VConstruct: A Computationally Efficient Method for Reconstructing Satellite Derived Chlorophyll A Data

by

Matthew Ehrler
B.Sc., University of Victoria, 2019

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

VConstruct: A Computationally Efficient Method for Reconstructing Satellite Derived Chlorophyll A Data

by

Matthew Ehrler
B.Sc., University of Victoria, 2019

Supervisory Committee

Dr. Neil Ernst, Co-supervisor
(Department of Computer Science)

Dr. Yvonne Coady, Co-supervisor
(Department of Computer Science)

Dr. Maycira Costa, Non-Departmental Member
(Department of Geography)

Supervisory Committee

Dr. Neil Ernst, Co-supervisor
(Department of Computer Science)

Dr. Yvonne Coady, Co-supervisor
(Department of Computer Science)

Dr. Maycira Costa, Non-Departmental Member
(Department of Geography)

ABSTRACT

The annual phytoplankton bloom is an important marine event. Its annual variability can be easily recognized by ocean-color satellite sensors through the increase in surface Chlorophyll-a concentration, a key indicator to quantitatively characterize all phytoplankton groups. However, a common problem is that the satellites used to gather the data are obstructed by clouds and other artifacts. This means that time series data from satellites can suffer from spatial data loss. There are a number of algorithms that are able to reconstruct the missing parts of these images to varying degrees of accuracy, with Data INterpolating Empirical Orthogonal Functions (DINEOF) being the most popular. However, DINEOF has a high computational cost, taking both significant time and memory to generate reconstructions. We propose a machine learning approach to reconstruction of Chlorophyll-a data using a Variational Autoencoder (VAE). Our method is 3-5x times faster (50-200x if the method has already been run once in the area). Our method uses less memory and increasing the size of the data being reconstructed causes computational cost to grow at a significantly better rate than DINEOF. We show that our method's accuracy is within a margin of error but slightly less accurate than DINEOF, as found by our own experiments and similar experiments from other studies. Lastly, we discuss other potential benefits of our method that could be investigated in future work, including generating data under certain conditions or anomaly detection.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank:

- Dr. Neil Ernst and Dr. Yvonne Coady for their mentorship, expertise, guidance and enthusiasm which was essential in completing this work.
- Dr. Christian Marchese, Dr. Maycira Costa, and Dr. Fernanda Giannini for providing their expert knowledge of oceanography.
- Ann Edelstein and Olive for providing support and motivation integral to finishing this thesis.

Chapter 1

Introduction

With over 500 Earth observation (EO) satellites launched in the past 50 years and around 140 planned missions listed on the CEOS database, the world is well into the era of "Big Earth Data" [8, 14, 45]. The main satellite mission referenced in this paper, Sentinel-3, collects data from a specific area near daily, with an average revisit time of <2 days [11]. Given that the size of EO datasets are constantly growing, computationally expensive EO algorithms, particularly those relying on time series, need to be reexamined for less expensive alternatives. Finding less computationally expensive alternatives to existing EO algorithms will save scientists time, money on expensive hardware, and will allow for the processing of higher resolution areas without sacrificing the size of area it can cover.

Modern Deep Learning (DL) techniques (i.e Neural Networks) have been identified by Imperatore and Riccio [18] to benefit Geoscience and Remote Sensing in three main areas: as a code accelerator for computationally expensive models, a model for processes one cannot model but do have data to teach, and as a model for classification problems. The main focus of DL research in Remote Sensing/Earth Observation is in the second two areas. The research in these areas has proven very successful, such as work by Anantrasirichai et al. [2] who designed a classifier for volcanic deformation

which could indicate eruption, or a model from Mudele et al. [34] for deriving the population of a type of mosquito from EO data. Models in the code accelerator category could provide a way to process the massive amount of new EO data being generated daily, at a much lower computational cost.

A prime example of a computationally expensive EO algorithm is the Data INterpolating Empirical Orthogonal Function (DINEOF) algorithm [4]. This algorithm is important to EO and climate change research, as it is one of the most used algorithms for reconstructing missing sections of oceanographic data [42]). It is commonly used to reconstruct Sea Surface Temperature (SST), Chlorophyll-a (Chl-a), and other satellite derived biogeochemical variables as they suffer from significant spatial data loss [39]. This data loss is due to clouds, sunglint, or other factors which can introduce error to the atmospheric correction process, rendering the confidence in the corrected data too low to be used [39]. DINEOF, though accurate, is fairly slow [42], and heavily reliant on temporal data [16]. Additionally no implementation suitable for massively parallel processing could be found and, as shown later, an implementation is unlikely to ever be found. These factors make processing the massive amount of EO data now accessible very resource intensive and time consuming. For instance, a reconstruction of a Globcolour derived Chl-a dataset of 6871 daily images, each consisting of 313x458 pixels, took roughly 96 hours on a M8ms Azure VM instance (8 cores, 218.75 GB of RAM) and cost approximately $200CAD in compute time alone (S.Pramlall, Personal Communication, August 16 2021).

In this work, a satellite-derived Chl-a reconstruction algorithm based on the Variational Autoencoder (VAE) from Kingma and Welling [22] as well as Attribute2Image's improvements in making generated images less random [44] is proposed. A VAE is a neural network with 2 main parts, an Encoder and Decoder. The Encoder takes an input image and learns how to compress it into a smaller latent space that can be

represented by a distribution. The Decoder then learns how to generate a realistic looking image by drawing from this distribution. Yan et al. [44]'s improvements in Attribute2Image give the Decoder some extra information or attributes (i.e., temperature) to use when generating the image, allowing the user to control what is generated. The dimensionality reduction in a VAE is similar to the Singular Value Decomposition (SVD) used in DINEOF, making it a good candidate for a replacement algorithm.

The process introduced in this work, VConstruct, is similar to Attribute2Image as well as Ivanov et al's inpainting process but modified to reconstruct biogeochemical data [19, 44]. VConstruct is tested using satellite-derived Chl-a data from the Salish Sea area surrounding Vancouver Island in Canada, and data from the Mediterranean sea. VConstruct is compared to the most popular method, DINEOF, using experiments modeled after Hilborn and Costa [16]'s experiments on satellite-derived Chl-a Data in the Salish Sea. The Salish Sea area was chosen as it shows a high variability in Chl-a concentration as well as high cloud cover, factors DINEOF is sensitive to [16]. The Mediterranean sea is used as well due to the general low cloud cover, which gives a large number of training samples [32]. The effects these different areas play on the quality of reconstruction is discussed. Computational costs of VConstruct and DINEOF including time and memory are also compared. Additionally, qualitative aspects such as generalizability and climate change resistance are compared.

The results show that deep learning algorithms are good candidates as code accelerators to replace or augment some resource intensive EO algorithms. VConstruct is able to reduce processing time by 3-5x depending on the input data and maintain similar but slightly less accuracy. Most of VConstruct's runtime comes from generating a reconstruction model for the given area, which can be reused for reconstructing any future days. If a VConstruct model has already been generated for that area, a new

set of images can be given to the model and reconstructed with roughly 50-200x less processing time when compared to DINEOF. A less resource intensive reconstruction algorithm will allow researchers to keep up with the massive influx of new EO data and will also lower the cost of entry with regards to time and hardware for anyone who wishes to process this data in the future. The same globcolour reconstruction discussed above could be performed with VConstruct locally on a mid tier consumer grade GPU in roughly 12 hours, likely less. On Azure this experiment could likely be run on a Standard_NV8as_v4 (1/4 of a Radeon Instinct MI25 Accelerator) for roughly $12CAD, or if this is too small an instance with a full Nvidia V100 would cost roughly $48CAD for the same time period.

Chapter 2

Background and Related Works

2.1 Chlorophyll-A

Chlorophyll-A (Chl-a) is considered an "Essential Climate Variable" for measuring and predicting climate systems and ocean health [5]. Chl-a is a commonly used metric to estimate phytoplankton biomass (measured in units of mg/m^3) and can be derived from ocean colour data gathered from satellites [1, 37]. Chlorophyll-a data can be used in studies of primary productivity [17], to detect changes in phytoplankton bloom phenology [28, 29, 41] or to locate harmful algae blooms, which can be fatal to marine life [39]. As climate change progresses, phytoplankton bloom phenology and the conditions for harmful algae blooms will be significantly impacted, making it all the more necessary to monitor them through EO data [33].

2.2 Other approaches to Chl-a reconstruction

The most popular approaches to Chl-a reconstruction are often based off finding Empirical Orthogonal Functions (EOFs) and using them to interpolate missing data. These methods are quite useful in remote areas with low satellite cover, like the Arctic

Ocean and Salish Sea [16, 28]. EOFs are usually calculated through Singular Value Decomposition, where the algorithms differ is in how each initially fills in the missing data and decides when to stop. Of the existing EOF algorithms DINEOF is the most accurate and commonly used [16, 26, 42]. DINEOF works by randomly generating candidates for the missing data and then refining those candidates via calculating EOFs until convergence. Once convergence is achieved it then adds a second EOF mode and does it all again. This process repeats until the accuracy converges through cross validating [39]. However DINEOF can be rather slow to process, requiring many more iterations then the next most accurate EOF algorithm, Recursively Subtracted Empirical Orthogonal Functions (RSEOF) [42]. It is slow enough that Taylor et al. [42] recommend using RSEOF to process large biogeochemical datasets. Liu and Wang [26] needed to split their 30 day dataset across 16 different parallel instances to process it in a reasonable amount of time, citing DINEOF's inefficiency as a limiting factor. Each of the 16 instances only had access to 1/16th of the available spatial data, limiting the potential maximum accuracy.

