UTRECHT UNIVERSITY

Department of Information and Computing Science

Applied Data Science master thesis

To what extent can we accurately predict water depth from overhead images using linear and non-linear methods, in a scaled laboratory estuary environment

First examiner:
Maarten Kleinhans

Candidate:
David Leahy

Second examiner:
Madlene Nussbaum

July 3, 2024

Acknowledgments

I would like to express my deepest gratitude to those who have supported me throughout the completion of this thesis.

First and foremost, I am profoundly grateful to my main supervisor, Dr. Maarten Kleinhans, for his invaluable guidance, expertise, and unwavering support. His insightful feedback and encouragement were instrumental in shaping this research.

I would also like to extend my sincere thanks to my second examiner, Dr. Madlene Nussbaum, for her thoughtful critiques and valuable suggestions that greatly improved the quality of this work.

Special thanks to Eise Nota for his assistance and willingness to answer my numerous questions. His support and cooperation were essential throughout this research process.

Thank you all for your support and contributions.

Abstract

Physical scale experiments improve our comprehension of fluvial, tidal, and coastal processes. However, acquiring accurate, precise, and continuous data on water depth has been difficult due to the limitations inherent in the measuring equipment.

In this paper, we predicted water depth in a two-fold process using different models, then tested the accuracy of these models. We used data from a physical scale estuary experiment using overhead imagery and Digital Elevation Models.

Analogous to real-world estuaries, much of the terrain is above water. This presented an issue in creating a water depth prediction model, as over 57% of the training data pixels were calculated to contain no water. First, we address the issue of zero-inflated data. We developed a binary classification model that predicts pixels with no water and with water, then use regression to predict the depth of pixels with water.

Binary classification models such as Logistic Regression, Random Forests, and Support Vector Machine Learning (SVM) were compared. These models were trained using overhead imagery from water depth calibration experiments. The overhead imagery was overlaid on water depth maps. The water depth maps were calculated from the calibration experiments using the (DEM) and increasing weir heights.

In the second step, regression models were developed and compared. Linear regression, Random Forests, and SVM models were trained on the pixels containing water in the water depth maps. The two-step Random Forest model was selected and applied to different overhead images under similar experimental conditions, the resulting predicted water depth maps produced promising results. The developed model was precise enough to detect geomorphological change by comparing overhead imagery 990 cycles apart.

The implication of the experimental data–model integration is that future experiments can derive water depth from overhead imagery in a simple, affordable, and labour-efficient manner.

Contents

1. Introduction

Estuaries are coastal ecosystems formed where rivers meet the sea. Influenced by tidal movements, estuaries show dynamic patterns of bars and channels. Understanding fluvial, tidal, and coastal processes is vital for managing these unique aquatic environments.

Physical scale experiments greatly enhance our understanding of estuarine processes, these models are sometimes referred to as *flumes*. Using physical scale models offers two primary advantages. Firstly, experiments provide precise control over initial and boundary conditions, facilitating rapid simulation of entire systems. This stands in stark contrast to the slow and dynamically changing conditions observed in field settings (where, for instance, a natural tidal cycle spans approximately 12 hours and 25 minutes apart). Secondly, physical models use real materials with their inherent laws and properties, in contrast to numerical models that rely on numerous parameters and approximations to simulate phenomena such as water flow and sediment transport.

However, acquiring accurate, precise, and continuous data on water depth on physical scale experiments has been a persistent challenge due to the limitations of current measuring equipment. This is because, at limited water depths, conducting flow measurements using submerged instruments without disturbing sediment poses significant technical challenges. This lack of reliable data hinders our ability to model and predict these complex processes accurately (Weisscher et al., 2020).

The focus of this study is to predict the water depths from overhead imagery. Focusing on the small depth of water of the physical experiment. In this paper, we will review how other researchers in the field have estimated water depth from similar experiments using overhead imagery and digital elevation models from laser scanning or stereo photography as described by (Weisscher et al., 2020; Leduc et al., 2019; Tal & Paola, 2010a; Upson, 2024)

In the world of remote sensing, simple linear models exist for extracting water depth from multispectral satellite imagery ((Geyman & Maloof, 2019; Lyzenga, 1978; Stumpf et al., 2003). These models are straightforward and can be accurate

in shallow depths. However, these models were designed to predict from satellite imagery (up to 13 bands for Sentinel-2 imagery) not on Red, Green, and Blue (RGB) channels and these linear algorithms fail to account for the non-linearity in empirical color-to-depth relationships, which are influenced by varying bottom types and greater depths. Part of this research is to determine despite these limitations if linear models are suitable at predicting water depth in laboratory scale experiments using overhead RGB imagery.

Machine learning models have been developed to address the issue of non-linearity in water depth prediction of coastlines, rivers and lakes. The typical approach is to train machine learning models on an area where satellite imagery is available and also detailed bathymetric information from LiDAR campaigns. The machine learning models that have proved more effective than linear prediction include: Support Vector Machine Learning (SVM) (Vojinovic et al., 2013; Mateo-Pérez et al., 2021; Yustisi Lumban-Gaol & Peters, 2022), Random Forests (Disanayaka Mudiyanselage et al., 2022; Zhang et al., 2020; Islam et al., 2023) and Convolutional Neural Networks (Ai et al., 2020; He et al., 2023) . These performance of these models vary with the accuracy of training data, difficulties in particular water depths and when water conditions differ. However, they show promise to predict water depth more accurately than a simple linear regression method. But have not been applied to laboratory scale experiments.

The purpose of this study is to investigate the possibility of accurately predicting water depth using overhead imagery, in physical scale estuarial experiments by comparing the performance of linear and non-linear models. We propose a novel approach that leverages binary classification models (Logistic Regression, Random Forests, and SVM) to classify if pixels contain water or no water based on the pixels RGB value. These models were trained using overhead imagery overlaid on water depth maps. The water depth maps were calculated using calibration data from Digital Elevation Models (DEM) weir height data.

In the second step of our approach, we develop regression models to predict water depths. Linear regression, Random Forests, and SVM models were trained on the pixels containing water and their performances were compared. Our results indicate that the two-step Random Forest model produces promising results, closely resembling the overhead imagery and providing accurate water depth predictions with a mean absolute error of 3.87 mm.

This research is significant because identifying accurate models to predict water depth from overhead imagery would not only expand the possibilities for analysing temporal changes in water depth but also revolutionize the monitoring of through-water topography at high resolution and frequency. This advancement allows for continuous observation of experimental topography without the need to stop the flow and drain the experiment for topographic scans, enabling the natural evolution of experimental conditions.

2. Literature Review

Below, we review the data collection and post-processing methods that are used to quantify water depth change and morphological changes in tidal flume experiments. In conjunction with their drawbacks and achievable level of accuracy.

2.1 Techniques for measuring bed elevation

Bed elevation can be measured in a number of ways, including point gauge surveys Peakall et al. (2007), scanning (M. G. Kleinhans et al., 2015; Van Dijk et al., 2014), ultrasonic echo sounding (Hoyal & Sheets, 2009; Stefanon et al., 2012), and structure-from-motion (SfM) photogrammetry (Weisscher et al., 2020; Leduc et al., 2019; Leuven et al., 2018), using georeferenced ground control points.

The point gauge survey technique was found to produce the most accurate, achieving precision up to ($\pm$0.1 mm) Peakall et al. (2007), and works with an undrained bed, but it has the disadvantage of being slow to measure. Scanning and ultrasonic echo can achieve accuracies of ($\pm$0.5 to 1 mm) and require a drained bed.

This is not ideal as bed topography may be disrupted during the draining and filling of the flume. Structure-from-motion (SfM) photogrammetry has been adopted in recent measurements of bed elevation, achieving accurate results within 1mm, this method also requires a drained bed (Weisscher et al., 2020; Leduc et al., 2019; Leuven et al., 2018).

2.2 Techniques for colour extraction of overhead imagery

To extract water depth from colour saturation, it can be beneficial to first transform the overhead imagery into a different colour space to augment the differences in colour saturation. An effective technique employed in literature, is to convert the RGB overhead imagery to the CIELAB L*a*b* colour space . Where L* is a scale for luminosity, a* is a scale from green to red, and b* is a scale from blue to yellow (Tal & Paola, 2010a; Leuven & Kleinhans, 2019).

Other colour transformations are possible. such as the HSV (Hue, Saturation

and Value) colour space which has been employed in RGB imagery of water, as performed by Champion et al. (2017). Both techniques can be effective at removing noise such as shadows and improve water depth prediction.

2.3 Techniques for water depth measurement in flumes

Water depth prediction can be derived from the dyed water colour saturation or from measured water levels. Previous studies have tried to extract water depth from colour saturation of dyed water (Upson, 2024; Leuven et al., 2018; Carbonneau et al., 2006; Tal & Paola, 2010b). The method involved using the correlation between water depth with the blueness derived from overhead imagery.

The method used by Leuven et al. (2018) was to fully submerge the sediment bed and capture overhead images. Following this, a Digital Elevation Model (DEM) was generated after the bed was drained slowly. The overhead images were converted to the CIELAB colour space and the B-band was extracted. A predictive relationship for water depth as a function of the B-band was developed. The results yielded a correlation between the B-band and water depth to approximately 1–5 mm, the MAE was not reported. So far the methods to predict water depth in flume experiments have only used linear regression methods.

An alternative technique to prediction water depth from overhead imagery is employing instrumentation such as an ultrasonic echo sounder, as performed by Tambroni et al. (2005) but this requires specialised instrumentation or using SfM photogrammetry to derive water depth, performed by Leduc et al. (2019). This technique was was accurate up 1mm but this system was implemented for unidirectional flow flume and a considerable disadvantage of requiring long processing times (5 hours for one image).

2.4 Machine learning models for water depth prediction

Machine learning models have been developed to address the challenge of non-linear water depth prediction applied to remote sensing data. This paper focuses on two machine learning methods, Random Forest and SVM.

In remote sensing, these models predict water depth by combining satellite imagery with water depth estimations, derived from a method that involves subtracting the surface water level from the bathymetric data obtained through LiDAR

campaigns. The machine learning models consistently outperform linear regression in water depth prediction (Vojinovic et al., 2013; Mateo-Pérez et al., 2021; Disanayaka Mudiyanselage et al., 2022; Zhang et al., 2020; Ai et al., 2020; He et al., 2023).

2.4.0.1 Random Forest

As described by Islam et al. (2023), a Random Forest is an ensemble learning method used for classification and regression tasks. A Random Forest is simply a collection of decision trees, each trained on a random subset of the data. A decision tree is a model that makes predictions by recursively splitting the data into subsets based on the value of input features, leading to a tree-like structure of decisions. In a Random Forest, the final prediction is made by averaging the predictions of individual trees (in regression) or by taking a majority vote (in classification). One can adjust the number of trees in the forest to balance between bias and variance, thereby optimizing performance .

2.4.1 Support Vector Machine Learning

Support Vector Machine (SVM) learning is a supervised machine learning algorithm used for classification and regression tasks. It works by finding the optimal hyperplane that best separates data points of different classes in a high-dimensional space, maximizing the margin between the closest points of the classes, known as support vectors. When data is not linearly separable, SVM can use kernel functions to transform the data into a higher-dimensional space where a linear separator can be found. A kernel function computes the dot product of data points in this higher-dimensional space without explicitly performing the transformation. Common kernel functions include linear, polynomial, and radial basis function (RBF). The performance of an SVM can be tuned by adjusting parameters such as the regularization parameter (C) and the choice of kernel function, As described by Mateo-Pérez et al. (2021).

2.4.2 Logistic Regression

Logistic regression is a statistical method used for binary classification problems, where the output represents probabilities that must lie between 0 and 1. By applying thresholds to these probabilities, the outcomes are classified as 0 or 1.

