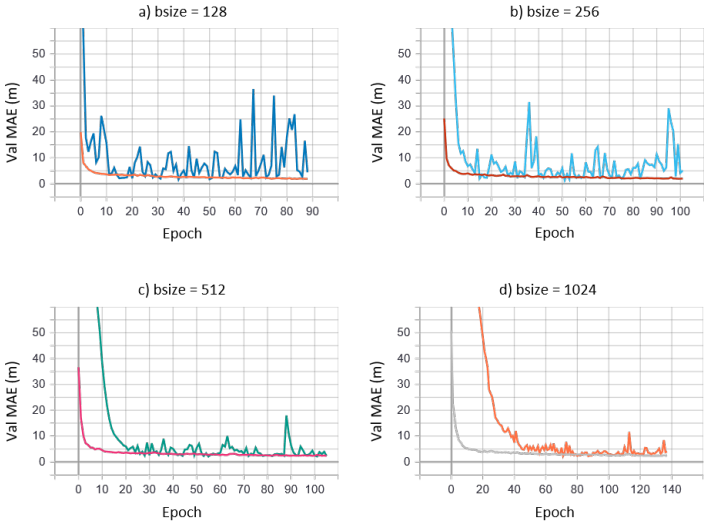


Figure A.16: Preliminary experiment for tuning the batch size.

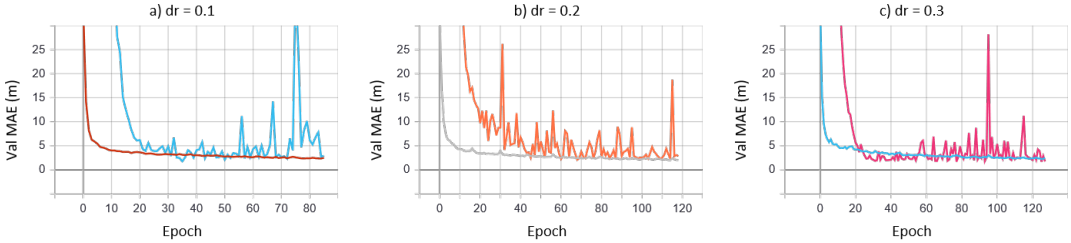


Figure A.17: Preliminary experiment for tuning the dropout rate.

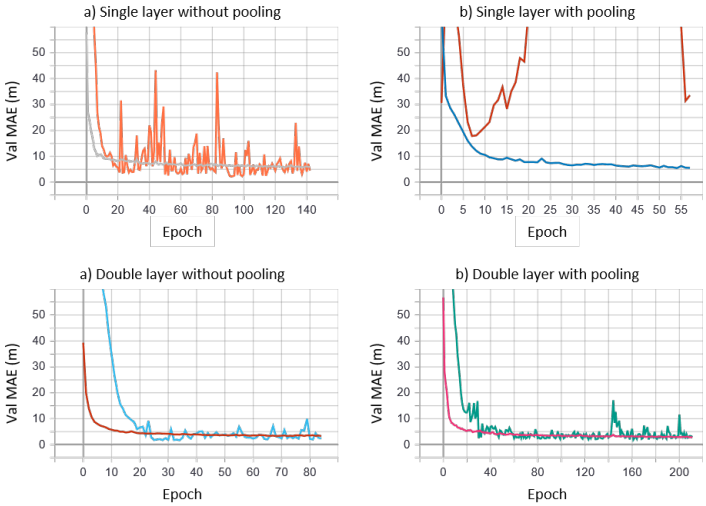


Figure A.18: Preliminary experiment to select CNN architecture using 5×5 window size.