Alternatively, Konik et al. [23] propose a non EOF based model for reconstruction that performs well, however it is highly dependent on having a model to simulate the value you are trying to reconstruct. This makes it less generalizable than EOF methods and ties its accuracy to the accuracy of the model.

2.3 Other Machine Learning approaches to Chl-a reconstruction

Machine Learning techniques have also been successful in reconstructing Chl-a data. For example, Park et al. [35] use a tree based model to reconstruct Chl-a in polar regions. This method is effective but similar to Konik et al. [23] requires knowledge of

the domain of the data being reconstructed to properly tune it. DINEOF and other EOF based methods do not have this draw back. Park et al. [35] state that the wrong predictors could introduce noise in their model, so a "...deep understanding of the factors that influence [Chl-a] is needed" [35]. These less generalizable models would potentially have to be rebuilt from the ground up for new areas or new biogeochemical variable, assuming they would even work in that area. Generalizable methods like DINEOF have been shown to work well in a large variety of situations with minimal tuning, making it a more useful tool to the end user as they likely will not want to change methods every time they work in a new area. This transferability, along with its accuracy, is likely why DINEOF is so widely adopted and why it is compared to in this work.

DINCAE is a very new machine learning approach to reconstructing data [3], which has also been shown to work on Chl-a [15]. DINCAE, like VConstruct uses Neural Networks. However DINCAE primarily focuses on accuracy and not computational cost. Like DINEOF, it also relies on temporal data which has a significant memory cost.

There are also several other machine learning techniques that perform tasks on Chl-a similar to reconstruction. Jin et al. [20] uses a neural network to predict Chl-a in a small area using hydrodynamic model data. Chen et al. [9] propose a method for correcting data that is gathered from EO satellites but has too low a confidence to be useful using a random forest.

Chapter 3

Model

The proposed model, VConstruct, is a deep learning model based on a Variational Autoencoder [22]. It consists of an Encoder, Decoder and Attribute neural network and is arranged as shown in Fig. 3.1. All network layers are fully connected layers with ReLU activation functions, a simple neural network architecture [43]. The exact dimensions of the layers are shown in Fig. 3.1.

A conventional VAE consists of just the Encoder and Decoder network and is attempting to optimize the objective function of:

$$logP(X) - D_{KL}[Q(z|X)||P(z|X)] = E[logP(X|z)] - D_{KL}[Q(z|X)||P(z)] \quad (3.1)$$

$P(X)$ is the likelihood of some data given some encoding X, $Q(z|X)$ is the Encoder network, $P(X|z)$ is the Decoder network, E is reconstruction loss and D_{KL} is Kullback-Leibler divergence [10, 22]. The Encoder network compresses an image down to a lower dimensional latent space and learns a distribution it can later sample from during testing when the complete image is unknown. The Decoder takes the output of the Encoder network, or random sample from the learnt distribution, and attempts to reconstruct the original image. Kullback–Leibler divergence and Reconstruction loss are used as loss functions. Kullback-Leibler divergence forces the learnt

distribution into a normal distribution and Reconstruction loss ensures minimal error between the input and output images [22].

The addition of the Attribute network is based off the work of Yan et al. [44] and Ivanov et al. [19]. The network extracts an attribute vector from a cloudy image which represents what is "known" about the information not obstructed by clouds. This attribute vector then influences the previously random image generation of the Decoder network so that it generates a potential reconstruction. To optimize the influence of the encoding the objective function (Equation 3.1) becomes that of a Conditional Variational Autoencoder (CVAE):

$$logP(X|c) - D_{KL}[Q(z|X,c)||P(z|X,c)] = E[logP(X|z,c)] - D_{KL}[Q(z|X,c)||P(z|c)] \tag{3.2}$$

c represents the attribute vector [10, 40].

These three networks make up the training configuration of VConstruct and can be seen in Fig. 3.1. When testing the Encoder network cannot be used as the complete image is not known, so the network is replaced with a random sample from the distribution learnt in training. The parts that switch out are indicated by the dashed lines. Once the model is fully trained the Encoder network can be entirely deleted, allowing for VConstruct to use less memory for generating actual reconstructions.

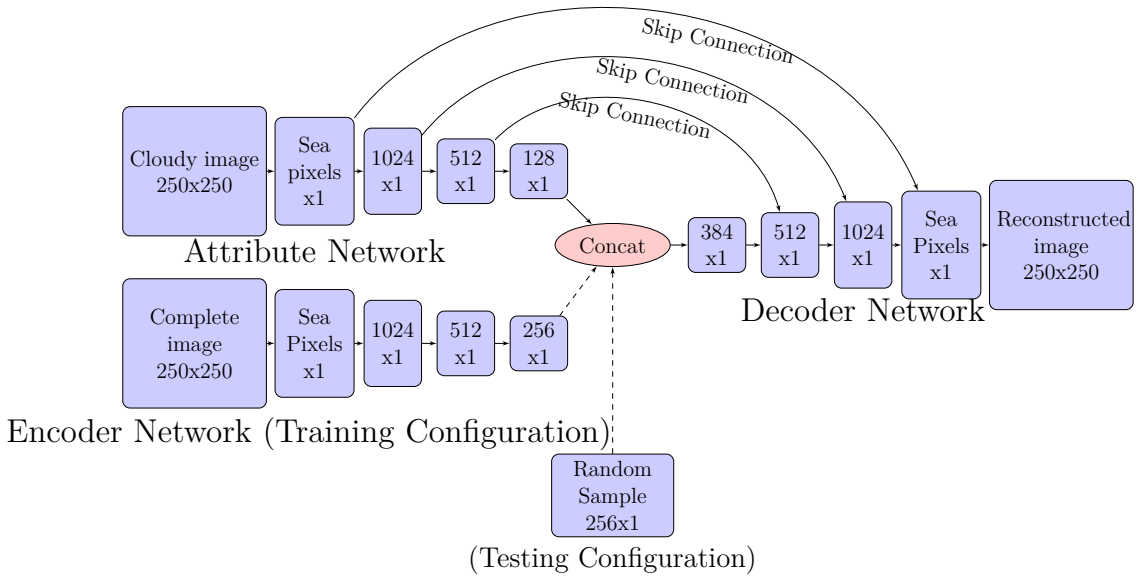

Figure 3.1: Configuration of VConstruct

Chapter 4

Model Validation

To prove the validity of VConstruct, it is shown to be able to provide accurate reproductions. This is tested by removing an area from a set of images and attempting to reconstruct it using both DINEOF and VConstruct, allowing accuracy to be tested under similar conditions. The area removed from the set of testing images is generated based off cloud cover in other images. This validation process is similar to Sirjacobs et al. [39].

4.1 Dataset and Preprocessing

The Salish Sea dataset comes from the Algae Explorer Project[1]. This dataset is daily mosaiced data spanning from 2016-04-26 to 2020-09-30, consisting of 1566 images total, and has a resolution of 300m. Algae Explorer uses data from Sentinel-3A coming from CODA [2] and is processed according to Giannini et al. [13]. Results for daily images are shown in 2 subsections of the Salish Sea, the Fraser River mouth as it is identified as a problem area for DINEOF [16]), and an area off the coast of Victoria as it is a good midpoint between the Mediterranean Sea and the Fraser River

[1] https://algaeexplorer.ca/
[2] https://www.eumetsat.int/coda

mouth in terms of complexity. The Victoria coast also served as a good diagnostic area due to its wide range of values and distinct shape.

For the Mediterranean Sea, the 8-day composite time series (from 1998 to 2021) of surface satellite-derived Chlorophyll-a at 4 km of spatial resolution was obtained from the GlobColour Project[3]. The surface chlorophyll-a values are gathered by using the Garver-Siegel-Maritorena (GSM) model [30] combining, when possible, data from different sensors (i.e., SeaWiFS, MERIS, MODIS, Sentinel-3, and VIIRS), and ultimately providing a merged product with an increased spatiotemporal coverage [31]. These data sources represent a variety of types of data, preprocessing steps, and instruments, broadening the testing scope. Though the dataset contained 1566 images, only 450 of them could be used in the tests below due to DINEOF's memory limitations. The data used is the first 450 images, ranging from 1998 to 2006.

The data for DINEOF is preprocessed using a similar process to Hilborn and Costa [16], removing major outliers and marking the clouds as appropriate. The preprocessing for the data for VConstruct uses a similar method to Han et al. [15], outliers are removed, clouds are marked, and the Chl-a anomaly is calculated to use as input.

4.2 Testing

Five days from each of the 3 geographic areas are selected at random from all days with low cloud cover. In total the testing set consists of 15 images. This allows artificial clouds to be added and accuracy be measured by comparing to the original complete image. The artificial clouds patterns added are cloud patterns extracted from other images, ensuring real world viability. Each image has thousands to tens of thousands of pixels removed, making the testing set rather large. For the purposes of testing in

[3]https://hermes.acri.fr/index.php

Section 5 all images are reconstructed, but statistical metrics are only calculated for these 5 to ensure accuracy is only reported for properly withheld data. The images coming from both DINEOF and VConstruct are interpolations of the entire image, including pixels that were not missing in the first place. It is common practice (but not always done) to use the reconstructions for only the missing pixels, combining them with the non missing pixels from the original incomplete image to make the final reconstruction [16, 26]. The combined image is used as the reconstruction when calculating the metrics below to ensure the metrics are only being calculated for the removed areas.

Images are compared with their reconstruction by calculating RMSE (Root Mean Squared Error) and mean ratio. RMSE is calculated via Equation 4.1 where x_i is the reconstructed value, and y_i is the true value. The logged version of the values are used for these calculations.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{4.1}$$

4.2.1 DINEOF Testing

Since DINEOF is not "trained" like ML models, conventional ML testing with test/ train/validate datasets cannot be performed. For this experiment the five testing images selected in preprocessing are overlaid with artificial clouds to create the testing set, which is then inserted back into the full set of images. Samples are shown in the Appendix, Figs. 4.1, A.1, and A.2. After running DINEOF, the pixels that were removed by the artificial cloud cover are compared with the respective pixels in the reconstructed output from DINEOF.