Logistic regression utilizes the sigmoid function, which is fitted using training

data. Once trained, the model converts inputs into probabilities. By applying a threshold to these probabilities (commonly 0.5), the model classifies the output as either 0 or 1.

2.5 Research question

Recent studies have established a correlation between the intensity of dyed water in a flume environment and the color saturation observed in overhead imagery (Leduc et al., 2019). Simultaneously, while linear models have demonstrated the capability to predict water depth from satellite imagery in remote sensing applications, non-linear models consistently achieve more accurate predictions.

Given these advancements, our research question emerges: *To what extent can we accurately predict water depth from overhead images using linear and non-linear methods?*

3. Data

3.1 Design of the Metronome facility

The Metronome is a flume set up in a laboratory designed to simulate tidal systems. It is unique among flumes worldwide as it is designed to periodically tilt, analogous to the ebb and flow of the tides. The system is permanently located at the Earth Simulation Laboratory of Utrecht University (UU).

The Metronome tilts periodically to simulate tidal movements that transport sediment and shape tidal morphologies. Its tilting axis is strategically positioned directly beneath the steel floor to minimize longitudinal motion. Four actuators operate in pairs to tilt the flume, with motion mirrored across the tilting axis. The Metronome is divided into riverside and seaward sections, as illustrated in Figure 3.1.

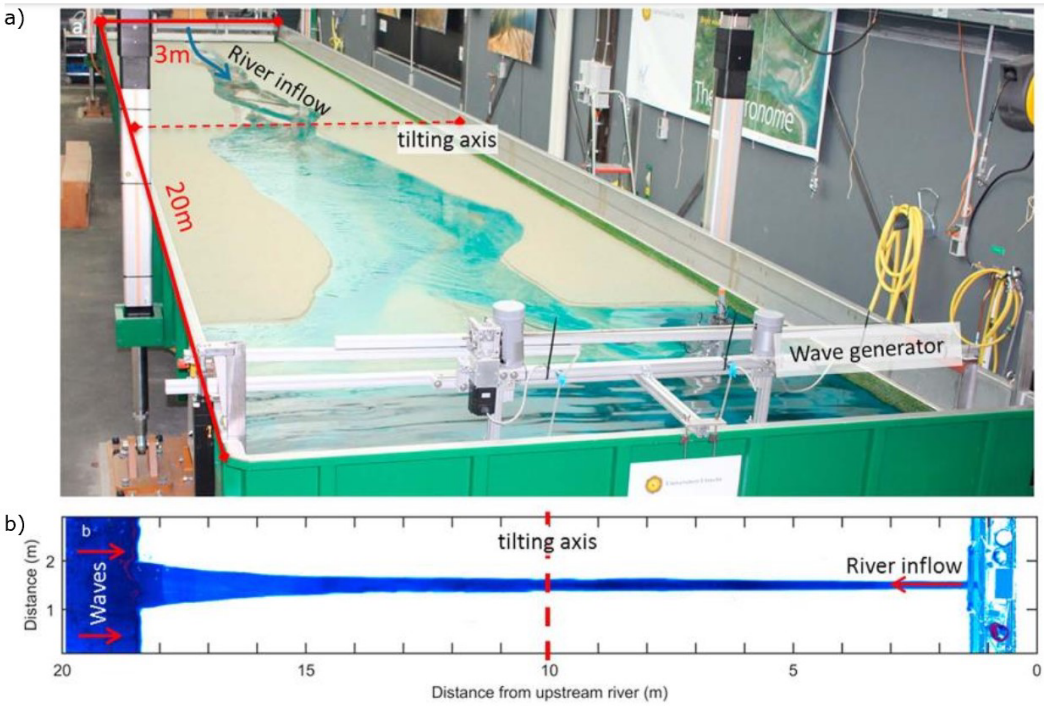


Figure 3.1: a) Example of a typical Metronome experiment. (b) Overhead image (with a blueness map applied for visualisation) of the experimental set-up (Leuven et al., 2018)

The Metronome is constructed with a steel basin that is 20 metres long and 3 metres wide, made from 4mm stainless steel. The side walls of the Metronome are folded to allow a bridge to smoothen the bed .

When the Metronome is not tilting, weirs are used to control the water elevation. A weir, which is effectively a small-scale dam, can be moved up or down precisely to regulate the water level on the upstream side, as water flow overtops the weir. The weir positioned on the seaward side is 3 metres wide and broad-crested, with a length of 0.06 meters and rounded edges resulting from the folded steel plate construction. Small actuators control the up and down movement of these weirs (M. G. Kleinhans, Leuven, et al., 2017). There is a slight overflow effect that occurs on top of the weir, resulting in the water elevation being slightly higher than the weir height. This is caused by the pumps pumping water into the seaward side, while the weirs maintains the water level.

To set up the experiment, water fills the system, controlled by adjusting the weir position. Once the metronome is filled, tilting is activated through the actuators, resulting in the simulation of tidal cycles that drive the morphological development of the sand.

3.2 Design of the experiment

Various setups are possible for the Metronome. In this thesis, data was available for one specific experiment, referred to as *Experiment 062*, which is described hereafter. Details of *Experiment 062*, are summarised in Table 3.1.

Parameter	**Value**	**Unit**
Tilting period	40	s
Maximum tilt gradient	0.008	m/m
Median sand grain size (D_{50})	0.52	mm
Sediment density (ρ_s)	2650	$\mathrm{kg\,m^{-3}}$

Table 3.1: Experimental parameters for tidal current generation.

In *experiment 062*, an initial simple channel shape was carved between rough industrial sand paper (median grain size of 0.7 mm) referred to as dikes. The dikes starting narrow and gradually widening to the width of the Metronome (see Figure 5.1). This initial setup facilitates the flow from the riverside boundary to the seaward side and back.

The sand is placed on top of a mat with artificial grass in the basin. Artificial grass is used as, past experiments by M. G. Kleinhans, Leuven, et al. (2017) showed that the smooth bottom surface of flumes increases downward erosion, forming scours. Implementing uniform roughness is essential to prevent scours across the Metronome.

The artificial grass was measured by M. Kleinhans et al. (2016) to be approximately 14 mm in height. The grass was glued in place and further secured with a layer of sand. Its stiffness ensures minimal bending even under the strongest experimental flows. The glue application prevents water from flowing under the grass (M. G. Kleinhans, Leuven, et al., 2017). The artificial grass contributes to the considerably darker color of the seafloor section (see Appendix 8.1).

To facilitate visualization of bar and channel morphology from overhead images, Brilliant Blue FCF colorant (E113) was used to dye the water. This organic dye, commonly employed as a food additive due to its non-toxic nature (EFSA, 2013), allows for clear distinction of features. However, determining the optimal dye concentration presents a challenge: too little dye results in minimal variation in water color across the experimental depth range, while excessive dye reduces light transmission through the water column, thereby compromising visibility for measuring bed topography (Leenman & Eaton, 2024). Past experiments by Huang et al. (2010) observed that beyond a certain depth, water color intensity plateaus regardless of depth, a threshold that decreases with higher dye concentrations. Thus, excessive dye concentration can compromise the accuracy of depth prediction methods, a factor essential for understanding water depth limits achievable with this dye concentration.

The concentration of the dye can be assumed to be consistent for both the experiments described in this paper at 5 mg/l. Maintaining this constant concentration is crucial, as fluctuations due to evaporation or water loss may necessitate recalibration of the depth-color relationship during or between experiments (Upson, 2024).

3.2.1 Calibration Experiments

This paper discusses *Experiment 062*, which is divided into two distinct parts: the *calibration experiment* and the *operational experiment*.

The *calibration experiment* begun after cycle 9000 in *experiment 062*. Where the bed was drained, then refilled incrementally to 4 different weir heights (the heights

can be found in Table 4.1). The metronome was not titled during the filling process and was level for this experiment. Each time the weir height was increased, overhead cameras captured images of the Metronome. Notably, small clumps of floating sand were observed on the water's surface during the experiment see Figure in Appendix 8.4.

3.2.2 Operational Experiments

The *Operational Experiment* refers to cycles 3499-4499 of *Experiment 062*. During this phase, images were captured between every tilting cycle, when the bed was level. However, due to the movement of water, ripples on the surface of the water were present in the captured images.

3.3 Data collection

3.3.1 Orthomosaisc

An orthomosaic is a seamlessly stitched image product created through photogrammetric orthorectification of a collection of images. This process corrects geometric distortions and balances colors to produce an accurately aligned mosaic dataset.

As described by Weisscher et al. (2020), seven industrial CMOS MAKO colour cameras capture the image data, that is processed into an orthomosaic. The placement of the cameras is illustrated in Figure 3.2. The cameras were positioned to be between 3.99-4.05 m from the co-aligned reference height of 0 (approximately the steel basin). Each camera has a dimension of 2,048 by 2,048 pixels.

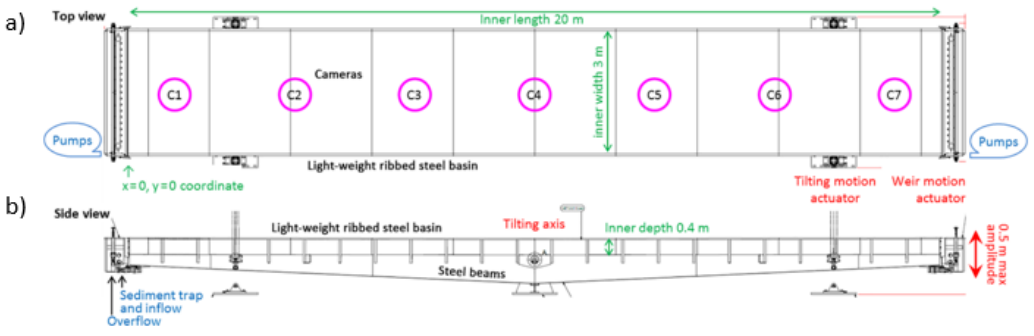


Figure 3.2: A top-down view of the CMOS MAKO camera placement and a side profile of the metronome showing the actuators placement. (M. G. Kleinhans, Leuven, et al., 2017)

According to Nota (2024a), the focal length of the cameras were calibrated in

the photogrammetry software to be between 2194-2239 pixels providing a spatial pixel resolution of 1.5–2 mm on the sediment surface of the Metronome.

Converting these seven images into one *orthomosaic* that can be accurately aligned with the Digital Elevation Model presents a challenging problem. This is due to image overlap (20%), lens distortion, color correction and alignment.

This challenge was addressed using a technique known as co-alignment with ground control points applied in research performed by Nota et al. (2022). Co-alignment refers to aligning multiple cameras at once instead of separately.

As described by Nota et al. (2022) photogrammetry refers to the process of aligning multiple images taken from different positions and angles into a unified coordinate system. Nota et al. (2022) found that the combined use of co-alignment and ground control points in photogrammetry is an effective method for achieving high relative and absolute accuracy. When this photogrammetric method is applied to orthomosaics in the Metronome, it produces images that can be aligned with the Digital Elevation Model.

Accurate alignment of images is just one aspect of producing reliable orthomosaics. When dealing with multiple orthomosaics to derive water height from imagery, maintaining consistent lighting conditions is crucial. The flume was illuminated at approximately 600 lux using daylight-colored fluorescent lights directed upward at a white, diffusive ceiling located about 4.5 meters above the flume floor. To minimize light reflection from the water surface in the photographs, a white photography backdrop cloth was placed between the ceiling and the flume.

3.3.2 Digital Elevation Model (DEM)

A laser camera was mounted on the gantry. The laser scanner registers the position of the laser line during a laser line procedure. The DEM is derived using Matlab software (Nota, 2024b).

It is challenging to precisely quantify the accuracy of the DEM model, as it is influenced by several factors including imperfect calibration, positioning, and orientation, with assumptions of constant roll, pitch and yaw positions, none of which are. It is expected that the errors in the DEM models will impact the accuracy of the water depth predictions.