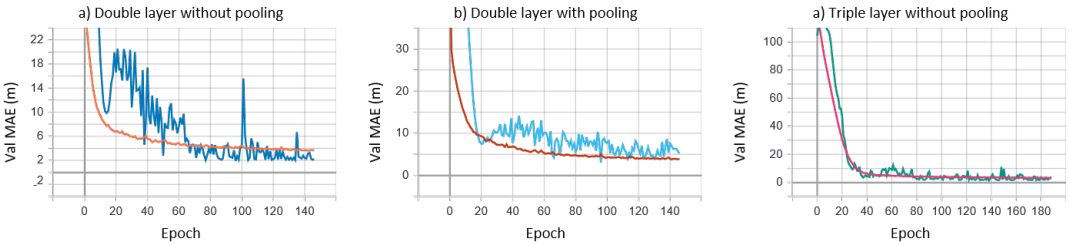


Figure A.19: Preliminary experiment to select CNN architecture using 7×7 window size.

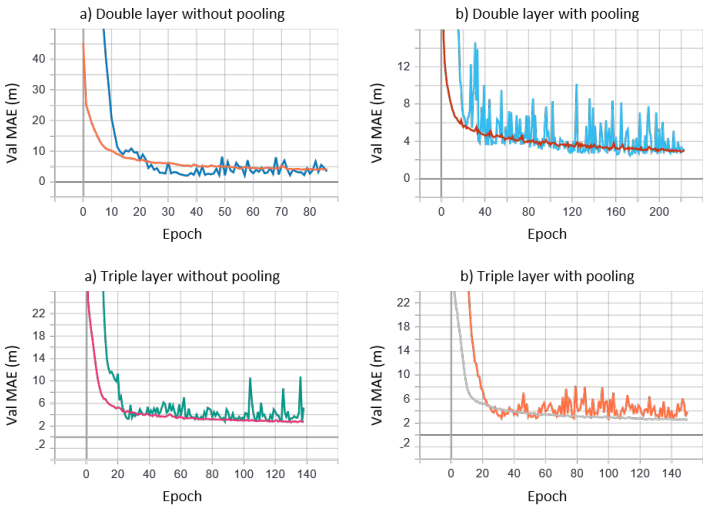


Figure A.20: Preliminary experiment to select CNN architecture using 9×9 window size.