This testing scheme slightly biases the experiment towards DINEOF as testing images are usually completely withheld when training deep learning models. DINEOF

needs access to the cloudy testing images in order to generate EOFs, whereas these images are withheld from VConstruct. This is unfortunately unavoidable but does not invalidate the results, as neither algorithm has access to the removed pixels.

4.2.2 VConstruct Testing

To train and test VConstruct the 5 testing images are completely removed from the dataset, they form the "testing" dataset. Of the remaining images those with low cloud cover are used as the "training" dataset and those with higher cloud cover to generate "training" clouds. VConstruct's model is learnt by showing it "training" images along with the same image with "training" clouds artificially inserted. There are many more "training" clouds then those used for testing, allowing VConstruct to learn how to reconstruct a wide variety of images. VConstruct uses these images to optimize the equations described in Section. 3.

The model is trained for 150 epochs (each training image is "shown" to the model 150 times with different cloud patterns). More epochs can improve accuracy as the model gets to see each image more, but they also increase run-time. For VConstruct, higher values did not seem to improve accuracy by much, which is why 150 is used. But as with all deep learning hyperparameters there is a chance other values or combinations will give better results.

After training the test dataset is used to test identically to DINEOF, reconstructing the same images with the exact same cloud patterns.

4.3 Accuracy Results

Table 4.1 presents the mean results of reconstructing the five randomly selected testing images in each area. RMSE (Root Mean Squared Error) and the ratio between the

Table 4.1: Testing Results. RMSE = Root mean squared error, mean ratio is the the mean of the reconstructed image to the mean of the true image. A ratio of 1 means there is no difference. Bolded values show the better score in each area for each metric

Area	**RMSE**				**Mean Ratio**	
	DINEOF		**VConstruct**		**DINEOF**	**VConstruct**
$(\#ofimages) \times (\#ofpixelslat) \times (\#ofpixelslong)(Area)$	$log_{10}\ mg\ m^{-3}$	$mg\ m^{-3}$	$log_{10}\ mg\ m^{-3}$	$mg\ m^{-3}$		
450x1040x385 (Mediterranean Sea)	**0.0695**	**1.17**	0.0881	1.22	0.9979	**1.0004**
1566x250x250 (Victoria Coast)	**0.0531**	**1.13**	0.0851	1.22	1.0140	**1.0002**
1566x250x250 (Fraser River Mouth)	**0.1008**	**1.26**	0.1347	1.36	**0.9827**	1.0302
Mean	**0.0744**	**1.19**	0.1026	1.27	**0.9982**	1.0102

mean of the reconstructed image and true image are reported.

The algorithms perform similarly, with DINEOF slightly outperforming VConstruct on all tests except for the mean ratio in the Victoria Coast and Fraser River Mouth. To determine if the level of error of VConstruct is acceptable, results from Table. 4.1 can be compared with other similar studies(e.g Hilborn and Costa [16], Liu and Wang [26]). These studies are performed under a variety of conditions, so it cannot be said whether or not whether they would be outperformed. But errors in Table. 4.1 can be compared to theirs to give an idea of "acceptable" error.

Liu et al. perform reconstructions with DINEOF, and report a mean ratio of .997 for coastal and inland waters which matches up with the Salish Sea areas and 1.012 for oligotrophic waters which matches up with Mediterranean sea [24]. The mean ratios are well within the .287 and .164 standard deviations reported by the authors, suggesting similar performance. Hilborn and Costa [16] report 0.11 RMSE ($log_{10}\ mg\ m^{-3}$) as their best accuracy when comparing satellite derived Chl-a to reconstructed Chl-a from DINEOF. This is similar to the mean of the measured RMSEs for VConstruct in Table. 4.1. Thus it can be concluded that VConstruct is performing nearly as well as the commonly used DINEOF approach, with, as explained in a subsequent section, significant advantages in other areas.

The testing images from the Mediterranean sea are shown in Fig. 4.1, samples

from other areas are found in the appendix in Fig.A.1 and A.2. The colour bar ranges were chosen manually to be able to compare reconstructions with full images the best. However due to the different ranges, colours should not be compared between different geographic areas. For comparison across different areas see Fig. 4.3. Brown represents land, white represents missing values. The Full Image is the original image before cloud cover is added. The Cloudy Image is the image with artificial clouds that is provided to the reconstruction algorithms to reconstruct. The VCon and Dineof images are the reconstructed images from the respective algorithm with that day's RMSE shown. Since the reconstructed images are the Full Image with the artificial cloud cover filled in, any values missing in the original full image will still be missing (though could be filled if needed). This ensures metrics are calculated for the removed area only.

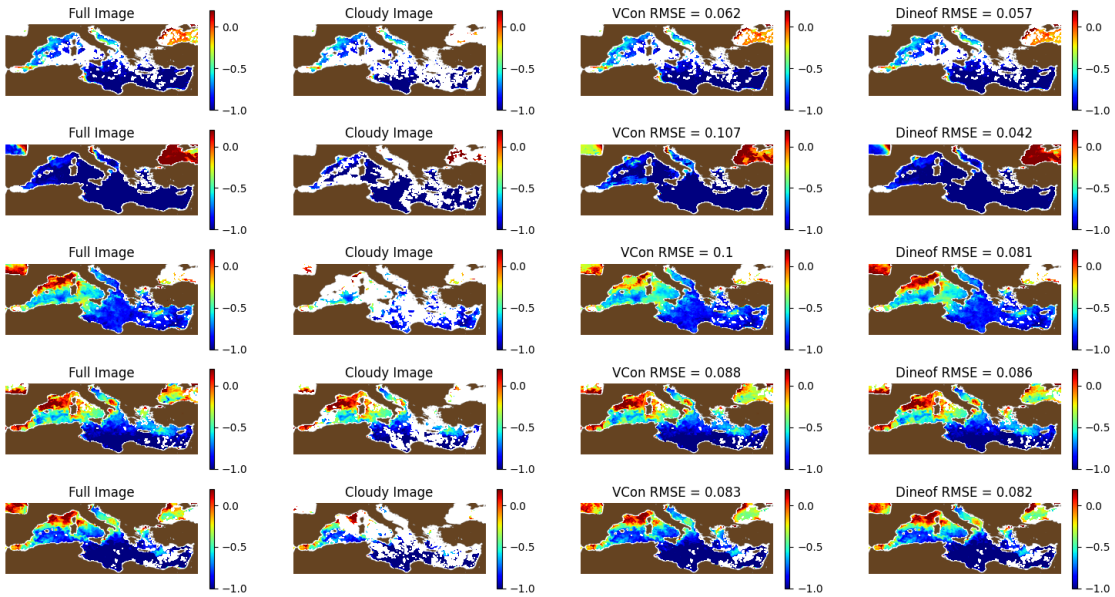

Figure 4.1: Reconstructed testing images for the Mediterranean Sea. Units are in $log(mg/m^3)$. Latitude ranges from 46.02°to 30.02°. Longitude ranges from -6.31°to 36.979°

Visually both reconstructions are fairly similar, the most significant difference in

Fig. 4.1 is VConstruct was unable to reconstruct the dark red area in the top left of the third image. A likely explanation is discussed in the next section. Both algorithms seem to do well reconstructing areas with both high and low concentrations.

4.4 Effects of Geographic Area on Results

Since the algorithms perform better in some area than others some reasons for this difference are explored. To see the difference more clearly, a plot comparing the reconstructed value with the actual value is shown in Fig. 4.2.

In the charts in Fig. 4.2 the closer the line of best fit is to $y = x$ the better the reconstruction is. From Fig. 4.2 and Table 4.1, the algorithms are both performing best in the Mediterranean Sea, followed by the Victoria Coast, then performing worst in the Fraser River Mouth. This matches Hilborn and Costa [16]'s finding that the Fraser River mouth is a problem area. The mean and coefficient of variation of Chl-a in mg/m^3 are plotted to provide a potential explanation in Fig. 4.3. The colourmaps for Figures 4.1,A.2,A.1 was chosen to show the most variation in that particular area to allow for easy comparison of reconstructions to the full image. The different ranges in colourbars however makes it hard to compare across the different areas. For Fig. 4.3 a consistent range is used to allow for comparison across areas, though patterns are harder to detect.

From Fig. 4.3 it can be seen that ranking the areas from lowest coefficient of variation to highest matches the order of reconstruction performance from highest to lowest. Areas with high variability are harder to reconstruct, matching Hilborn and Costa [16]'s findings. An additional reason for the difference in performance could be the amount of data present, ordering the areas by number of ocean pixels present matches the order of reconstruction performance. The Mediterranean Sea has 180505,

Victoria Coast has 38129, and Fraser has 34175. This makes sense as the more data present the easier it is to learn spatial patterns. There are not enough data points to make any strong conclusions about the effect temporal length has on performance. However it is interesting to note that the Mediterranean Sea dataset has roughly one third the amount of time points as the other areas but performs significantly better.

A similar matching ranking can be observed by looking at how the actual values are distributed in Fig. 4.2. The best performing Mediterranean sea appears to have its actual values fairly evenly distributed across its range (-1 to 1). The middle performing Victoria Coast is also fairly distributed across its range (0 to 1) but has somewhat of a gap around the .5 mark. The worst performing Fraser river has the majority of its values fairly concentrated in one small section of its range of values (0-2). This shows the importance of a good representative training set, areas and samples not represented by the training set will not be reconstructed as well.