Before the *calibration experiments* began, the dry bed of the Metronome was

screed to process a Digital Elevation Model (DEM), referred to as the *calibration DEM*, which was used for subsequent analyses. Throughout the experiment, the Metronome was not tilted and water was gradually filled. Maintaining a reasonable assumption that the *calibration DEM* remains consistent across the four water depths.

4. Method

4.1 Preparation of the data

The following steps outline the method of preparing the data: removing shadows from the orthomosaics, stacking the data, estimating the offset in the water depth maps, explaining the method used to estimate the water depth in the *calibration experiments*, explaining how the water height is predicted and comparing the water height predicted using the *operational images.*

4.1.1 Shadow Removal

There are visible shadows on the water surface of the orthomosaics in both *calibration* and *operational* orthomosaics. Due to the gantry and the actuators. These shadows can interfere with the accuracy of water depth models. To mitigate this issue, as demonstrated by Leuven et al. (2018), the orthomosaics were converted to LAB (CIELAB) color space images. The shadows are quite well isolated to the L* Band and are largely gone from the other two bands.

To remove shadows, Leuven et al. (2018) extracted the B band of the L*A*B colour space to predict water depth. In this experiment, to preserve as much data as possible, the L band was set to 127 (the constant middle point of the L* band) and subsequently converted back to the RGB color space. These colour transformations were performed in Python using the OpenCV (Open Source Computer Vision Library) (Bradski, 2000) and converted to a "Tagged Image File Format" to allow for stacking onto the NetCDF (Network Common Data) file format of the DEM. The *orthomosaics* that have a flattened L* band and converted back to RGB values are hereby referred to as *converted orthomosaics*.

4.1.2 Stacking

The four *converted orthomosaics* from the *calibration experiment* were converted to a dataframe consisting of 2,400,000 rows with an x, y and three RGB columns. The *calibration DEM* was imported, flipped and converted to a dataframe (2,400,000 rows). The orthomosaic dataframe was then converted to the same dimensions of

the the DEM. A stack could be developed by merging the dataframes based on x, y coordinates, without data loss.

4.1.3 Offset Calculation

The method described is to calculate an estimation of water depth for each pixel in the four *calibration stacks*. This is done using the known weir heights as listed in Table 4.1.

Weir Height	Value	Offset
Depth 0	46.36 mm	27.0 mm
Depth 1	75.70 mm	36.5 mm
Depth 2	82.73 mm	35.5 mm
Depth 3	94.95 mm	39.0 mm

Table 4.1: Weir heights and corresponding values with offsets.

Before detailing the process used to calculate water height, it is crucial to note that the weir height does not directly correspond to the absolute water height from the flume floor. An *offset* exists, likely due to the presence of artificial grass and water overflow atop the weir. The specific offset values for each depth are listed in Table 4.1 were determined through an iterative trial-and-error process.

This involved plotting water depth maps with varying offset values, comparing through visual inspection the water depth map at each depth with the respective overhead image at each depth, until a high degree of resemblance was achieved. While this method produced water maps with a close resemblance, it introduced a source of potential inaccuracy. Refer to Figure 5.1b.

4.1.4 Water Depth Calculation

Algebraic expressions to estimate water height on a pixel by pixel basis. Using the following equations in mm:

The relationship between H, z_w, and k is given by:

$$H = z_w - k \tag{4.1}$$

Here, H represents the total height from the steel floor to the top of the water

level, z_w is the water height of the weir, and k is combined offset caused by weir, the correction from visual inspection and the height of the artificial grass on the seabed lying.

The relationship between h and H, where z_b denotes the bed height, can be expressed as:

$$h = H - z_b \tag{4.2}$$

In this equation, h denotes the water depth from the terrain of the Metronome to the top of the water level at that pixel, and z_b is the height of the terrain at the pixel (DEM).

Further preprocessing includes incorporating conditional expressions to convert all negative water measurements to zero. This allows for the creation of a binary classification map distinguishing pixels as either *wet* or *dry*.

$$\textit{Water height} = 0 \;\; (\textit{dry})$$
$$\textit{Water height} > 0 \;\; (\textit{wet})$$

The area of interest lies between the dikes. A mask was drawn of this area and used to extract the area of interest.

To use this data in a machine learning model, the data had to be restructured into a long format where all four depths are appended onto each other along with their respective RGB values for each pixel. Where, for each pixel, there are four water depth values and four sets of RGB values.

4.2 Approach to zero-inflated data

Initial preprocessing revealed that the data was heavily imbalanced. When all four *calibration* experiments were stacked, 57.8% of the rows were found *dry*. A heavily imbalanced dataset can impact the accuracy of regression models. To address this issue, a two-step approach was developed.

Each analysis method (Linear Analysis, Random Forest, and SVM) was split into two models. The first model uses all four water depth stacks to train a binary classifier (Logistic Regression, Random Forest and SVM) to determine whether there is water or not, we refer to these models as *binary regression models*. The sec-

ond set of models we refer to as *regression models* (Linear Regression, Random Forest and SVM). The *regression models* aim to to predict the depth of *wet* pixels (predicted from the binary classification models) using the inputs of RGB columns.

4.3 Models

Efforts were made to compare the models as fair as possible. For each model the same training and testing dataset were used. With the exception of the SVM model which used a smaller proportion of the training data due to computational limitations.

Model	**Binary Classification Model**		**Regression Model**	
	Training	**Testing**	**Training**	**Testing**
Linear Model	768,000	117,596	11,745	73,237
Random Forest	768,000	117,596	223,707	73,237
SVM	76,800	117,596	22,370	73,237

Table 4.2: Number of samples used in training and testing linear and non-linear models. The linear model refers to the logistic regression and linear extrapolation technique.

4.3.1 Logistic Regression and Linear regression

The linear regression model was trained using a balanced training set consisting of 96,000 randomly sampled pixels from four water depth stacks. The set was balanced with 50% wet pixels and 50% dry pixels, each represented by RGB values for classification. The binary classification data was split into 80% training and 20% testing datasets (768,000 and 117,596 pixels, respectively). A random seed was set to ensure reproducibility across all subsequent random samplings.

The data was trained using the "lm" function as part of the base R package ((R Core Team, 2023)). Where the probability threshold was set at default (0.50).

For the linear regression model, only depth 1, depth 2 and depth 3 were used. Depth 0 is not used for regression training as it contains the most *dry* cells and due to the offset used (27.0mm) used in calculating to calculate H, is considerably lower compared to the other 3 depths (36.5mm, 35.5mm and 39.0mm), it would likely impact the results.

For the linear regression model, lines were generated for each pixel corresponding to three increasing water depths. The correlation coefficient (R^2) and slope were computed iteratively for each line, then plotted against the blue band.

Initial plots of the blue-band versus water height showed significant noise, prompting subsequent filtering. This study aims to evaluate the presence of a robust linear relationship. Therefore, pixels were filtered to include only those with $R^2 > 0.80$ and slopes less than 10 (as slopes exceeding this threshold were likely attributable to noise).

The equation of the line was fitted on the filtered data and used as a simple linear prediction model that could input a blue-light band value and output a water depth prediction.

4.3.2 Random Forest

The R *Ranger* package by (Wright & Ziegler, 2017), was used for the Random Forest models. For the Random Forest binary classification model. The data (768,000 and 117,596 pixels) was the same train/test data as the logistic regression model. The model was grown using 500 trees set as a classification Random Forest.

For the Random Forest regression model, pixels with depth 0 were excluded, and *dry* pixels were filtered out. A total of 296,944 pixels were randomly selected. Eighty percent (223,707 pixels) of the data were used for training, and the remaining 20 percent (73,237 pixels) were used for model evaluation. To build the regression Random Forest 500 trees were grown.

Feature importance was calculated for Random Forest Model. Feature importance is a technique used to understand which variables (features) in the dataset have the most influence on the model's predictions. The importance is typically measured by how much each feature contributes to decreasing the Gini impurity when building the decision trees. Gini impurity measures the degree or probability of a particular element being wrongly classified when it is randomly chosen.

4.3.3 Support Vector Machine

The R *Caret* package was used for the SVM models (Kuhn, 2008). Only 10% (76,800 pixels) of the same training data used for the Random Forest and Logistic Regression was used in the training dataset due to computational resource constraints. The binary classification model was specified as type-C classification and used the linear kernel. Where type-C classification specifically refers to a scenario where the goal is to mean shape of the decision boundary used to separate classes was linear.

To compute the SVM regression models, only 10% (76,800 pixels) of the same

training data used for the Random Forest was used. The choice of using a radial basis function (RBF) kernel with a *gamma* value of 0.1 in our SVM regression model was aimed at balancing model complexity and generalization. The RBF kernel is well-suited for capturing non-linear relationships in data, making it effective for our task. A *gamma* value of 0.1 indicates a moderate level of model complexity, allowing the SVM to generalize well while still capturing intricate patterns in the data. To optimize the *gamma* value, experimentation and cross-validation can be done to refine the model, but was not done here.

4.4 Evaluation of model performance

The binary and regression models were evaluated using different methods.

4.4.1 Evaluation of binary classification models

For the evaluation of the classification models, all models were assessed using the same test data. Each classification model was evaluated using *accuracy* and a *confusion matrix*. *Accuracy* is defined as the number of correct predictions divided by the total number of predictions.

$$A = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(\hat{y}_i = y_i) \tag{4.3}$$

Where, $\mathbb{I}(\hat{y}_i = y_i)$ will be 1 if the predicted label $\hat{y}_i$ matches the actual label y_i, and 0 otherwise.

A confusion table shows the counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions.

	Predicted	
Actual	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Table 4.3: Confusion matrix showing the counts of true positive (TP), false negative (FN), false positive (FP), and true negative (TN) predictions.

The recall (or sensitivity) of a classification model is defined as:

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{4.4}$$

4.4.2 Evaluation of regression models

To evaluate the performance of each model, MAE (Mean Absolute Error), MedAE (Median Absolute Error), MSE (Mean Squared Error), and RMSE (Root Mean Squared Error) are used. All metrics are calculated using the same test data across the three models.

Mean Absolute Error (MAE) is defined as the average of the absolute differences between the predicted and actual values. Where y_i represents the actual values, $\hat{y}_i$ represents the predicted values, and n is the number of samples.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4.5}$$

The Median Absolute Error(MedAE) is the median of the absolute differences between predicted values and actual values. We can use it as a bias metric partner to the MAE to see if the is overestimating or underestimating the predictions.

$$\text{MedAE} = \text{median}\left(|y_i - \hat{y}_i|\right) \tag{4.6}$$

$$\text{Bias Metric Partner} = -1 * \text{MedAE} \tag{4.7}$$

A positive *bias metric* suggests the model is overpredicting, while a negative value suggests it is underpredicting.

Mean Squared Error (MSE) is defined as the average of the squared differences

between the predicted and actual values.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{4.8}$$

Root Mean Squared Error (RMSE) is defined as the square root of the average of the squared differences between the predicted and actual values.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{4.9}$$

The predicted results were plotted against the observed calculated values in the test set.

Furthermore the MEC was used to evaluate the performance of a predictive model, it measures how well the model's predictions match the observed calculated data. An MEC of 1 indicates perfect predictions. An MEC of 0, indicates the predictions have the same variance as the observed data (Nussbaum et al., 2023) .

The Model Efficiency Coefficient (MEC) is defined as:

$$\text{MEC} = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{5.1}$$

where:

- MEC is the Model Efficiency Coefficient.
- O_i are the observed values.
- P_i are the predicted values.
- $\bar{O}$ is the mean of the observed values.
- n is the number of observations.

4.5 Prediction error maps

To evaluate all three regression methods, we applied their respective binary classification and regression models to *calibration image 2*. Prediction error maps were generated for each model by subtracting predicted water heights from actual water heights for every cell, enabling us to compare their results. *Calibration image 2* was chosen because it has clearly defined dry and wet sections, facilitating easier visualization and comparison.