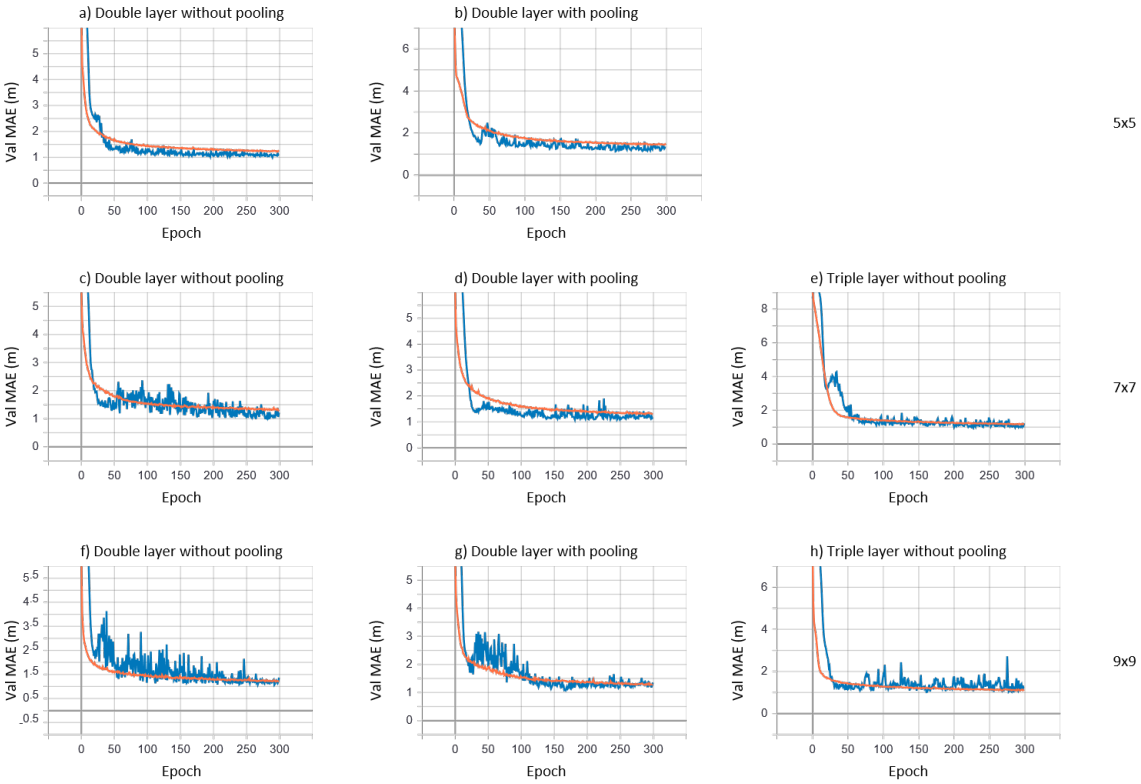


Figure A.21: CNN training using different architectures and different window sizes with the RGB band.

Bibliography

Ai, B., Wen, Z., Wang, Z., Wang, R., Su, D., Li, C., and Yang, F. (2020). Convolutional neural network to retrieve water depth in marine shallow water area from remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:2888–2898.

Antoine, D. and Morel, A. (1999). A multiple scattering algorithm for atmospheric correction of remotely sensed ocean colour (MERIS instrument): Principle and implementation for atmospheres carrying various aerosols including absorbing ones. *International Journal of Remote Sensing*.

Appeldoorn, R., Ballantine, D., Bejarano, I., Carlo, M., Nemeth, M., Otero, E., Pagan, F., Ruiz, H., Schizas, N., Sherman, C., and Weil, E. (2016). Mesophotic coral ecosystems under anthropogenic stress: a case study at Ponce, Puerto Rico. *Coral Reefs*.

Auret, L. and Aldrich, C. (2012). Interpretation of nonlinear relationships between process variables by use of random forests. *Minerals Engineering*.

Bauer, L. J., Kimberly, E., Roberson, K. K., Kendall, M. S., Tormey, S., and Timothy A. Battista (2012). Shallow-Water Benthic Habitats of Southwest Puerto Rico. *NOAA Technical Memorandum NOAA NOS NCCOS*, 155:37.

Bejarano, I. and Appeldoorn, R. (2013). Seawater turbidity and fish communities on coral reefs of puerto rico. *Marine Ecology Progress Series*, 474:217–226.

Briceño, H. O. and Boyer, J. N. (2015). 2014 Annual Report of the Water Quality Monitoring Project for the Water Quality Protection Program of the Florida Keys National Marine Sanctuary. Technical Report May 2015.

Brockmann, C., Doerffer, R., Peters, M., Stelzer, K., Embacher, S., and Ruescas, A. (2016). Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. In *European Space Agency, (Special Publication) ESA SP*.

Caballero, I. and Stumpf, R. P. (2019). Retrieval of nearshore bathymetry from Sentinel-2A and 2B satellites in South Florida coastal waters. *Estuarine, Coastal and Shelf Science*.

Casal, G., Harris, P., Monteys, X., Hedley, J., Cahalane, C., and McCarthy, T. (2020). Understanding satellite-derived bathymetry using Sentinel 2 imagery and spatial prediction models. *GIScience and Remote Sensing*.

Doerffer, R. and Schiller, H. (2007). The MERIS case 2 water algorithm. *International Journal of Remote Sensing*.

EPA (2018). What are the trends in the extent and condition of coastal waters and their effects on human health and the environment?

European Space Agency (2015). SENTINEL-2 User Handbook.

Gao, J. (2009). Bathymetric mapping by means of remote sensing: Methods, accuracy and limitations. *Progress in Physical Geography*, 33(1):103–116.

Hamylton, S. M., Hedley, J. D., and Beaman, R. J. (2015). Derivation of high-resolution bathymetry from multispectral satellite imagery: A comparison of empirical and optimisation methods through geographical error analysis. *Remote Sensing*, 7(12):16257–16273.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December:770–778.

International Hydrographic Organization (2008). IHO Standards for Hydrographic Surveys (S-44) 5th Edition.

Kabiri, K. (2017). Discovering optimum method to extract depth information for nearshore coastal waters from Sentinel-2A imagery-case study: Nayband Bay, Iran. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4W4):105–110.