Further investigating the effect of variation and having a good training set, a sample of 2 individual testing images from the testing set for the Victoria Coast are shown. In the "Victoria VCon" plot in Fig. 4.2 a number of poorly reconstructed pixels can be seen in the higher range. Separating the correlation charts into those for individual days (samples in Fig. 4.4), it can be seen that this is likely coming from one poorly reconstructed day in Victoria (See also Fig. A.1). This day has significantly higher Chl-a values then all other testing days, and is shown compared to a more normal day from the testing set in Fig. 4.4 (The rest of the testing days can be seen in Fig. A.1). VConstruct performs significantly worse on this anomalous day when compared to another more normal day, or the composite (DINEOF too but to a lesser degree). The poor performance is likely due to the fact that there are very few days like this one in the training set, making it difficult to learn the patterns present. This negative factor of VConstruct can however be leveraged to make an

effective anomaly detector as will be discussed in Section 6.

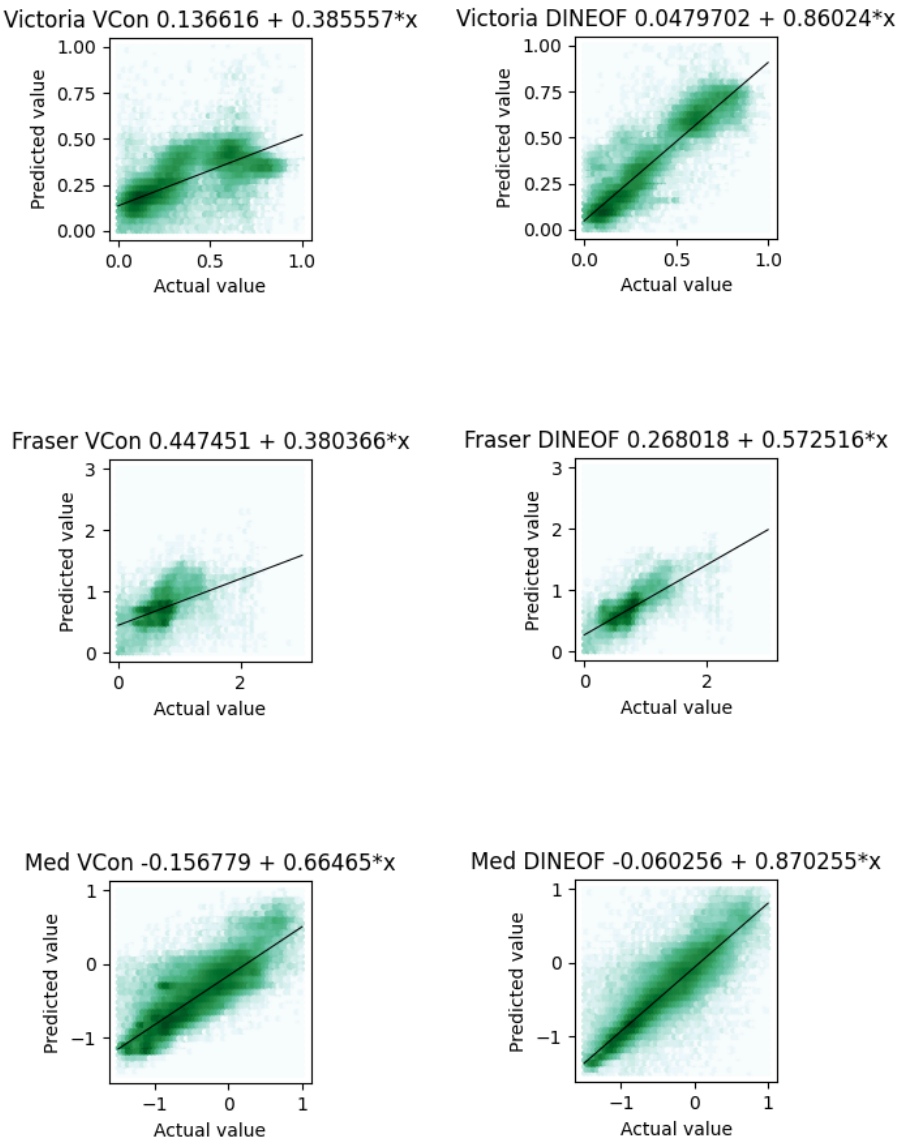

Figure 4.2: Correlation between actual and predicted values for each algorithm in each geographic area. Units are $log(mg/m^3)$. Values plotted are a composite of all 5 testing days. The darker the green the more values in that area. The black line represents the line of best fit and the equation is provided in the graph title in the form $y = b + mx$.

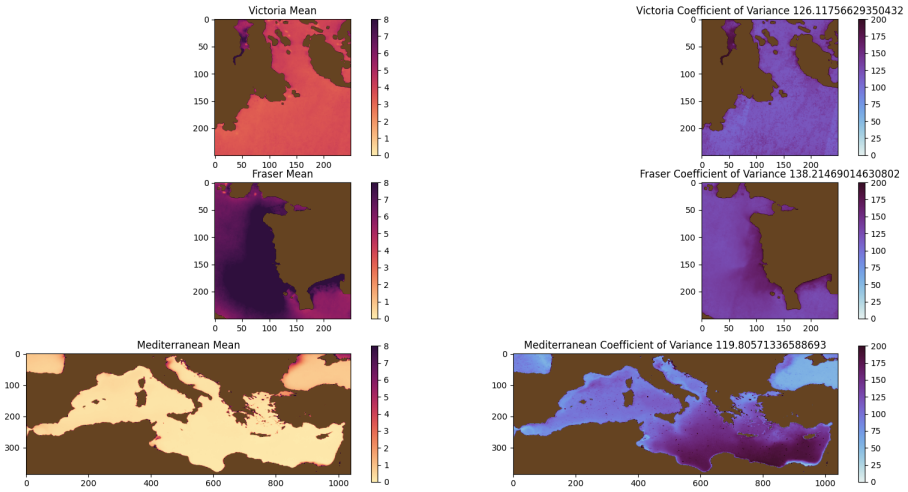

Figure 4.3: The mean of each geographic area's Chl-a measurements are on the left. Units are mg/m^3. The coefficient of variation for each area is shown on the left, the units used in the calculation were mg/m^3

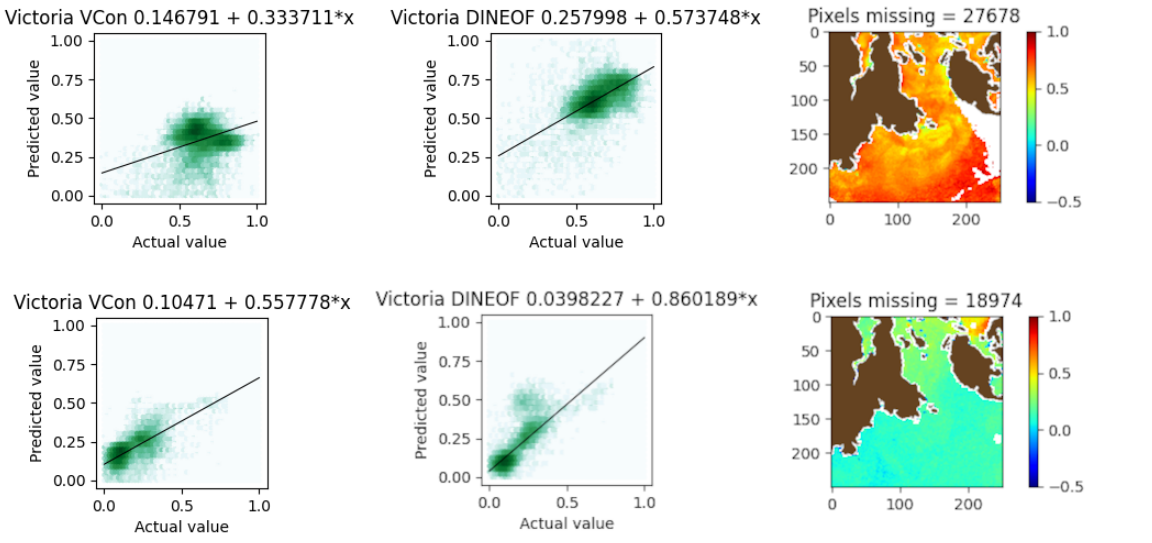

Figure 4.4: Correlation chart for 2 different testing days in Victoria. Units are $log(mg/m^3)$. The darker the green the more values in that area. The black line represents the line of best fit and the equation is provided in the graph title in the form $y = b + mx$.

Chapter 5

Computational Costs

Model accuracy is only one relevant comparison, albeit the most important one. In this section computational costs are investigated in two different areas: running time, and memory cost (both in terms of image spatial size and temporal size). These areas were chosen as they will likely affect the end user the most. Overall, VConstruct takes approximately 3-5x less compute time (50-200x when trained) and up to 2x less memory (even more on longer datasets) to perform its calculations. This reduction in computational cost will save researchers time and money as well as let them perform work on high resolution images with less concessions in terms of area or time covered. These reductions in computational costs are also shown to get bigger on datasets larger then those tested.

Memory usage and time is measured in Table 5.1 by running both algorithms on a local machine with an Intel Core i5-9400 CPU, 8Gb of RAM, 8Gb of swap space, and GeForce RTX 2060 GPU.