For the Random Forest model, prediction error maps were generated for all 3 depths.

4.6 Applying models to operational images

The two-step Random Forest model was selected to predict water depth on *operational images*. A subset of the *operational images* was selected. Images were chosen from tidal cycles (3509-4499). To reduce processing time, every tenth image from a total of 990 tidal cycles was selected, resulting in a subset of 99 images.

Each image underwent the same preprocessing steps as the calibration images. First, the images were scaled to match the metronome dimensions. Next, they were transformed into the LAB color space, where the L band was fixed at 127 to standardize brightness. Subsequently, the images were converted back to the RGB color space.

After preprocessing, a Random Forest Binary Classification Model was employed to predict whether each cell in the image was wet or not. Following this classification, a regression model using the RGB values was applied to predict the water depth for each cell.

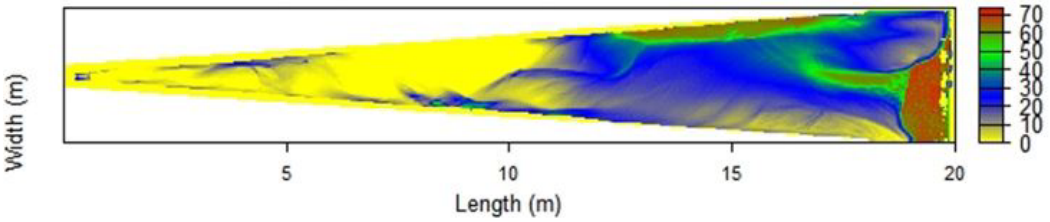
This iterative process was repeated for all 99 images, allowing us to systematically assess the model's performance across different tidal cycles.

Using the time step of 990 cycles. A water prediction map of cycle 4499 was subtracted from cycle 3509. To illustrate if the model is precise enough to detect water depth change.

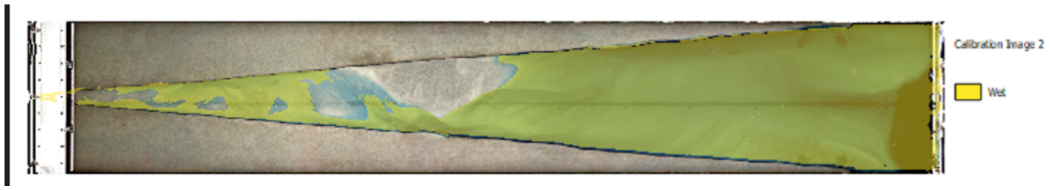
5. Results

5.1 Water depth maps

Four water depth maps were calculated; an example of water depth 2 is shown in Figure 5.1a. A binary *wet/dry* map is shown in Figure 5.1b, which filters out the dry pixels and keeps all *wet* pixels as a binary 1. It is then overlaid onto the orthomosaic of calibration image 2.



(a) Depth 2 water depth calculated using weir height and the digital elevation model.



(b) Calculated dry and wet cells where water depth is ≤ 0 mm, overlaid onto the RGB image.

Figure 5.1: Example of water depth maps for water depth 2

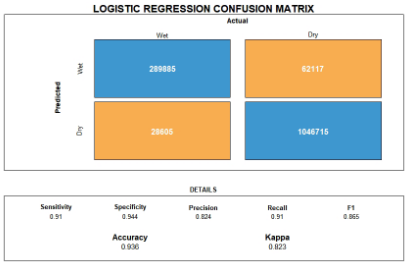
It is clear from Figure 5.1b that there is a correctly identified *dry* area, but it incorrectly identifies some areas where there is visually water as *dry*.

5.2 Binary classification Results

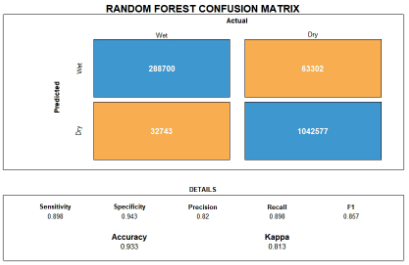
The results of classification models using Logistic Regression, Random Forest and SVM are shown two fold. First the results are presented with confusion matrices, MAE. Second the error map of calibration image 2. As mentioned, the SVM model was trained on 10% of the training the other models used.

5.2.1 Model Performance

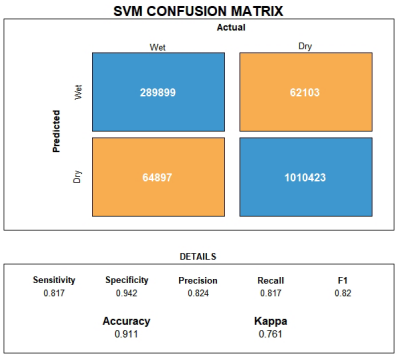
To evaluate the performance of the binary classification models, confusion matrices were generated for each algorithm. Confusion matrices are essential tools for assessing the effectiveness of classifiers by summarizing their predictions against actual class labels. Each matrix visualizes the counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).



(a) Confusion Matrix for Logistic Regression



(b) Confusion Matrix for Random Forest



(c) Confusion Matrix for SVM

Figure 5.2: Confusion Matrices for the binary classifier

5.2.2 Error prediction maps

The Random Forest binary predictor model was used to predict *wet* and *dry* cells and compared to the calculated *wet* and *dry* cells. The results were overlaid on the original RGB image.

The model incorrectly classifies some areas as *dry* in Figure 5.1b, which are actually *wet*. Although the model correctly predicts these pixels as *wet* from their RGB values, they are evaluated as false positives. Dikes and borders of the Metronome (specifically the bridge on the upstream side, the elevation sensors and the wavemaker on the downstream side) are also evaluated as incorrect predictions. These

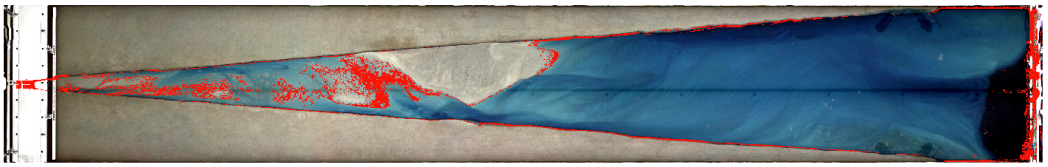


Figure 5.3: Incorrectly predicted pixels using the Random Forest model were mapped back onto the original orthomosaic. Red indicates an incorrectly predicted pixel.

borders can be removed by future masking so are of little concern.

5.3 Regression Results

The results of the regression can be compared using the MAE, Bias Partner Metric, RSME and MSE values of each model. The models are tested on the same data. But SVM was trained on 10% of the training data of the RF model due to time required for model training.

Model	MAE(mm)	Bias Metric (mm)	RMSE(mm)	MSE(mm^2)
Linear extrapolation	10.38	3.50	16.09	263.58
Random Forest	3.87	0.44	6.53	42.64
SVM	5.77	-0.63	12.33	152.23

Table 5.1: Performance Metrics of Depth Prediction Models tested using 73,237 pixels

The negative *bias metric* suggests that the SVM model underpredicts, while the other models overpredict. All models seem to be affected by large errors (high MSE), and Random Forest has the lowest MAE.

5.3.1 Linear Regression

Part of this research is to establish if there is a linear relationship that exists between colour intensity and height and can this relationship be used to accurately predict water height.

Using the initial predictions of *wet* and *dry* cells from the linear regression model. The predicted *wet* cells were selected. The intensity of colour at each band (RBG) was plotted against their water height increase over the 3 cycles.

The resulting lines were calculated, and their coefficients were fitted. The initial plots of band-to-water height exhibited significant noise and were subsequently filtered out. This research aims to assess the presence of a robust linear relationship. Therefore, pixels were filtered to include only those with a ($R^2 > 0.80$) and slopes

less than 10 (as slopes greater than this were likely due to noise).

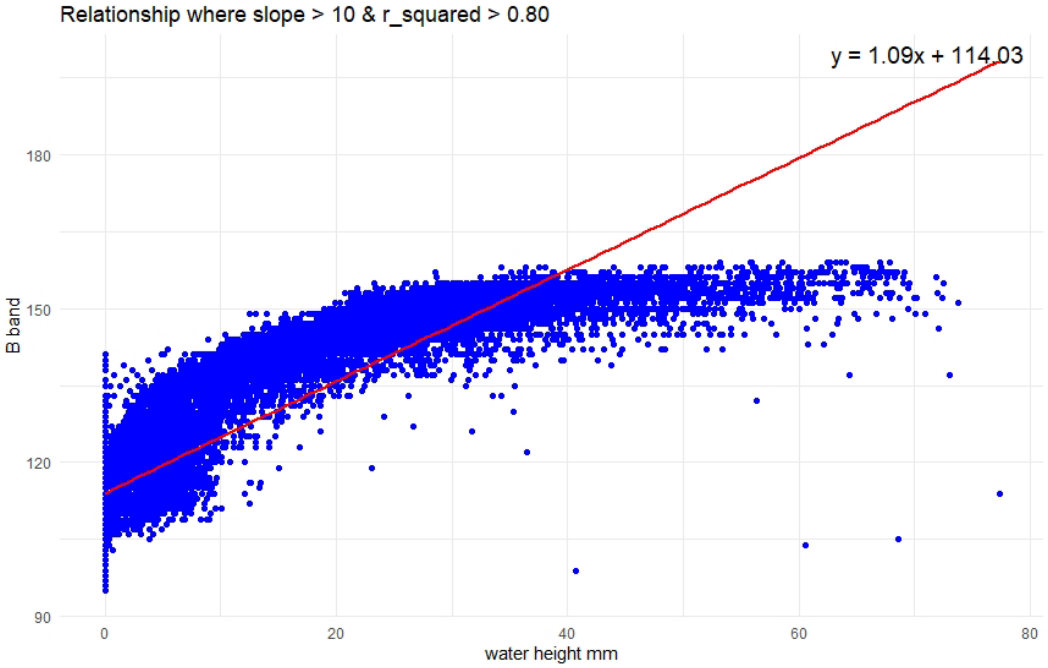
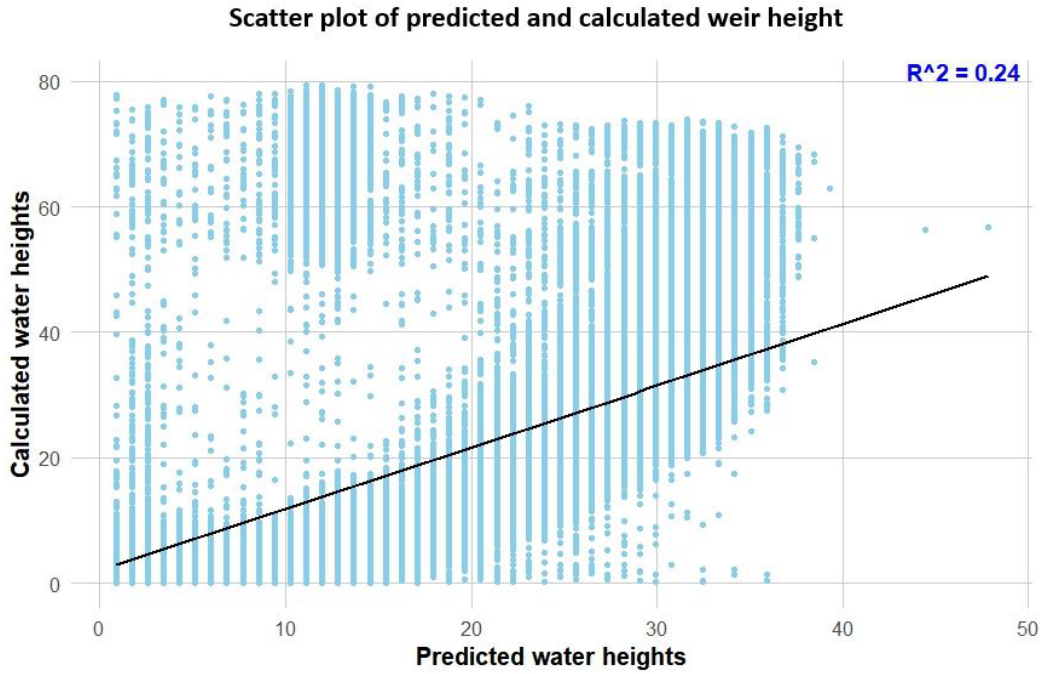


Figure 5.4: Plot of pixel depth with RGB-B band values

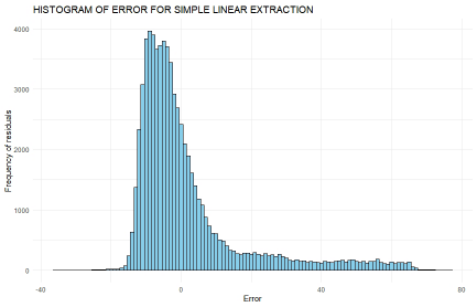
The graph was plotted this way for visualisation purposes. To form a simple usable linear extraction the equation of the line of best fit can be rewritten as Equation 5.1.

