

Predicting sedimentation in Lake Alajuela

*Constructing a machine learning pipeline to predict morphological behavior in
the man-made Lake Alajuela reservoir*

Lars Marinus Langhorst

Thesis to obtain the degree of Master of Science in Geomatics
Delft University of Technology
Delft, June 2022



DELFT UNIVERSITY OF TECHNOLOGY

THESIS TO OBTAIN THE TITLE OF

MSc. GEOMATICS

Predicting sedimentation in Lake Alajuela

Author:

Lars Marinus LANGHORST

Supervisors:

Ken Arroyo OHORI

Hugo LEDOUX

June 23, 2022

Lars Marinus Langhorst: *Predicting sedimentation in Lake Alajuela* (2022)

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

The work in this thesis was carried out in the:



3D geoinformation group
Delft University of Technology



Department of Cartography
Panama Canal

CANAL DE PANAMÁ

Supervisors: Ken Arroyo Ohori
Hugo Ledoux
Co-reader: Giorgio Agugiaro

Abstract

To keep the Panama Canal operational year round, the water level in the canal is maintained above the minimum depth. As water is passed through the locks and into the oceans on either side, the canals main reservoir loses fresh water, and this water is then replaced by the canals secondary reservoir which is Lake Alajuela. The Lake Alajuela reservoir is supplied of water by two large rivers, bringing in a sediment flux and decreasing the reservoirs capacity over time. Sedimentation sets a time limit to the use of the reservoir, and understanding this process and predicting the future morphological changes to occur in the reservoir are key to ensure continuation of operations in the Panama Canal.

To predict the quantities of local sedimentation in the reservoir, a machine learning model is trained with features extracted from bathymetric models covering various years. These features are computed values that correlate with morphological processes. As such, the adaptiveness of a machine learning model along with the ability to extract morphological features from any elevation model provides the possibility to predict future sedimentation where other methods with empirical and numerical methods do not suffice due to a lack of data and restrictive parameters.

Three types of machine learning algorithms and 21 features were initially tested for the purpose of the morphological modelling, during which the SVR model was the most successful. The testing was done on the Río Chagres basin as well as the Río Pequení basin, both located in Lake Alajuela. Extensive hyperparameter tests were done to optimize and further test the performance of the SVR. Depending on the study area and the scale of the morphological behavior occurring, slightly different sets of features were most effective. Nevertheless, in both study areas tested the runoff model has proven to be a key factor for predicting the sedimentation and achieved a 70 to 80% accuracy for predicting zones of low or high sedimentation.

The same model trained with the data from 1997 to 2018 was then used to predict 6 years into the future where more dynamic morphological behavior will occur according to the model. A downstream moving sediment front in the Río Chagres basin is recognized and predicted to have moved 500 meters between the years 2018 and 2024. Such a sediment front travelling too far into the reservoir could have catastrophic consequences for all operations of the Panama Canal, making the awareness and anticipation of its progression highly important.

Acknowledgements

I would like to thank my first and second supervisors Ken Arroyo Ohori and Hugo Ledoux for the feedback, guidance and insight throughout the entire process of this thesis. Additionally, I am thankful to the Cartography Department of the Panama Canal for the opportunity to perform such an influential and large scale research, and the support and information I have received from Fernando Bolivar, Jaime Rodriguez, and Javier Huertas.

Contents

1	Introduction	1
1.1	Objective	3
1.2	Scope	3
1.3	Data	4
1.4	Thesis outline	5
2	Background and Related work	6
2.1	Morphology	6
2.1.1	Sedimentation in Lake Alajuela	8
2.1.2	Current modelling methods	8
2.2	Machine learning	9
2.2.1	Machine learning and morphological parameters	10
2.2.2	Support Vector Regression	14
3	Methodology and Data	17
3.1	Data used	18
3.2	Programming details	21
3.2.1	FME	21
3.2.2	Unity	21
3.2.3	Python	22
3.3	Data preparation	22
3.3.1	Calibration	22
3.3.2	Thinning and triangulation	23
3.3.3	Clipping	23
3.3.4	Grid projection	23
3.4	Feature computation	23
3.4.1	Hydrological features	24
3.4.2	Geometric features	27
3.4.3	Temporal features	28
3.5	Model and feature selection	29
3.5.1	Feature selection criteria	29
3.5.2	Model selection	32
3.6	Hyper-parameter tuning	39
3.7	Future prediction in time steps	44
4	Results and Analysis	47
4.1	Accuracy analysis	47
4.1.1	Numeric	47
4.1.2	Probabilistic	50
4.1.3	Robustness to noise	51

4.1.4	Adaptability	53
4.2	Limitations of data availability	53
4.2.1	Years of data	53
4.2.2	Accuracy of DEM's	53
4.3	Comparison to common methods	54
4.3.1	Numerical models	54
4.3.2	Morphology ML prediction models	54
4.4	Model application	55
4.4.1	Hydrological analysis	55
4.4.2	Río Chagres	57
4.4.3	Río Pequení	60
4.5	Reproducibility	62
5	Conclusion and Discussion	63
5.1	Research overview	63
5.2	Contributions	65
5.3	Limitations	66
5.4	Recommendations and future work	67
A	What did not work	69
A.1	Planform and profile curvature features	69
A.2	MAT	69
A.3	Temporal features	69
B	Feature importances	70
C	Model testing results	72
C.1	Río Chagres	72
C.2	Río Pequení	77
D	Hyperparameter tuning results	83
D.1	Poly kernel degree tests	83
D.1.1	Río Chagres	83
D.1.2	Río Pequení	84
D.2	Rbf kernel C and epsilon tests	85
D.2.1	Río Chagres	85
D.2.2	Río Pequení	89