Time is measured as wall clock time of the process from start to finish. Memory is measured as the maximum Resident Set Size plus Swap Space used by the process. Maximum memory is used as average memory usage does not matter if the process cannot be run. The total memory column for VConstruct in Table 5.1 is the maximum

Table 5.1: Computational costs of VConstruct and DINEOF. * indicates code was killed by Out of Memory killer before completion

	DINEOF		VConstruct				VConstruct Reconstruct			
Sample Size (TimexWidthxHeight)	**Time**	**Memory**	**Time**	**Memory**	**GPU**	**Total**	**Time**	**Memory**	**GPU**	**Total**
1040x1040x385 (Rolling Avg)	*	*	66min	4.8GB	5.2GB	7.9GB	56s	3.4GB	3GB	6.4GB
550x1040x385 (Rolling Avg)	1621min*	10.9GB*	34min	4.8GB	5.2GB	7.9GB	34s	3.4GB	3GB	6.4GB
1040x1040x385	*	*	41min	4.8GB	5.2GB	7.9GB	56s	3.4GB	3GB	6.4GB
550x1040x385	110min	11.9GB	20min	4.8GB	5.2GB	7.9GB	34s	3.4GB	3GB	6.4GB
250x1040x385	40min	5.1GB	6min	4.8GB	5.2GB	7.9GB	18s	3.4GB	3GB	6.4GB
1566x500x500	*	*	21min	4GB	4.1GB	6.8GB	33s	3.1GB	2.5GB	5.7GB
1000x500x500	*	*	15min	4GB	4.1GB	6.8GB	21s	3.1GB	2.5GB	5.7GB
500x500x500	23min	7.4GB	8min	4GB	4.1GB	6.8GB	15s	3.1GB	2.5GB	5.7GB
250x500x500	11min	5GB	4min	4GB	4.1GB	6.8GB	10s	3.1GB	2.5GB	5.7GB
1566x400x400	40min	12.3GB	16min	3.7GB	3.1GB	5.8GB	18s	3GB	1.7GB	4.6GB
1000x400x400	21min	9.4GB	10min	3.7GB	3.1GB	5.8GB	12s	3GB	1.7GB	4.6GB
500x400x400	14min	5.2GB	6min	3.7GB	3.1GB	5.8GB	9s	3GB	1.7GB	4.6GB
250x400x400	7min	4GB	4min	3.7GB	3.1GB	5.8GB	7s	3GB	1.7GB	4.6GB
1566x250x250	20min	5.7GB	6min	3.2GB	1.9GB	4.6GB	9s	2.8GB	1.4GB	4.2GB
1000x250x250	10min	4.7GB	4min	3.2GB	1.9GB	4.6GB	7s	2.8GB	1.4GB	4.2GB
500x250x250	6min	3GB	2min	3.2GB	1.9GB	4.6GB	5s	2.8GB	1.4GB	4.2GB
250x250x250	4min	1.9GB	1min	3.2GB	1.9GB	4.6GB	3s	2.8GB	1.4GB	4.2GB

combine main and GPU memory used simultaneously. VConstruct's implementation includes a step at the very end where it saves the network back to disk. This transfer is not included in the measurements as it is not a required part of the algorithm, and if needed could be performed in a way that would not increase the overall memory usage (by transferring the model in small chunks and saving them piece by piece). The same data as the Model Validation section for these measurements.

5.1 Computational Time

Time is the amount of time it takes to from when the respective algorithm starts running to when the results are saved to disk. With VConstruct the neural net can be pretrained and generate reconstructions on demand from this saved model. For this reason the amount of time it takes for VConstruct to generate a reconstruction without training is reported. DINEOF does not have a similar concept and the whole algorithm needs to be run every time.

From Table 5.1 it can be seen that the difference in computational time is signif-

icant. Comparing a full run of both algorithms shows a speed of roughly 3-5x with more significant speedups on larger datasets. When VConstruct is already trained speedups of 50-200x are seen.

To verify this trend will continue as data scales further, both theoretical and experimental time complexity will be examined. The brunt of work in VConstruct is calculating the weights of the neural network which each need a constant amount of work to compute. The number of weights corresponds directly to the number of pixels in the input image, so increasing the image size linearly increases the time complexity. This is backed up by the experimental results in Figure 5.1 part A as it shows a clear linear relation between image size and computation time. VConstruct scales similarly as the dataset grows in the temporal dimension; each additional time point adds a constant amount of work to calculate (one set of matrix calculations for the network and then one optimization iteration). This is backed up in Fig 5.1 part B.

The theoretical time complexity of DINEOF is hard to pin down as the main loop of the program is dependent on how fast the data converges and not the input size. This is evident when comparing the run time of 550x1040x385 (Rolling Average) to 550 x 1040 x 385 as this data is identical except for the rolling average and has a nearly 15x difference in run time. The rolling average test was killed before completion so it is potentially an even greater difference.

To simplify, the time complexity of SVD can be used as a lower bound since it is the bottleneck operation of each DINEOF loop. These SVD operations are performed until convergence, then an additional EOF mode is added, repeating the process until an optimal solution is found. The user can also set a maximum number of iterations or EOF modes to speed up run time at the cost of accuracy. Because of this using the time complexity of SVD will better generalize across different datasets and if the

number of iterations or EOF modes are limited by the user. Comparing to the time complexity of SVD will also generalize to other variants of DINEOF or EOF based methods that are designed to be faster, like Ping et al. [36]'s VE-DINEOF, as SVD (or similar operations) are integral to the EOF process and cannot be optimized away. This implementation of DINEOF uses truncated SVD which according to Li et al. [25] has time complexity of $O(2mn^2 + n^3 + n + mn)$ where m is image size and n is the number of time points. Some other EOF based methods use eigen decomposition instead of SVD, this is an equally computationally expensive operation with a cubic runtime.

Scaling image size increases SVD linearly making that the lower bound, experimental results show a much higher polynomial scaling. Increasing the number of images being processed scales the time of SVD cubically which is supported by the experimental results but not as clearly on the scale of this experiment.

The vast majority of VConstruct's run time comes from its training process. This is where it goes through the data and learns the appropriate weights for the network. These weights can be saved and reused on new data given it comes from the same geographical area. This vastly improves computation time, beating DINEOF by over 50-200x as seen in Table 5.1.

Overall VConstruct is considerably faster then DINEOF especially when it is already been trained. Given that VConstruct scales considerably better than DINEOF the performance gap will only continue to grow with VConstruct vastly outperforming DINEOF on large datasets.

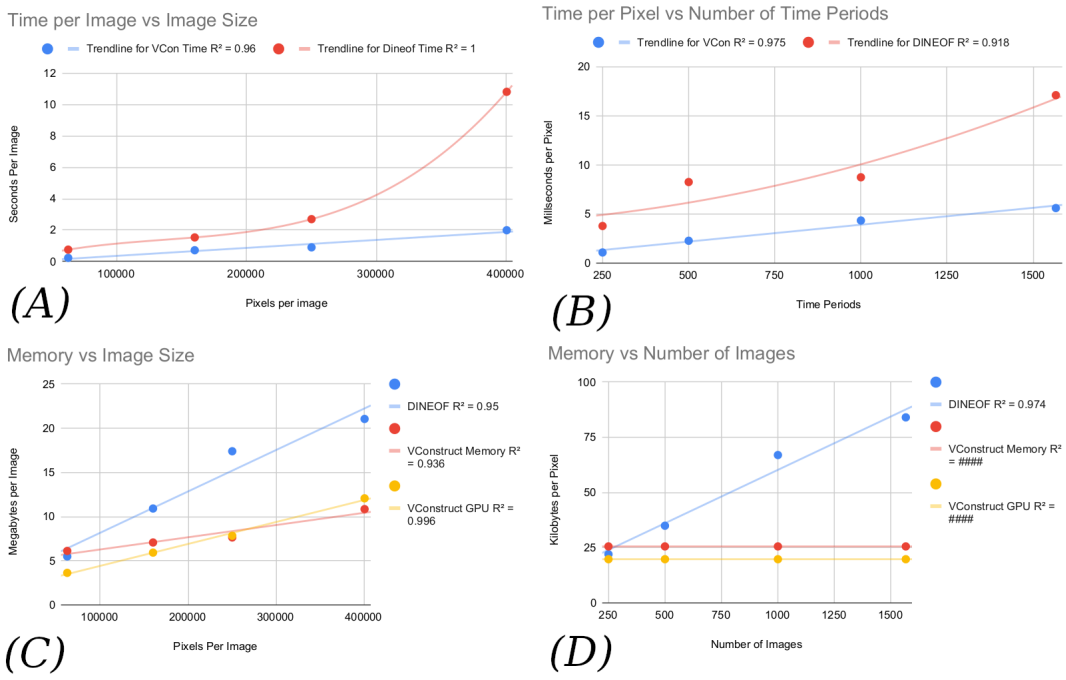

Figure 5.1: *(A)* Time Cost for Larger Datasets. DINEOF has a cubic trendline, VConstruct is Linear. *(B)* Time Cost for Longer Datasets. DINEOF has a quadratic trendline, VConstruct is Linear. *(C)* Memory Cost for Larger Datasets. All trendlines are linear. *(D)* Memory Cost for Longer Datasets. DINEOF's trendline is linear, VConstruct's are constant.

5.2 Spatial Memory

For the purposes of this paper, spatial memory is defined as the memory cost of increasing the spatial area or number of pixels being reconstructed. This is relevant as as the resolution of Earth Observation systems are increased, the limitations of current algorithms will require smaller areas to be reconstructed, losing valuable information. Similarly to time, both the total memory and the memory used only during reconstruction are reported, as these can differ for VConstruct. A significant amount of memory is used to calculate gradients during training VConstruct which can be skipped if the model is pre-trained. Additionally, the Encoder Network (Fig. 3.1) is not needed on a pre-trained model.

Overall VConstruct takes significantly less memory than DINEOF on all but the smallest datasets. To verify this scales, the information from Table 5.1 is used to calculate the amount of memory needed per image as they get larger, this is graphed in Fig. 5.1 part C.