Kanno, A., Koibuchi, Y., and Isobe, M. (2012). Shallow Water Bathymetry from Multispectral Satellite Images: Extensions of Lyzenga's Method for Improving Accuracy. *Coastal Engineering Journal*, 53(4):431–450.

Kanno, A. and Tanaka, Y. (2012). Modified lyzenga's method for estimating generalized coefficients of satellite-based predictor of shallow water depth. *IEEE Geoscience and Remote Sensing Letters*, 9(4):715–719.

Kanno, A., Tanaka, Y., Kurosawa, A., and Sekine, M. (2013). Generalized Lyzenga's Predictor of Shallow Water Depth for Multispectral Satellite Imagery. *Marine Geodesy*, 36(4):365–376.

Kaufman, Y. J. and Sendra, C. (1988). Algorithm for automatic atmospheric corrections to visible and near-ir satellite imagery. *International Journal of Remote Sensing*.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*.

LaRocque, P. and West, G. (1999). Airborne laser hydrography: an introduction.

Larsen, M. C. and Webb, R. M. (2009). Potential effects of runoff, fluvial sediment, and nutrient discharges on the coral reefs of Puerto Rico. *Journal of Coastal Research*.

Lyzenga, D. R. (1978). Passive remote sensing techniques for mapping water depth and bottom features. *Applied Optics*, 17(3):379.

Lyzenga, D. R. (1985). Shallow-water bathymetry using combined lidar and passive multispectral scanner data. *International Journal of Remote Sensing*, 6(1):115–125.

Lyzenga, D. R., Malinas, N. P., and Tanis, F. J. (2006). Multispectral bathymetry using a simple physically based algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8):2251–2259.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., and Gascon, F. (2017). Sen2Cor for Sentinel-2.

Manessa, M. D. M., Haidar, M., Hartuti, M., and Kresnawati, D. K. (2018). Determination of the Best Methodology for Bathymetry Mapping Using Spot 6 Imagery: a Study of 12 Empirical Algorithms. *International Journal of Remote Sensing and Earth Sciences (IJReSES)*, 14(2):127.

Manessa, M. D. M., Kanno, A., Sekine, M., Haidar, M., Yamamoto, K., Imai, T., and Higuchi, T. (2016). Satellite-Derived Bathymetry Using Random Forest Algorithm and Worldview-2 Imagery. *Geoplanning: Journal of Geomatics and Planning*, 3(2):117.

Matthew, M. W., Adler-Golden, S. M., Berk, A., Richtsmeier, S. C., Levine, R. Y., Bernstein, L. S., Acharya, P. K., Anderson, G. P., Felde, G. W., Hoke, M. L., Ratkowski, A. J., Burke, H.-h. K., Kaiser, R. D., and Miller, D. P. (2000). Status of atmospheric correction using a MODTRAN4-based algorithm. In *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*.

Mayer, B. and Kylling, A. (2005). *Technical note: The libRadtran software package for radiative transfer calculations - Description and examples of use*.

Misra, A., Vojinovic, Z., Ramakrishnan, B., Luijendijk, A., and Ranasinghe, R. (2018). Shallow water bathymetry mapping using Support Vector Machine (SVM) technique and multispectral imagery. *International Journal of Remote Sensing*, 39(13):4431–4450.

Moore, G. K. (1980). Satellite remote sensing of water turbidity. *Hydrological Sciences Bulletin*.

Mulder, A. E. (2020). *Semantic Segmentation of RGB-Z Aerial Imagery Using Convolutional Neural Networks*. PhD thesis, Delft University of Technology.

National Ocean Service (NOS) (2018). Hydrographic Surveys Specifications and Deliverables (2018 Edition). Technical Report May.

NOAA (2016). Exploring the Deep Ocean with NOAA.

NOAA (2017a). Benthic Habitat Mapping in Puerto Rico and the U.S. Virgin Islands for a Baseline Inventory.

NOAA (2017b). History of Hydrographic Surveying.