Using the linear model. The predicted water depths were compared to the calculated water depths in the test dataset, then the distribution of the residuals were plotted. Additionally, the distribution of residuals greater than $\pm 2\sigma$ (standard deviations) from the mean were highlighted.

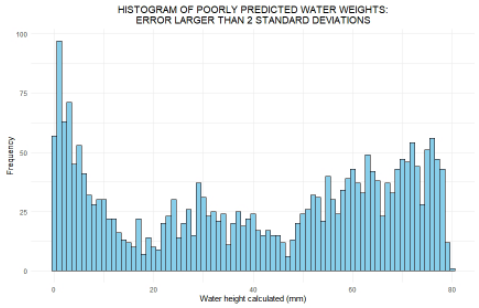
$$h = \frac{B - 114.03}{1.09} \tag{5.1}$$



(a) Scatter plot of Results, with line of best fit



(b) Histogram of residuals showing the distribution of errors



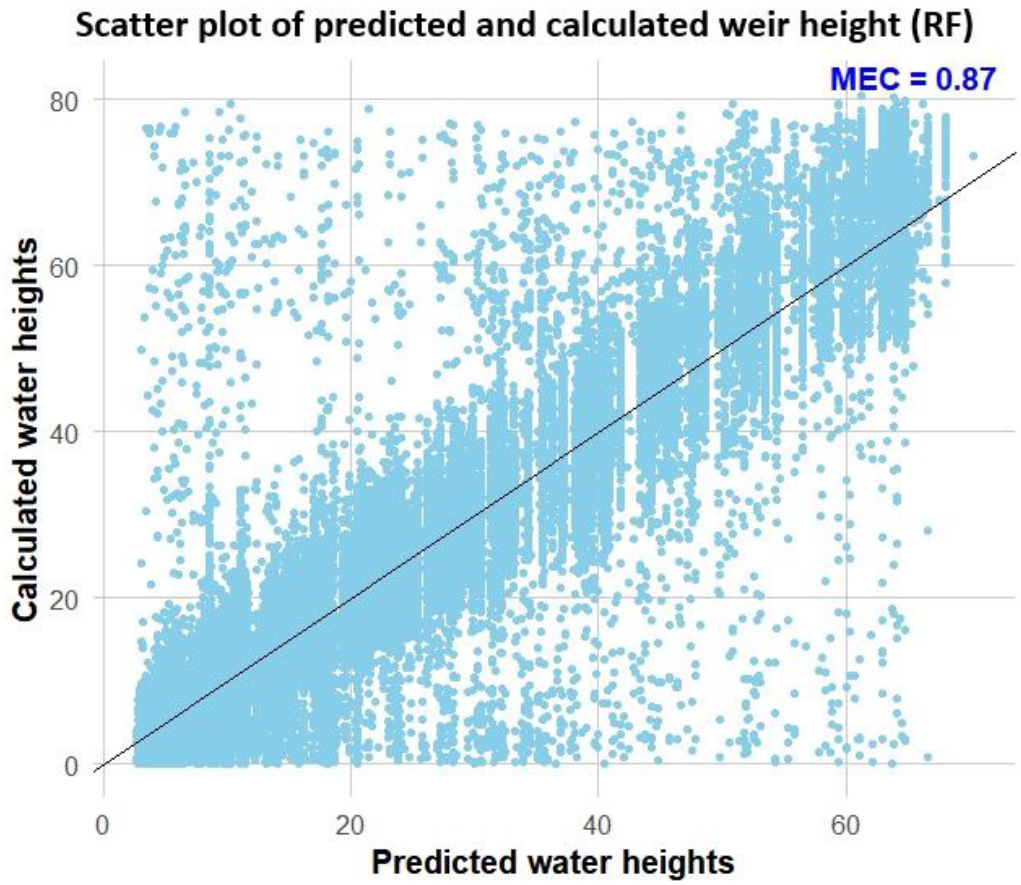
(c) Distribution of residuals greater than $\pm 2\sigma$ from mean error

Figure 5.5: Comparison of histograms and poor predictions

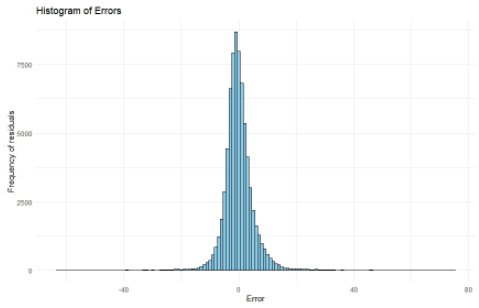
A low MEC (Model Efficiency Coefficient) indicates that the linear regression model performs poorly. The positive skew in the distribution of errors suggests that the model tends to underpredict. It also overpredicts as the tail of the distribution extends, it highlights the limitations of the linear assumption, with the errors becoming increasingly spread out.

5.3.2 Random Forest

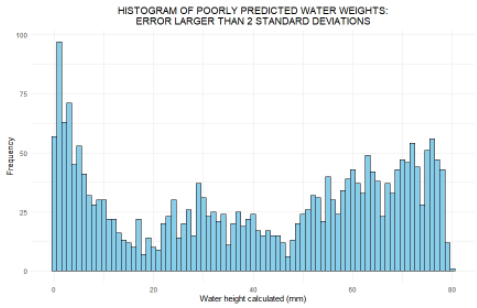
The Random Forest model predicted versus calculated water heights were plotted. Then the residuals were plotted. As was feature importance.



(a) Scatter plot of results, with line of best fit



(b) Histogram of residuals showing the distribution of errors



(c) Distribution of residuals greater than $\pm 2\sigma$ from mean error

Figure 5.6: Comparison of histograms and poor predictions

The scatter plot and the distribution of residuals showing both overpredictions and underpredictions indicate that while the model is generally accurate, it does have instances where it deviates from the actual values. The MAE of 3.87 quantifies the average magnitude of the errors in the model's predictions, without considering their direction. This value provides a sense of how much, on average, the predictions deviate from the observed values. The MSE of 42.64 indicates that the model's errors are relatively large in magnitude, as MSE amplifies the impact of

larger errors by squaring them. This means that although the MEC is high, suggesting a generally good model fit, there are significant outliers where the model prediction errors are substantial. When we consider c) the distribution of residuals greater than $\pm 2\sigma$ from mean error. It indicates that the model has high prediction errors particularly at very low and very high water depth.

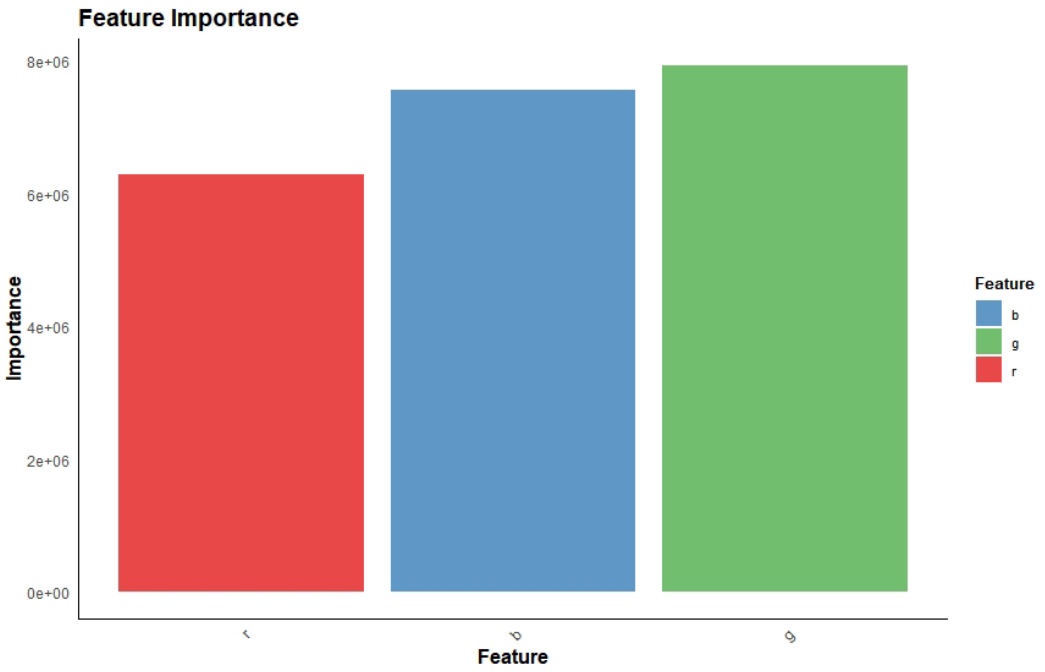
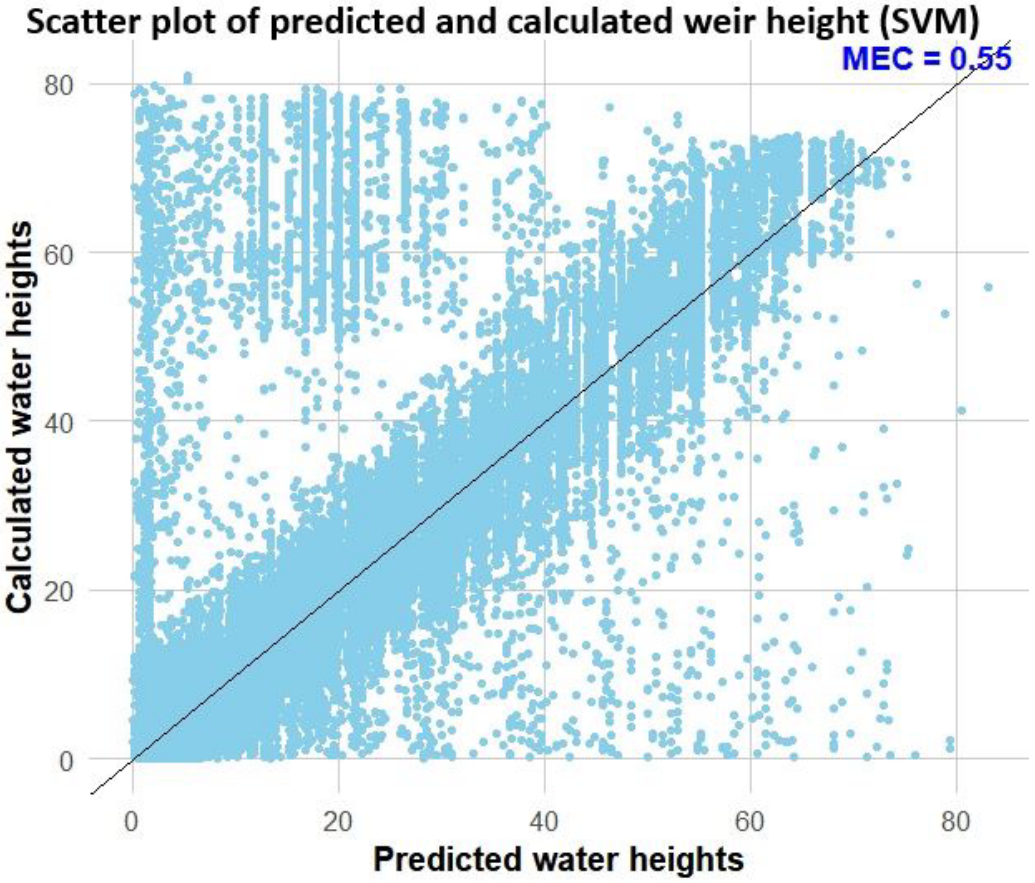


Figure 5.7: Feature Importance of RGB Colour bands in the Random Forest Model for predicting water height.