In Fig. 5.1 part C both algorithms appear to scale linearly with image size. This matches up with the theoretical complexity of VConstruct, as image size increases, a proportional number of weights are added to the network, each with a constant amount of space needed. Similarly to time complexity DINEOF's space complexity is hard to pin down as it is somewhat dependant on the input data itself. $O(3n^2 + 3n + 2mn)$ is the space complexity of truncated SVD from Li et al. [25] with n being number of time points and m being image size. Using this as a lower bound as with time this backs up the linear growth of DINEOF.

Overall VConstruct uses less memory and scales just as well as DINEOF.

5.3 Temporal Memory

For the purposes of this paper, temporal memory is defined as the memory cost of increasing the length of time reconstructed. This is relevant as new data is being gathered every day and the most accurate results will be achieved by using the longest time period possible [16].

From Table 5.1 it can be seen that as the length of time being reconstructed increases, VConstruct's memory usage remains constant, whereas DINEOF's increases. This can be more clearly seen in Fig. 5.1 part D.

This is in line with the theoretical space complexity of VConstruct. Since images are provided to the network in constant size batches it does not matter how many there are, as they are not all needed at the same time. This makes VConstruct's

memory usage constant in regard to number of images. In Fig. 5.1 part D, DINEOF appears to scale roughly linearly with number of images. This does not match up with the polynomial $O(3n^2 + 3n + 2mn)$ complexity from Li et al. [25]. This could be due to perhaps a difference in implementation or just not having enough data-points to see the polynomial behaviour yet. On short datasets DINEOF uses less memory than VConstruct, but given VConstruct's constant scaling it performs exceptionally well on long datasets.

Chapter 6

Other Benefits of VConstruct

VConstruct has a significantly lower computational cost with accuracy similar to, but slightly less than DINEOF. Additionally, the VAE structure provides some other interesting benefits.

The first major benefit is an even further reduction of computational time. VAEs and deep learning in general benefit greatly from parallelization on GPUs, so VConstruct will be able to make great use of new GPU hardware as it comes out. GPU improvements have been a major area of development recently due to the boom in deep learning approaches. By comparison, DINEOF has no current parallel implementation and will likely not have one that provides a significant improvement in performance. Most of the DINEOF algorithm builds off previous parts and therefore must be done in series. While its SVD operation is parallelizable, it is unknown whether it would be enough to amortize the cost of using a GPU. VConstruct will overall make better use of new hardware. For example, VAEs are quite efficient on quantum annealing machines [21].

There is also potential for increased accuracy, potentially outperforming DINEOF. Barth et al. [3] introduces a DL based approach to Chl-a reconstruction focused on improving accuracy which beats DINEOF. They do not use a VAE but the approaches

are similar enough that some of their techniques could be applied here to improve accuracy.

The attribute vector, an input to the VAE, currently stores the representation of the cloudy image, but has greater potential. It is a condensed representation of a cloudy image used to condition the output of the Decoder network. However, this vector could be changed to represent other things, for example Yan et al. [44] use their attribute vector to represent things like hair colour when generating faces. If a similar technique were applied to VConstruct the attribute vector could encode oceanographic information like wind speed or sea surface temperature. This could allow for potential Chl-a images to be generated under certain conditions which could be useful for modelling.

The nature of conditional VAEs make them easy to enhance with other data sources as well. The attribute vector does not need to come from the cloudy Chl-a image; another data source could be used to generate the attribute vector or to augment the existing attribute vector. An interesting example of this is SAR data coming from Sentinel-1 or other satellites, which has been shown to have a relation to Chl-a [6]. Given that SAR penetrates through clouds this is a prime data source to assist with reconstruction.

One final area VConstruct might provide benefits is in detecting anomalous data in time series satellite data. VAEs have previously been very successful in detecting avalanche deposits from satellite data in the past [38]. Using a similar technique, harmful algae blooms could be detected in an automated fashion. In the training process VConstruct learns a distribution of images and then assigns a score that represents how well it did in reconstructing that image. First VConstruct is trained on a set of images that contain no harmful algae blooms and then a reconstruction of an image with a suspected HAB is attempted. After, the same loss score as used in

training can be calculated, if this score is good it means the image is well represented by the learnt distribution and likely not anomalous. If the score is bad it means the image is not well represented by the distribution and is likely anomalous. See section 4.4 for an example.

Chapter 7

Discussion

In addressing some limitations of this work, the implementations of the 2 algorithms are discussed. The DINEOF algorithm was written in Julia and uses an official implementation from GHER [1] and VConstruct is written using Python and Pytorch. Both Julia and PyTorch use C/C++ in their backends so should be comparable in terms of time and space. The parts of code written in pure Python will, in theory, be slower than Julia, but this does not seem to make any difference in the results.

VConstruct uses GPU memory primarily, which is not as available in as high amounts as RAM. If VConstruct's GPU memory requirements exceed the amount that is available it can be set to use RAM only. This mode will use slightly less RAM than the amount listed in the "Total" column of Table 5.1 (slightly less as there is no overhead for GPU setup and data transferring). This method still uses less RAM then DINEOF on large datasets but will be considerably slower than the GPU version. Additionally given that VConstruct is parallel it can be run across multiple GPUs or even a cluster of computers each with multiple GPUs, greatly increasing the amount of GPU memory available. There is also some degree of tuning the user can do to trade speed and memory. Currently the algorithm is tuned to be balanced but could,

[1] http://modb.oce.ulg.ac.be/mediawiki/index.php/DINEOF

if desired, be tuned to trade speed for memory or vice versa.

Of the several geographic areas tested, no significant difference in performance compared to DINEOF is seen, besides the handling of anomalous days that have no representative training data. Anomalous days are by definition rare so should not impact reconstruction accuracy significantly. From this it can be assumed that VConstruct is generalizable across different geographic areas, assuming the model is trained on that area.

Both DINEOF and VConstruct are likely equally resistant to climate change, as they use the same dataset as an input and have similar internal processes [37]. Other models including those of Barth et al. [3], Park et al. [35], and Konik et al. [23] use relations between various oceanographic datasets in their predictions. These relationships could break down due to climate change, potentially reducing the accuracy of these models or rendering them unusable.

For the purposes of reducing computational cost as much as possible VConstruct does not use any temporal data as input, and only reconstructs based off spatial information. This reduces the computational cost significantly as the temporal information can be provided beforehand. This comes with the downside of not being able to reconstruct days with no spatial information. DINEOF is able to reconstruct days with no spatial information but the official documentation from GHER recommends not doing this, as it can produce unreliable reconstructions [2].

DINEOF can also be used to reconstruct other types of oceanographic data like sea surface temperature [12]. Though VConstruct is not tested on other types of datasets it should generalize well given sufficient training data. The models of both Yan et al. [44] and Ivanov et al. [19], which VConstruct is based on, perform well on multiple complex datasets including faces, birds, and MNIST numbers. From this it follows

[2]http://modb.oce.ulg.ac.be/mediawiki/index.php/DINEOF

that VConstruct should do well on other datasets. An interesting dataset to test on would be one with a slower rate of change like the Normalized Difference Vegetation Index (NVDI) which captures details about vegetation growth. VConstruct needs to map the dataset to a smaller range in preprocessing to function well, this mapping is more effective with less anomalous values. A dataset with a slower rate of change like NVDI should have less anomalous values, potentially giving better accuracy. The pipeline is fairly generalizable so all that is needed is to download the dataset and then update the preprocessing scripts to the specifics of that dataset.

Although DINEOF is the most widely used algorithm for reconstructing satellite data, the DINCAE approach from Barth et al. [3] is the current state of the art. DINCAE significantly outperforms DINEOF and, since DINEOF has better accuracy then VConstruct, likely VConstruct as well in accuracy. Barth et al. [3] do not explore the efficiency of their method as thoroughly as this work but provide enough for some comparisons to be made. Barth et al. [3] report a 3.5x speed up when compared to DINEOF, which is comparable to the lower end of speedups achieved by VConstruct. Their tests were performed on a more powerful GPU than tests in this work were, so it is likely VConstruct has the better speed. They do not report memory usage but the input required to their network is 8x the size of the input to VConstruct due to temporal data and other required inputs. This factor suggests VConstruct is more efficient in terms of memory. The DINCAE model could not be successfully be integrated into this work's testing and data processing pipeline, so a more precise comparison could not be provided, this is an area of future work. DINCAE's architecture does provide it some of VConstruct's benefits outlined in Section. 6 but not all. DINCAE is parallel so can utilize new hardware and can utilize additional data sources. DINCAE cannot perform the same anomaly detection or image generation under other conditions that VConstruct can without

major changes. VConstruct is more resistant to climate change as it does not depend on the correlation between satellite-derived Chl-a and other ocean variables [37].

The likely best solution for Chl-a reconstruction is an ensemble based approach, incorporating the best of each method. There is a few ways this could be accomplished, the first being to use DINEOF and VConstruct together. VConstruct could be used to quickly generate a preview of the reconstruction, and then, if more detail is needed, DINEOF could be applied. Alternatively they could be used in series, VConstruct could be used as a correction layer for DINEOF as it could be conditioned such that DINEOF's reconstructions are used in place of the cloudy image and the true full image is kept as the desired ground truth. This approach would be more computationally expensive but could result in improved accuracy. A complex but likely incredibly effective ensemble based approach would be to add a step that analyzes the incoming dataset and applies the most appropriate method for reconstruction. For example if the area was polar Park et al. [35]'s method could be applied, if it was large or had low variation VConstruct would work, DINEOF could be applied to smaller areas that do not have another approach or if high accuracy was required.