Perkins, T. (2012). Speed and accuracy improvements in FLAASH atmospheric correction of hyperspectral imagery. *Optical Engineering*.

Perkins, T., Adler-Golden, S., Matthew, M., Berk, A., Anderson, G., Gardner, J., and Felde, G. (2005). Retrieval of atmospheric properties from hyper and multispectral imagery with the FLAASH atmospheric correction algorithm. In *Remote Sensing of Clouds and the Atmosphere X*.

Polcyn, F., W, L. B., and I, J. S. (1970). The measurement of water depth by remote sensing techniques. page 37.

Quadros, N. (2016). Technology in Focus: Bathymetric Lidar.

Richter, R., Louis, J., and Berthelot, B. (2012). Sentinel-2 MSI – Level 2A Products Algorithm Theoretical Basis Document. *European Space Agency, (Special Publication) ESA SP*, 49(0):1–72.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351:234–241.

Sagawa, T., Yamashita, Y., Okumura, T., and Yamanokuchi, T. (2019). Shallow water bathymetry derived by machine learning and multitemporal satellite images. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 8222–8225.

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*.

Schläpfer, D., Borel, C. C., Keller, J., and Itten, K. I. (1998). Atmospheric precorrected differential absorption technique to retrieve columnar water vapor. *Remote Sensing of Environment*.

Shettle, E. P. and Fenn, R. W. (1975). MODELS OF THE ATMOSPHERIC AEROSOLS AND THEIR OPTICAL PROPERTIES. In *AGARD Conf Proc*.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–14.

Sonogashira, M., Shonai, M., and Iiyama, M. (2020). High-resolution bathymetry by deep-learning-based image superresolution. *PLoS ONE*, 15(7):1–19.

Spitzer, D. and Dirks, R. (1986). Remote sensing for resources development and environmental management: Classification of bottom composition and bathymetry of shallow waters by passive remote sensing. In *Proceedings of the Seventh International Symposium, Enschede, Netherlands*, volume 2, pages 775–777.

Stumpf, R. P., Holderied, K., and Sinclair, M. (2003). Determination of water depth with high-resolution satellite imagery over variable bottom types. *Limnology and Oceanography*, 48(1part2):547–556.

The Hawaii State Department of Health (2018). 2018 State of Hawaii Water Quality Monitoring and Assessment Report. Technical Report July 2018.

The Puerto Rico Environmental Quality Board (PREQB) (2016). Puerto Rico 305(b)/303(d) Integrated Report. (November):243.

Tonion, F., Pirotti, F., Faina, G., and Paltrinieri, D. (2020). A machine learning approach to multispectral satellite derived bathymetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 5(3):565–570.

Traganos, D., Poursanidis, D., Aggarwal, B., Chrysoulakis, N., and Reinartz, P. (2018). Estimating satellite-derived bathymetry (SDB) with the Google Earth Engine and sentinel-2. *Remote Sensing*, 10(6):1–18.

UN (2017). Factsheet: People and Oceans.

University of Hawai'i (2013). Ohau Coastal Geology.

Vanhellemont, Q. and Ruddick, K. (2014). Turbid wakes associated with offshore wind turbines observed with Landsat 8. *Remote Sensing of Environment*.

Vermote, E., Tanré, D., Deuzé, J. L., and Herman, M. (2006). Second Simulation of a Satellite Signal in the Solar Spectrum - Vector (6SV). *Spectrum*.

Vinayaraj, P. (2017). Development of Algorithms for Near-shore Satellite Derived Bathymetry Using Multispectral Remote Sensing Images.

Vinayaraj, P., Raghavan, V., and Masumoto, S. (2016). Satellite-Derived Bathymetry using Adaptive Geographically Weighted Regression Model. *Marine Geodesy*.

Wang, Z. J., Turko, R., Shaikh, O., Park, H., Das, N., Hohman, F., Kahng, M., and Polo Chau, D. H. (2021). Cnn explainer: Learning convolutional neural networks with interactive visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1396–1406.

Colophon

This document was typeset using LaTeX, using the KOMA-Script class `scrbook`. The main font is Palatino.