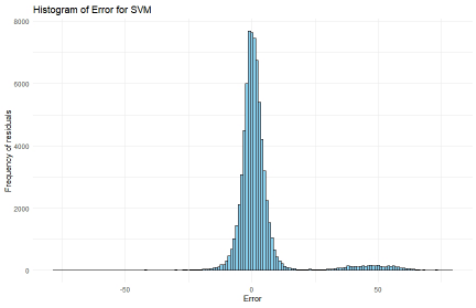
The G band of the *converted orthomosaics* is the most important feature in the RGB band in predicting water height.

5.3.3 SVM

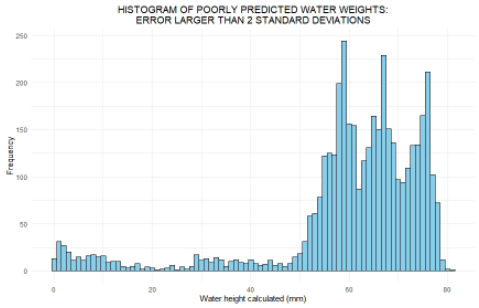
The SVM model predicted versus calculated water heights were plotted. Then residuals were plotted.



(a) Scatter plot of results, with line of best fit



(b) Histogram of residuals showing the distribution of errors



(c) Distribution of residuals greater than $\pm 2\sigma$ from mean error

Figure 5.8: Comparison of histograms and poor predictions

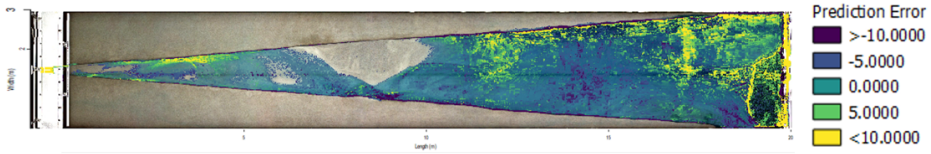
An MEC of 0.58 suggests that the model has a moderate level of accuracy in predicting the observed data. This value indicates that while the model captures some of the underlying patterns in the data, there are still considerable discrepancies between the predicted and observed values.

The distribution of residuals, particularly the negative skew in residuals greater than $\pm 2\sigma$ from the mean error, suggests that large errors of the model occur for

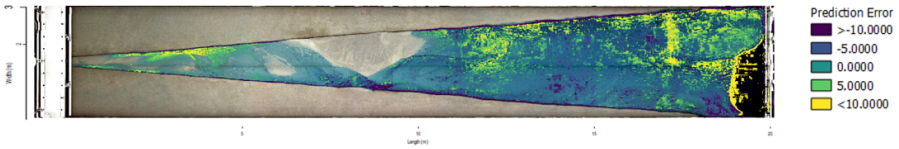
deep water depths. This means the model tends to underpredict or overpredict significantly at these depths, leading to the larger errors observed.

5.4 Comparing errors

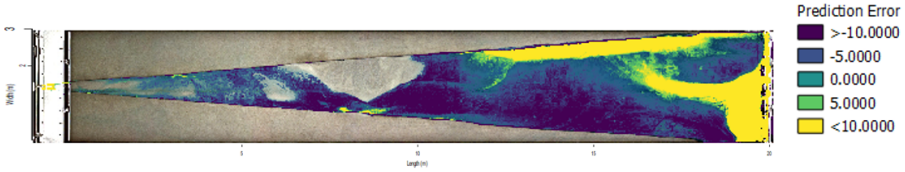
To visually compare how the models are performing the models (binary and regression) were reapplied to the calibration 2 image and overlaid with the original RGB image. The error was calculated on all 4 depths for the RF model and can be found in Appendix 8.3.



(a) Prediction error of Random Forest model



(b) Prediction error of SVM



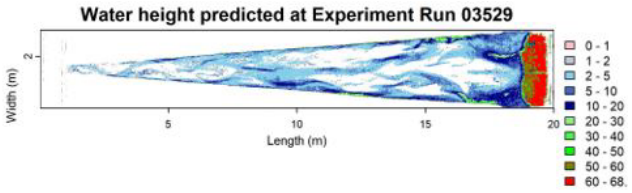
(c) Prediction error of Linear Extrapolation

Figure 5.9: Comparison of Prediction Errors

5.5 Applying the Random Forest model to operational images

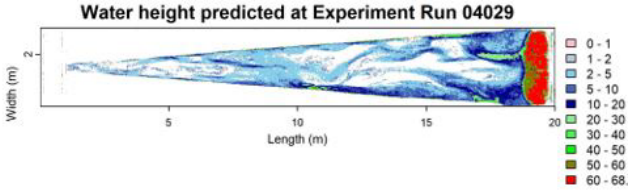
The two-step Random Forest model is applied to the *operational images*. Below are three images: one taken at the beginning, one in the middle, and one at the end of the process. The respective orthomosaics and water depth predictions are shown. An accompanying video is included See Supplementary Video in the supporting information of this document showing the 99 images.

(a) Orthomosaic at cycle 3509



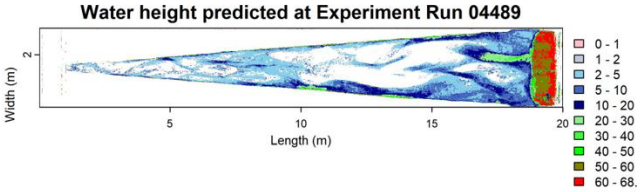
(b) Water depth prediction at cycle 3509

(c) Orthomosaic at cycle 4029



(d) Water depth prediction at cycle 4029

(e) Orthomosaic at cycle 4489



(f) Water depth prediction at cycle 4489

Figure 5.10: Orthomosaic and corresponding water depth predictions at cycles 3509, 4029, and 4489

5.5.1 Visualising morphological change using water depth predictions

It becomes important to determine if the predictions of the Random Forest produces results that are precise enough to detect change.

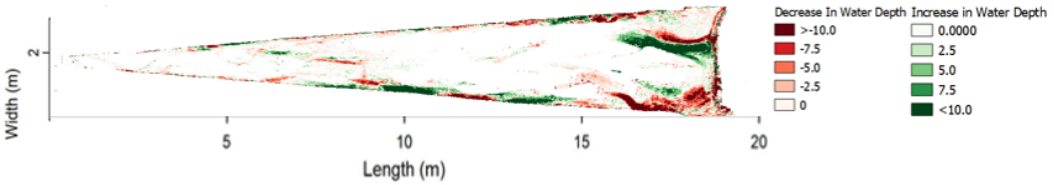


Figure 5.11: The difference in water predictions of cycle 4489 and 3509

The green color on the map indicates predicted deeper water depths, which may suggest scour development, while red indicates shallower water depths, potentially indicating sand deposition

6. Discussion

6.1 Water depth calibration

The accuracy of the predictions depends significantly on the precision of the calibration dataset. If the height of the DEM is not accurate, neither is the calculated depth of the water. These depths are what the model is trained on, so the accuracy of the DEM is crucial for accurate predictions.

The *calibration DEM* used to calculate bed height z_b introduces potential errors due to lens distortions, positional variations, and orientation inaccuracies. As the same DEM is used consistently, these errors remain temporally uniform across the entire calibration dataset, affecting the overall accuracy of the predictions. But spatially the errors in the DEM are dependant on the laser camera intrinsics and extrinsics. Future developments of the software used to compute these DEM models could reduce the error.

To compute the water depth, the weir height z_w, was adjusted by a value k, to account for both the weir overflow and artificial grass offset. The accuracy of this offset is also crucial to water depth prediction. This offset was subtracted from the weir height.

A nonlinear value of k between the values of 27.0 to 39.0 mm across four calibration images was deemed too large for effective model training. To minimize errors, images 1 to 3, with offsets of 36.5 to 39.0 mm, were used for regression training. However, since precise height measurement is less critical for binary classification, image 0 was also included in the training dataset for that purpose.

For this experiment the value of k was essentially estimated (through a visual trial and error process) which in turn introduces another source of error into the water depth predictions. Experiments could be done to measure the overflow height above the weir at different weir heights and explore relationship between weir overflow and weir height. The resulting misalignment between the calculated water height map and orthomosaic is evident in Figure 5.1, where certain sections of water are not adequately covered.

6.2 Binary Classification

This misalignment primarily accounts for the main source of error in the binary classification of "wet" and "dry" pixels. As depicted in Figure 5.3, the model correctly identifies certain cells as "wet" based on RGB values of the *converted orthomosaics*, but due to calibration issues, these cells are incorrectly classified as "dry". Misclassification of dikes, specific seafloor areas, and Metronome borders also occur however, these are less concerning as they can be filtered out in subsequent experiments using a more refined vector mask.

All three models perform satisfactorily in this task, aiming to filter out *dry* cells without losing any *wet* cells. In this context, missing a positive prediction (false negative) is more critical than falsely predicting a positive outcome (false positive). Therefore, recall is the most important metric.

Logistic Regression, Random Forest, and SVM achieve Recall scores of 0.91, 0.90, and 0.87, respectively, indicating their effectiveness in minimizing false negatives by accurately identifying most "wet" cells. This high Recall ensures that these models are reliable for predicting "wet" cells, with misclassified cells primarily being false positives.

However, the misalignment in calibration images might affect the models' reliability and transferability. Given the promising regression results, future calibration experiments can address and refine these issues to enhance model performance.

6.3 Regression results

The Random Forest model was chosen to predict on the *operational* images. This was due to the low MAE 3.87 mm compared to 10.38mm and 5.77mm (linear extrapolation and SVM).

For the linear regression technique, the technique was investigated as it offers the advantage of being computationally efficient and easy to implement.

However, under these conditions, the results are unacceptably unreliable. A linear relationship between the B-band and water depth appears to exist (in particular depths), but extracting this relationship is challenging due to data noise and depth limits where linearity holds and it appears limited to a narrow range of depth. Beyond a depth of 40 mm, water depth increases without a proportional increase in

B-band intensity. While marginal improvements might be possible with different bands or band indices (Lyzenga, 1978), there are inherent limitations.

The results indicate that despite being previously used for water depth prediction (Upson, 2024; Leuven et al., 2018) (MAE not reported), the model achieves visually similar water depth prediction maps to the overhead imagery. However, given the reported MAE of (10.38 mm) at this concentration of water dye, a linear assumption of water depth should not be considered accurate.

Thus, the research focused on predicting water depth using non-linear methods, concluding that linear models are not suitable for this task with overhead photos. Consequently, the rest of this discussion focuses on non-linear methods.