Besides the obvious decrease in cost there are a few other practical use cases for VConstruct's speed. The first being as an exploratory tool, both DINEOF and VConstruct require little tuning as they do not require *a priori* knowledge about the dataset and their base settings do quite well with regards to accuracy. However some changes to preprocessing or to the settings the algorithms do have may provide better accuracy (i.e removing certain months with low confidence data, changes to the atmospheric correction algorithm, etc). The high cost of running DINEOF makes it prohibitive to explore these options, where VConstruct can as each individual run is much cheaper. The other case is in a tool like Algae Explorer[3] which provides

[3]https://algaeexplorer.ca/

daily images of algae in the Salish Sea area. If Algae Explorer wanted to provide a reconstructed version of their data this would not be possible in real time with DINEOF as the algorithm would need to run over every image every day. However VConstruct makes this possible since a model can be generated for the area Algae Explorer covers and then reused for each new image, reducing run time significantly.

VConstruct is stochastic as its reconstructions rely on a random sample from the latent space to generate. Because of this, reconstructing the same image multiple times may provide different results. This may be desirable as there may be multiple possible ways to fill the gaps and sampling the latent space enough times would allow a wide variety to be generated. From these the best reconstruction could be selected based on whatever criteria the user wishes or an error map could be generated. However if determinism is required it could be achieved by saving the latent variable used to reconstruct the image and reusing whenever that image is reconstructed. This would require more space but not more RAM as only one latent variable would need to be saved in RAM at once.

A common problem in machine learning is its black box nature [27]. Lundberg and Lee [27] provide significant steps forward with machine learning interpretability, however this is primarily focused on classification tasks and not generative ones. VConstruct is somewhat of a black box due to this. The attribute vector and vector from the latent space are not inherently human understandable as the meaning of each value and their relations are chosen to best reduce the input data to the latent space. Because of this, it is tough to say exactly why VConstruct made the reconstruction the way it did. Some work has been done in "disentangling" the latent variable into human understandable values, but investigating this was outside the scope of this work and would be an interesting area of future work [7].

Chapter 8

Conclusion

A reconstruction method for Chl-a data based off a Variational Autoencoder is introduced and shown to provide similar, but slightly less, accuracy then the widely used DINEOF. Furthermore, VConstruct significantly outperforms DINEOF in computational costs, taking 3-5x less time (50-200x less time if the network has already been trained) than DINEOF. It is also shown that the memory cost is less than and scales better or just as well as DINEOF. This is quite useful as this is the era of "Big Earth Data" and massive amounts of EO data are being generated daily [14]. Lastly, some other benefits of VConstruct that could be explored in future work are outlined, including generating images under certain conditions or anomaly detection.

Bibliography

[1] S. Alvain, C. Moulin, Y. Dandonneau, and F. M. Bréon, "Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery," *Deep Sea Research Part I: Oceanographic Research Papers*, vol. 52, no. 11, pp. 1989–2004, Nov. 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0967063705001536

[2] N. Anantrasirichai, J. Biggs, F. Albino, P. Hill, and D. Bull, "Application of machine learning to classification of volcanic deformation in routinely generated insar data," *Journal of Geophysical Research: Solid Earth*, vol. 123, no. 8, pp. 6592–6606, 2018. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JB015911

[3] A. Barth, A. Alvera-Azcárate, M. Licer, and J.-M. Beckers, "DINCAE 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations," *Geoscientific Model Development*, vol. 13, no. 3, pp. 1609–1622, Mar. 2020. [Online]. Available: https://gmd.copernicus.org/articles/13/1609/2020/

[4] J. M. Beckers and M. Rixen, "Eof calculations and data filling from incomplete oceanographic datasets," *Journal of Atmospheric and Oceanic Technology*, vol. 20, no. 12, pp. 1839 – 1856, 2003. [Online].

Available: https://journals.ametsoc.org/view/journals/atot/20/12/1520-0426_2003_020_1839_ecadff_2_0_co_2.xml

[5] S. Bojinski and M. M. Verstraete, "(PDF) The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy," *ResearchGate*, 2014. [Online]. Available: https://www.researchgate.net/publication/271271716_The_Concept_of_Essential_Climate_Variables_in_Support_of_Climate_Research_Applications_and_Policy

[6] M. Bresciani, M. Adamo, G. De Carolis, E. Matta, G. Pasquariello, D. Vaičiūtė, and C. Giardino, "Monitoring blooms and surface accumulation of cyanobacteria in the curonian lagoon by combining meris and asar data," *Remote Sensing of Environment*, vol. 146, pp. 124–135, 2014, liege Colloquium Special Issue: Remote sensing of ocean colour, temperature and salinity. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425713003283

[7] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, "Understanding disentangling in β-vae," 2018.

[8] CEOS, "Committee on earth observation satellites (ceos) database." [Online]. Available: http://database.eohandbook.com

[9] S. Chen, C. Hu, B. B. Barnes, Y. Xie, G. Lin, and Z. Qiu, "Improving ocean color data coverage through machine learning," *Remote Sensing of Environment*, vol. 222, pp. 286–302, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425718305753

[10] C. Doersch, "Tutorial on variational autoencoders," 2021.

[11] ESA, "Sentinel-3 olci coverage." [Online]. Available: https://dragon3.esa.int/web/sentinel/user-guides/sentinel-3-olci/coverage

[12] U. Ganzedo, A. Alvera-Azcárate, G. Esnaola, A. Ezcurra, and J. Sáenz, "Reconstruction of sea surface temperature by means of dineof: a case study during the fishing season in the bay of biscay," *International Journal of Remote Sensing*, vol. 32, no. 4, pp. 933–950, 2011. [Online]. Available: https://doi.org/10.1080/01431160903491420

[13] F. Giannini, B. P. Hunt, D. Jacoby, and M. Costa, "Performance of olci sentinel-3a satellite in the northeast pacific coastal waters," *Remote Sensing of Environment*, vol. 256, p. 112317, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425721000353

[14] H. Guo, "Big earth data: A new frontier in earth and information sciences," *Big Earth Data*, vol. 1, no. 1-2, pp. 4–20, 2017. [Online]. Available: https://doi.org/10.1080/20964471.2017.1403062

[15] Z. Han, Y. He, G. Liu, and W. Perrie, "Application of DINCAE to Reconstruct the Gaps in Chlorophyll-a Satellite Observations in the South China Sea and West Philippine Sea," *Remote Sensing*, vol. 12, no. 3, p. 480, Jan. 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/3/480

[16] A. Hilborn and M. Costa, "Applications of DINEOF to Satellite-Derived Chlorophyll-a from a Productive Coastal Region," *Remote Sensing*, vol. 10, no. 9, p. 1449, Sep. 2018. [Online]. Available: https://www.mdpi.com/2072-4292/10/9/1449

[17] Y. Huot, M. Babin, F. Bruyant, C. Grob, M. S. Twardowski, and H. Claustre, "Does chlorophyll a provide the best index of phytoplankton biomass for primary productivity studies?" *Biogeosciences Discussions*, vol. 4, no. 2, pp. 707–745, Mar. 2007. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00330232

[18] P. Imperatore and D. Riccio, "Geoscience and remote sensing: New achievements." IntechOpen, 2010, ch. 7. [Online]. Available: https://books.google.ca/books?id=C5uUDwAAQBAJ

[19] O. Ivanov, M. Figurnov, and D. Vetrov, "Variational autoencoder with arbitrary conditioning," in *International Conference on Learning Representations*, 2019.

[20] D. Jin, E. Lee, K. Kwon, and T. Kim, "A deep learning model using satellite ocean color and hydrodynamic model to estimate chlorophyll-a concentration," *Remote Sensing*, vol. 13, no. 10, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/10/2003

[21] A. Khoshaman, W. Vinci, B. Denis, E. Andriyash, H. Sadeghi, and M. H. Amin, "Quantum variational autoencoder."

[22] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv:1312.6114 [cs, stat]*, May 2014, arXiv: 1312.6114. [Online]. Available: http://arxiv.org/abs/1312.6114

[23] M. Konik, M. Kowalewski, K. Bradtke, and M. Darecki, "The operational method of filling information gaps in satellite imagery using numerical models," *International Journal of Applied Earth Observation and Geoinformation*, vol. 75, pp. 68–82, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0303243417302805

[24] D. Kotta and D. Kitsiou, "Chlorophyll in the eastern mediterranean sea: Correlations with environmental factors and trends," *Environments*, vol. 6, no. 8, 2019. [Online]. Available: https://www.mdpi.com/2076-3298/6/8/98

[25] X. Li, S. Wang, and Y. Cai, "Tutorial: Complexity analysis of singular value decomposition and its variants," 2019.

[26] X. Liu and M. Wang, "Filling the gaps of missing data in the merged viirs snpp/noaa-20 ocean color product using the dineof method," *Remote Sensing*, vol. 11, no. 2, 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/2/178

[27] S. Lundberg and S. Lee, "A unified approach to interpreting model predictions," *CoRR*, vol. abs/1705.07874, 2017. [Online]. Available: http://arxiv.org/abs/1705.07874

[28] C. Marchese, C. Albouy, J.-E. Tremblay, D. Dumont, F. D'Ortenzio, S. Vissault, and S. Bélanger, "Changes in phytoplankton bloom phenology over the north water (now) polynya: a response to changing environmental conditions," *Polar Biology*, vol. 40, 09 2017.