6.3.1 SVM and Random Forest

Considering the SVM model, the Metronome generates a substantial volume of images, necessitating consideration of computational efficiency. While SVM was explored in this study, compromises were necessary due to computational constraints. The SVM model was trained on one-tenth of the dataset used for other models but was tested on the same test set. Although the results themselves do not necessarily disqualify SVM, the unequal comparison and the computational time required for training and prediction were prohibitive. Therefore, Random Forest (RF) was selected for predicting *operational* images.

The results indicate that the regression models perform well overall, despite the high RMSE and MSE (Table 5.1). RMSE, and MSE are particularly sensitive to outliers. These outliers may stem from various sources, such as measurement errors from the mask capturing border areas and dikes, model inaccuracies, imperfections in the orthomosaics, the dye reaching the upper limit to colour change with increasing depth.

The Random Forest model shows the lowest MSE (42.64 mm). However, due to the narrow ranges of the Metronome, it is critical to identify outliers and understand their distribution across different water depths. Analysis of the histogram of errors suggests a normal distribution. Filtering the residuals and plotting errors greater than 2 standard deviations reveals that shallow depths and deep depths exhibit the highest errors.

Examining the error map of the Random Forest prediction on *calibration image 2* (Figure 6.1), highlights specific regions where the model's predictions diverge

significantly from actual depths.

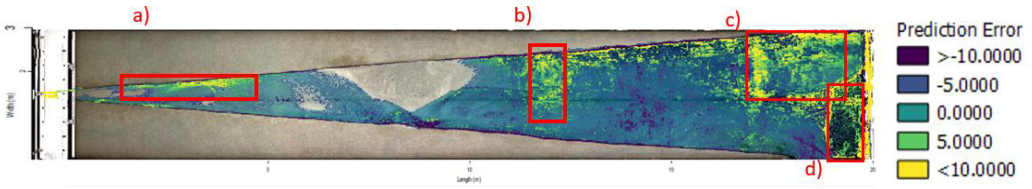


Figure 6.1: The difference in water predictions of cycle 4489 and 3509. Riverside opening (*a*), section in the middle (*b*, upper right (*c*), and on the seafloor (*d*)

The error in (d) is not concerning, as the seafloor depth does not affect the changing morphological features of the Metronome. It can be effectively masked out in future applications. Similarly, the area near the river mouth (a) can also be masked out.

To explain the under-prediction in (b) and (c), we referred to the calculated water depth maps and compared the areas of large errors with the calculated depth in those regions, as detailed in in Appendix section 8.2. It was observed that these under-predictions begin when the calculated depth exceeds 30 mm. In these cases, the model predicts approximately 10 mm or less water depth at the same pixel, suggesting that > 30 mm threshold, the calibration experiment can under-predict depth. Another consideration is colour preprocessing technique used in creating the *orthomosaics* (Weisscher et al., 2020) is not optimised, this could contribute to the underpredictions, due to inconsistent colouring. Future research will likely resolve this issue by applying white balancing to all overhead cameras.(Nota, 2024c).

The issue of under-prediction may be partially addressed by repeating the experiment with a lower concentration of dye. However, this effect is observed only in certain pixels, and overall, the predictions perform well, with MAE values of 3.87 mm. It is difficult to compare the performance of this model with other research. As no other research was found on the topic of using machine learning models to predict water depth in laboratory flumes. But this research does indicate that the method applied successfully in remote sensing (Vojinovic et al., 2013; Mateo-Pérez et al., 2021; Disanayaka Mudiyanselage et al., 2022; Zhang et al., 2020) can be applied successfully to laboratory scale experiments. Meaning future research could also apply more advanced models such as Convolutional Neural Networks (Ai et al., 2020).

6.4 Application to operational imagery

The model, when applied to operational images, demonstrates precision in depicting changes in water depth in moving images. This promising capability suggests that future research could utilize this model for channel emergence, scour detection, and monitoring sand bar deformation.

By examining areas where water height increase greater than 10 mm (Figure 5.11), a scour (dark green) is notably emerging at the seabed entrance after just 990 cycles. This observation is supported by the orthomosaic from cycle 4489 (Figure 5.10f).

Conversely, water height decrease, can be used to detect sand deposition. A sand bar is observed to be growing in the top right corner of the image, evidenced by a decrease in water depth (red), indicating sediment deposition over tidal cycles. These changes are more discernible using water depth change maps compared to orthomosaics. With an extended timescale, these findings are expected to become even more pronounced.

The experiment demonstrates promising results that can potentially be applied to other experimental orthomosaics, provided consistent preprocessing steps are undertaken (such as LAB transformation and pixel alignment), along with maintaining uniform dye concentration and lighting conditions. Building upon and improving the techniques preformed by Weisscher et al. (2020).

Further improvements could include repeating the calibration and model training with a broader range of depths, using a weaker concentration of dye, ensuring removal of floating sand particles from the calibration imagery, refining the mask to exclude narrower areas (such as removing the dikes and seafloor), accurately determining weir height offsets at different depths, and enhancing the calibration of the Digital Elevation Model (DEM).

Future applications could include implementing edge detection methods (Pirzada & Siddiqui, 2013) to delineate estuary water boundaries and identify shapes of continuous channels. Additionally, smoothing techniques using K-Nearest Neighbours (Milan et al., 2016) could enhance the visualization of longer and continuous water channels.

7. Conclusion

This project aimed to develop and apply linear and non-linear models to derive water depth data from overhead imagery in the Metronome. By utilizing DEM data and known weir height increases in *calibration experiments*, we generated estimations of water depth per pixel. We overlaid the pixel values in the *converted orthomosaics* onto the water depths and evaluated several machine learning models for predicting water depth.

The results indicate that Random Forest exhibited the most promising performance in predicting water depth, achieving a superior MAE of 3.87 mm with a recall of 0.90 for *wet* cells, compared to Support Vector Machine (MAE: 5.77 mm, recall: 0.87). Random Forest also demonstrated significantly better computational efficiency.

It was found that linear extrapolation of depth produced an inaccurate results as the dataset method is more prone to noise and the linear relationship is limited to depth (max 40mm).

The application of the Random Forest model to *operational images* shows promise in change detection, scour identification, sand bar deformation, feature analysis and water height analysis. Future research could further enhance this by edge detection and K-Nearest Neighbour methods.

Challenges for the Random Forest model included its sensitivity to outliers. (MSE: 42.64mm), and the model is prone to under-predicting depth larger then 30mm. This could be addressed in future experiments by using a weaker dye concentration and adding more depths to the calibration images. Addressing these issues could lead to more accurate predictions and broader applicability.

Overall, this study contributes to challenge of predicting water depth from overhead imagery. By leveraging the method described in this paper, researchers can advance the understanding of scaled morphological features of Metronome under different conditions.

8. Appendix A

8.1 Calibration RGB orthomosaics

Figure 8.1: Calibration Orthomosaic depth at weir height 46.36 mm. Red box identifying seabed area.

Depth 1

Figure 8.2: Calibration Orthomosaic depth at weir height 75.70mm

Depth 2

Figure 8.3: Calibration Orthomosaic depth at weir height 82.73mm

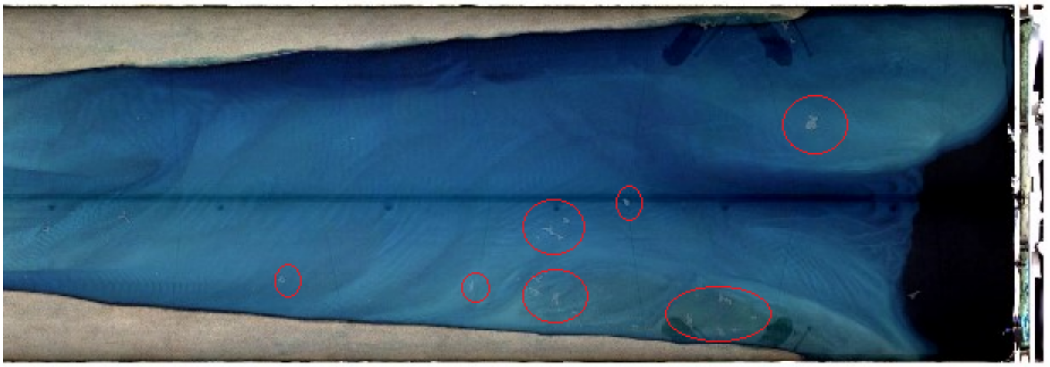


Figure 8.4: Zoom of depth 2 image to show visible floating sand clumps

Depth 3:

Figure 8.5: Calibration Orthomosaic depth at weir height 94.85mm

8.2 Random Forest Predictions water depth on calibration images:

Depth 1

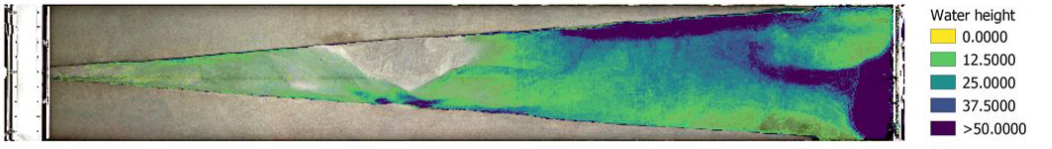


Figure 8.6: Random Forest Water Depth Prediction for image 1

Depth 2

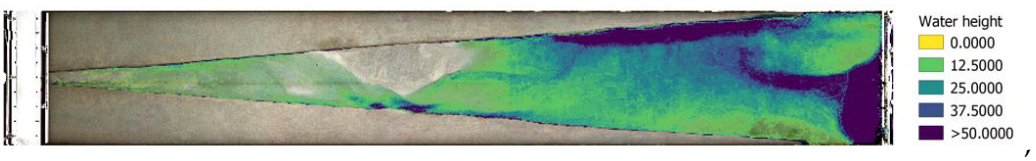


Figure 8.7: Random Forest Water Depth Prediction for image 2

Depth 3:

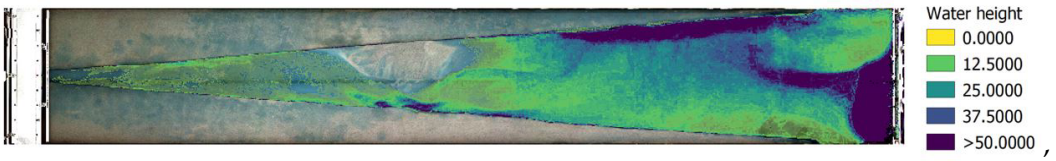


Figure 8.8: Random Forest Water Depth Prediction for image 3

8.3 Random Forest Prediction error on calibration images:

Depth 1:

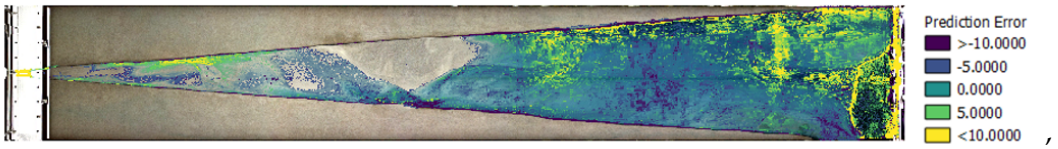


Figure 8.9: Random Forest Prediction error at weir height 75.70mm

Depth 2:

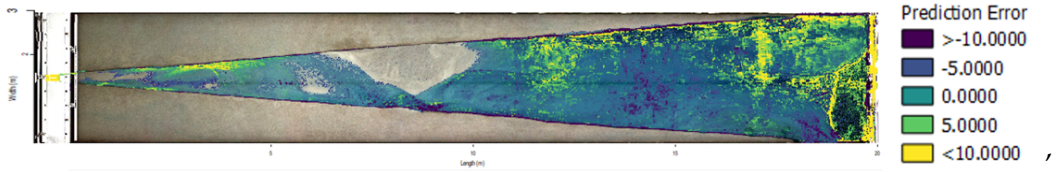


Figure 8.10: Calibration Orthomosaic depth at weir height 82.73mm

Depth 3:

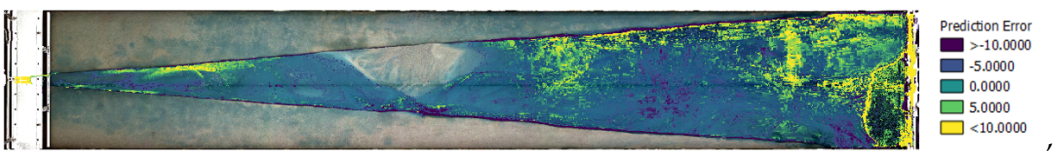


Figure 8.11: Calibration Orthomosaic depth at weir height 94.85mm

8.4 Water depth prediction on operational images using Random Forest

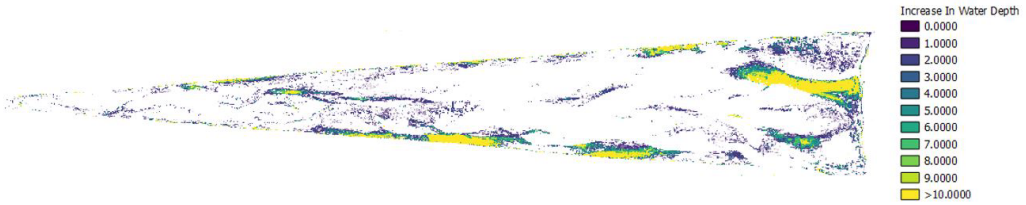


Figure 8.12: Water height density distribution curvees.