[29] C. Marchese, L. Castro de la Guardia, P. G. Myers, and S. Bélanger, "Regional differences and inter-annual variability in the timing of surface phytoplankton blooms in the labrador sea," *Ecological Indicators*, vol. 96, pp. 81–90, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1470160X18306538

[30] S. Maritorena, D. A. Siegel, and A. R. Peterson, "Optimization of a semianalytical ocean color model for global-scale applications," *Appl. Opt.*, vol. 41, no. 15, pp. 2705–2714, May 2002. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-41-15-2705

[31] S. Maritorena, O. H. F. d'Andon, A. Mangin, and D. A. Siegel, "Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues," *Remote Sensing of Environment*, vol. 114, no. 8, pp. 1791–1804, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425710001136

[32] N. Mayot, F. D'Ortenzio, M. Ribera d'Alcalà, H. Lavigne, and H. Claustre, "Interannual variability of the mediterranean trophic regimes from ocean color satellites," *Biogeosciences*, vol. 13, no. 6, pp. 1901–1917, 2016. [Online]. Available: https://bg.copernicus.org/articles/13/1901/2016/

[33] S. K. Moore, V. L. Trainer, N. J. Mantua, M. S. Parker, E. A. Laws, L. C. Backer, and L. E. Fleming, "Impacts of climate variability and future climate change on harmful algal blooms and human health," in *Environmental Health*, vol. 7, no. S2. Springer, 2008, p. S4.

[34] O. Mudele, F. M. Bayer, L. F. R. Zanandrez, A. E. Eiras, and P. Gamba, "Modeling the temporal population distribution of *ae. aegypti* mosquito using big earth observation data," *IEEE Access*, vol. 8, pp. 14 182–14 194, 2020.

[35] J. Park, J.-H. Kim, H.-c. Kim, B.-K. Kim, D. Bae, Y.-H. Jo, N. Jo, and S. H. Lee, "Reconstruction of Ocean Color Data Using Machine Learning Techniques in Polar Regions: Focusing on Off Cape Hallett, Ross Sea," *Remote Sensing*, vol. 11, no. 11, p. 1366, Jan. 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/11/1366

[36] B. Ping, F. Su, and Y. Meng, "An improved dineof algorithm for filling missing values in spatio-temporal sea surface temperature data," *PLOS ONE*, vol. 11, p. e0155928, 05 2016.

[37] S. Sathyendranath, R. J. Brewin, T. Jackson, F. Mélin, and T. Platt, "Ocean-colour products for climate-change studies: What are their ideal characteristics?" *Remote Sensing of Environment*, vol. 203, pp. 125–138, 2017, earth Observation of Essential Climate Variables. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425717301670

[38] S. Sinha, S. Giffard-Roisin, F. Karbou, M. Deschatres, A. Karas, N. Eckert, C. Coléou, and C. Monteleoni, "Variational autoencoder anomaly-detection of avalanche deposits in satellite sar imagery," in *Proceedings of the 10th International Conference on Climate Informatics*, ser. CI2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 113–119. [Online]. Available: https://doi.org/10.1145/3429309.3429326

[39] D. Sirjacobs, A. Alvera-Azcárate, A. Barth, G. Lacroix, Y. Park, B. Nechad, K. Ruddick, and J.-M. Beckers, "Cloud filling of ocean colour and sea surface temperature remote sensing products over the Southern North Sea by the Data Interpolating Empirical Orthogonal Functions methodology," *Journal of Sea Research*, vol. 65, no. 1, pp. 114–130, Jan. 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1385110110001036

[40] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28. Curran Associates, Inc., 2015. [Online]. Available: https://proceedings.neurips.cc/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf

[41] K. D. Suchy, N. Le Baron, A. Hilborn, R. I. Perry, and M. Costa, "Influence of environmental drivers on spatio-temporal dynamics of satellite-derived chlorophyll a in the strait of georgia," *Progress in Oceanography*, vol. 176, p. 102134, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0079661118303136

[42] M. H. Taylor, M. Losch, M. Wenzel, and J. Schröter, "On the Sensitivity of Field Reconstruction and Prediction Using Empirical Orthogonal Functions Derived from Gappy Data," *Journal of Climate*, vol. 26, no. 22, pp. 9194–9205,

Nov. 2013. [Online]. Available: https://journals.ametsoc.org/jcli/article/26/22/9194/34073/On-the-Sensitivity-of-Field-Reconstruction-and

[43] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Aug 2018. [Online]. Available: https://doi.org/10.1007/s13244-018-0639-9

[44] X. Yan, J. Yang, K. Sohn, and H. Lee, "Attribute2Image: Conditional Image Generation from Visual Attributes," *arXiv:1512.00570 [cs]*, Oct. 2016, arXiv: 1512.00570. [Online]. Available: http://arxiv.org/abs/1512.00570

[45] X. Yao, G. Li, J. Xia, J. Ben, Q. Cao, L. Zhao, Y. Ma, L. Zhang, and D. Zhu, "Enabling the big earth observation data via cloud computing and dggs: Opportunities and challenges," *Remote Sensing*, vol. 12, no. 1, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/1/62

Appendix A

Reconstructed Figures

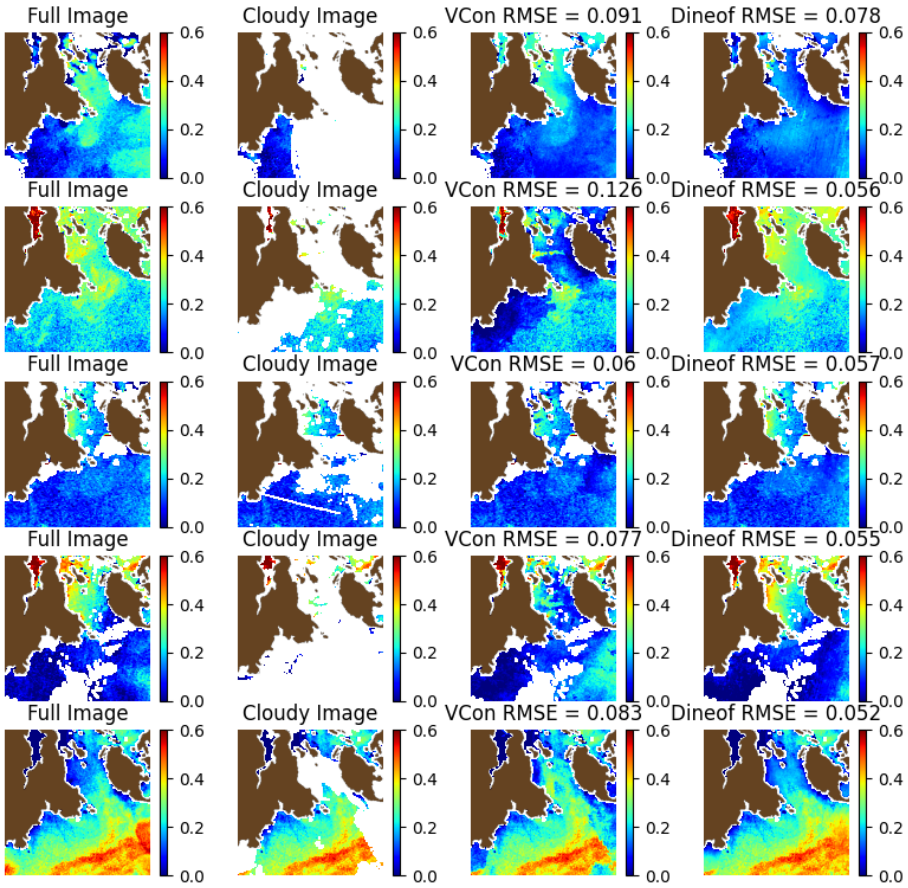

Figure A.1: Reconstructed testing images for the Victoria Coast

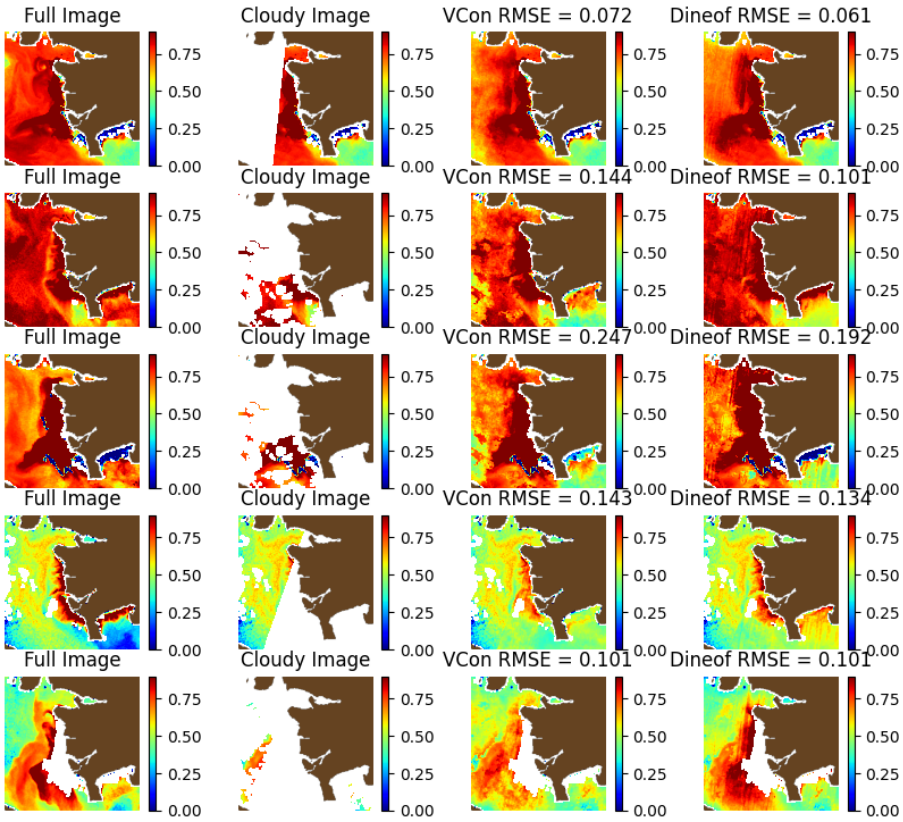

Figure A.2: Reconstructed testing images for the Fraser River Mouth