Figure 8.13: Predicted water height decrease between 4489 and 3509.

Bibliography

Weisscher, S. A. H., Boechat-Albernaz, M., Leuven, J. R. F. W., Van Dijk, W. M., Shimizu, Y., & Kleinhans, M. G. (2020). Complementing scale experiments of rivers and estuaries with numerically modelled hydrodynamics. *Earth Surface Dynamics*, *8*(4), 955–972. https://doi.org/10.5194/esurf-8-955-2020

Leduc, P., Peirce, S., & Ashmore, P. (2019). Short communication: Challenges and applications of structure-from-motion photogrammetry in a physical model of a braided river. *Earth Surface Dynamics*, *7*(1), 97–106. https://doi.org/10.5194/esurf-7-97-2019

Tal, M., & Paola, C. (2010a). Effects of vegetation on channel morphodynamics: Results and insights from laboratory experiments. *Earth Surface Processes and Landforms*, *35*(9), 1014–1028. https://doi.org/https://doi.org/10.1002/esp.1908

Upson, M. (2024). A method using remote sensing to determine water depth on a laboratory scale. *Earth Surface and Water*, (2737957).

Geyman, E. C., & Maloof, A. C. (2019). A simple method for extracting water depth from multispectral satellite imagery in regions of variable bottom type. *Earth and Space Science*, *6*(3), 527–537. https://doi.org/https://doi.org/10.1029/2018EA000539

Lyzenga, D. R. (1978). Passive remote sensing techniques for mapping water depth and bottom features. *Appl. Opt.*, *17*(3), 379–383. https://doi.org/10.1364/AO.17.000379

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

Vojinovic, Z., Abebe, Y., Ranasinghe, R., Vacher, A., Martens, P., Mandle, D., Frye, S., Ettinger, E., & Zeeuw, R. (2013). A machine learning approach for estimation of shallow water depths from optical satellite images and sonar measurements. *Journal of Hydroinformatics*, *15*, 1408–1424. https://doi.org/10.2166/hydro.2013.234

Mateo-Pérez, V., Corral-Bobadilla, M., Ortega-Fernández, F., & Rodríguez-Montequín, V. (2021). Determination of water depth in ports using satellite data based on machine learning algorithms. *Energies*, *14*(9). https://www.mdpi.com/1996-1073/14/9/2486

Yustisi Lumban-Gaol, K. A. O., & Peters, R. (2022). Extracting coastal water depths from multi-temporal sentinel-2 images using convolutional neural networks. *Marine Geodesy*, *45*(6), 615–644. https://doi.org/https://doi.org/10.1080/01490419.2022.2091696

Disanayaka Mudiyanselage, S., Abd-Elrahman, A., Wilkinson, B., & Lecours, V. (2022). Satellite-derived bathymetry using machine learning and optimal sentinel-2 imagery in south-west florida coastal waters. *GIScience Remote Sensing*, *59*, 1143–1158. https://doi.org/10.1080/15481603.2022.2100597

Zhang, J., Li, S., & Wang, M. (2020). Water depth inversion based on landsat-8 date and random forest algorithm. *Journal of Physics: Conference Series*, *1437*(1), 012073. https://doi.org/10.1088/1742-6596/1437/1/012073

Islam, K. I., Elias, E., Carroll, K. C., & Brown, C. (2023). Exploring random forest machine learning and remote sensing data for streamflow prediction: An alternative approach to a process-based hydrologic modeling in a snowmelt-driven watershed. *Remote Sensing*, *15*(16). https://doi.org/10.3390/rs15163999

Ai, B., Wen, Z., Wang, Z., Wang, R., Su, D., Li, C., & 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*, *PP*, 1–1. https://doi.org/10.1109/JSTARS.2020.2993731

He, C., Jiang, Q., Tao, G., & Zhang, Z. (2023). A convolutional neural network with spatial location integration for nearshore water depth inversion. *Sensors*, *23*(20). https://doi.org/10.3390/s23208493

Peakall, J., Ashworth, P. J., & Best, J. L. (2007). Meander-Bend Evolution, Alluvial Architecture, and the Role of Cohesion in Sinuous River Channels: A Flume Study. *Journal of Sedimentary Research*, *77*(3), 197–212. https://doi.org/10.2110/jsr.2007.017

Kleinhans, M. G., van Scheltinga, R. T., van der Vegt, M., & Markies, H. (2015). Turning the tide: Growth and dynamics of a tidal basin and inlet in experiments. *Journal of Geophysical Research: Earth Surface*, *120*(1), 95–119. https://doi.org/https://doi.org/10.1002/2014JF003127

Van Dijk, W., Lageweg, W., Kleinhans, M., Baar, A., & Rutten, J. (2014). Bank pull or bar push: What drives scroll-bar formation in meandering rivers? *Geology*, *42*, 319–322. https://doi.org/10.1130/G35192.1

Hoyal, D. C. J. D., & Sheets, B. A. (2009). Morphodynamic evolution of experimental cohesive deltas. *Journal of Geophysical Research: Earth Surface*, *114*(F2). https://doi.org/https://doi.org/10.1029/2007JF000882

Stefanon, L., Carniello, L., D'Alpaos, A., & Rinaldo, A. (2012). Signatures of sea level changes on tidal geomorphology: Experiments on network incision and retreat. *Geophysical Research Letters*, *39*(12). https://doi.org/https://doi.org/10.1029/2012GL051953

Leuven, J. R. F. W., Braat, L., van Dijk, W. M., de Haas, T., van Onselen, E. P., Ruessink, B. G., & Kleinhans, M. G. (2018). Growing forced bars determine nonideal estuary planform. *Journal of Geophysical Research: Earth Surface*, *123*(11), 2971–2992. https://doi.org/https://doi.org/10.1029/2018JF004718

Leuven, J. R. F. W., & Kleinhans, M. G. (2019). Incipient tidal bar and sill formation. *Journal of Geophysical Research: Earth Surface*, *124*(7), 1762–1781. https://doi.org/https://doi.org/10.1029/2018JF004953

Champion, B., Jamshidi, M., & Joordens, M. (2017). Depth estimation of an underwater object using a single camera. *KnE Engineering*, *2*, 112. https://doi.org/10.18502/keg.v2i2.603

Carbonneau, P., Lane, S., & Bergeron, N. (2006). Feature based image processing methods applied to bathymetric measurements from airborne remote sensing in fluvial environments. *Earth Surface Processes and Landforms*, *31*, 1413–1423. https://doi.org/10.1002/esp.1341

Tal, M., & Paola, C. (2010b). Effects of vegetation on channel morphodynamics: Results and insights from laboratory experiments. *Earth Surface Processes and Landforms*, *35*(9), 1014–1028. https://doi.org/https://doi.org/10.1002/esp.1908

Tambroni, N., Bolla Pittaluga, M., & Seminara, G. (2005). Laboratory observations of the morphodynamic evolution of tidal channels and tidal inlets. *Journal of Geophysical Research: Earth Surface*, *110*(F4). https://doi.org/https://doi.org/10.1029/2004JF000243

Kleinhans, M. G., Leuven, J., Braat, L., Markies, H., Simmelink, A., Roosendaal, C., van Eijk, A., Vrijbergen, P., & van Maarseveen, M. (2017). Turning the tide: Comparison of tidal flow by periodic sea level fluctuation and by periodic bed tilting in scaled landscape experiments of estuaries. *Earth Surface Dynamics*, *5*(4), 731–756. https://doi.org/10.5194/esurf-5-731-2017

Kleinhans, M. G., Leuven, J. R. F. W., Braat, L., & Baar, A. (2017). Scour holes and ripples occur below the hydraulic smooth to rough transition of movable beds. *Sedimentology*, *64*(5), 1381–1401. https://doi.org/https://doi.org/10.1111/sed.12358

Kleinhans, M., Leuven, J., Braat, L., van der Vegt, M., van Maarseveen, M., Markies, H., Roosendaal, C., & van Eijk, A. (2016). Reversing tidal flow and estuarine morphodynamics in the metronome laboratory flume. https://diamweb.ewi.tudelft.nl/pecs2016/abstracts/MaartenKleinhans.pdf

EFSA. (2013). Scientific opinion on the safety and efficacy of brilliant blue fcf (e133) as a feed additive for cats and dogs. *EFSA Journal*, *11*(7), 3288. https://doi.org/https://doi.org/10.2903/j.efsa.2013.3288

Leenman, A. S., & Eaton, B. C. (2024). Remote sensing of laboratory rivers. *Earth Surface Processes and Landforms*, *49*(1), 58–81. https://doi.org/https://doi.org/10.1002/esp.5577

Huang, M. Y., Huang, A. Y., & Capart, H. (2010). Joint mapping of bed elevation and flow depth in microscale morphodynamics experiments. *Experiments in Fluids*, *49*, 1121–1134.

Nota, E. (2024a). Personal communication.

Nota, E., Nijland, W., & de Haas, T. (2022). Improving uav-sfm time-series accuracy by co-alignment and contributions of ground control or rtk positioning. *International Journal of Applied Earth Observation and Geoinformation*, *109*, 102772. https://doi.org/https://doi.org/10.1016/j.jag.2022.102772

Nota, E. (2024b). Matlab code for dem development [Code used for digital elevation model development. Not publicly available. Contact the author for access.].

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

R Core Team. (2023). R: A language and environment for statistical computing.

Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*(1), 1–17. https://doi.org/10.18637/jss.v077.i01

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Nussbaum, M., Zimmermann, S., Walthert, L., & Baltensweiler, A. (2023). Benefits of hierarchical predictions for digital soil mapping—an approach to map bimodal soil ph. *Geoderma*, *437*, 116579. https://doi.org/https://doi.org/10.1016/j.geoderma.2023.116579

Nota, E. (2024c). Future paper on white balancing overhead cameras.

Pirzada, J., & Siddiqui, A. (2013). Analysis of edge detection algorithms for feature extraction in satellite images. *International Conference on Space Science and Communication, IconSpace*, 238–242. https://doi.org/10.1109/IconSpace.2013.6599472

Milan, A., Valadan Zoej, M. J., & Mokhtarzade, M. (2016). A new method for road detection in urban areas using high-resolution satellite images and lidar data based on fuzzy nearest-neighbor classification and optimal features. *Arabian Journal of Geosciences*, *9*. https://doi.org/10.1007/s12517-016-2374-1