List of Figures

1.1	Satellite image in false colour of the Panama Canal, all water visualized as white	1
1.2	Satellite image of the Lake Alajuela watershed in false color, all water visualized as white	4
2.1	Copernicus satellite images of the Río Chagres basin in the dry season (left) and the rainy season (right) - (Buchhorn et al., 2020)	6
2.2	Common Backwater curves (Blom, 2021)	7
2.3	Old photograph of the creation of the Madden Dam taken in 1933	8
2.4	Parameters extracted from DEM by Asadi et al. (2021)	11
2.5	Parameters extracted from DEM by Asadi et al. (2021)	12
2.6	Improvement of predictions by Asadi et al. (2021)	12
2.7	Predictor grid median prediction error (Graw et al., 2021)	13
2.8	Support vector regression one dimension (Awad and Khanna, 2015)	14
2.9	Under and overfitting lines (Awad and Khanna, 2015)	15
3.1	Research Pipeline	17
3.2	Chagres river mouth in different datasets of different quality	19
3.3	Sedimentation levels of datasets	20
3.4	Visual representation of morphological changes in Lake Alajuela in the years 1997 to 2018	21
3.5	Averaged runoff scores	25
3.6	Length of flow path for dataset years	25
3.7	Distance to flow path for dataset years	26
3.8	Weighted Manhattan distance to river mouth for dataset years	26
3.9	Relative aspect angle to flow path for dataset years	26
3.10	Curvature on scales	28
3.11	Predictions for range of features at Río Chagres	30
3.12	Predictions for range of features at Río Pequení	31
3.13	RMSE values [m]	32
3.14	Prediction probabilities for Río Pequení	34
3.15	Prediction probabilities for Río Chagres	35
3.16	Radar graph of feature importances in percentage for both river basins	37
3.17	Prediction probabilities for Río Chagres using various kernels	39
3.18	Prediction probabilities for Río Chagres with rbf kernel and range of different parameters	40
3.19	Prediction probabilities for Río Chagres with poly kernel and range of different degrees	41
3.20	Prediction probabilities for Río Pequení with different kernels	42
3.21	Prediction probabilities for Río Pequení with rbf kernel and range of different parameters	43

3.22	Prediction probabilities for Río Pequení with poly kernel and range of different degrees	44
3.23	Data flow	45
3.24	Río Chagres prediction for 2018 using time steps	46
4.1	Histogram of errors made in prediction for Río Pequení	47
4.2	Histogram of errors made in prediction for Río Chagres	48
4.3	2D Histogram of errors made in prediction for Río Pequení	48
4.4	2D Histogram of errors made in prediction for Río Chagres	49
4.5	Indications of different gross errors	49
4.6	Prediction probabilities for Río Pequení	50
4.7	Prediction probabilities for Río Chagres	51
4.8	Overview of tests done with added noise	52
4.9	Sedimentation in Río Chagres Mouth in periods 1997-2008, 2008-2012, 2012-2018	56
4.10	Heights of Río Pequení Mouth in different years	56
4.11	Differences in height of Río Pequení Mouth between different years	57
4.12	Differences in height of Río Pequení extended into main reservoir between different years	57
4.13	The morphological changes that have occurred in the Río Chagres mouth and the predicted changes for 6 years after the last measured data	58
4.14	The scaled steps of that make up the Río Chagres prediction	58
4.15	Heights of Río Chagres predicted into future years	59
4.16	Differences in height of Río Pequení predicted for periods into future years	60
4.17	Heights of Río Pequení predicted into future years	61
B.1	Features ranked on importance for Río Chagres RFR prediction	70
B.2	Features ranked on importance for Río Pequení RFR prediction	71

List of Tables

2.1	Overview of Morphological modelling parameters often used	9
3.1	Overview of datasets	18
4.1	Noise test results	52
4.2	Volumes of sedimentation in Río Chagres basin	59
4.3	Volumes of sedimentation in Río Pequení basin	61

Acronyms

ML	Machine Learning
MLA	Machine Learning Algorithm
DEM	Digital Elevation Model
TIN	Triangulated Irregular Network
GIS	Geographic Information System
RF	Random Forest
RFR	Random Forest Regressor
SVR	Support Vector Regressor
SVC	Support Vector Classifier
SVM	Support Vector Machine
MLPR	Multi-Layer Perceptron Regression
ANN	Artificial Neural Network
MUSLE	Modified Universal Soil Equation
MLR	Multiple Linear Regression
SAR	Sediment Accumulation Rate
GP	Gaussian Process
PCA	Principal Component Analysis
ESVM	Evolutionary Support Vector Machine
MASE	Mean Absolute Scaled Error
RMSE	Root Mean Squared Error
LiDAR	Light Detection and Ranging
LoD	Level of Detail

Chapter 1

Introduction

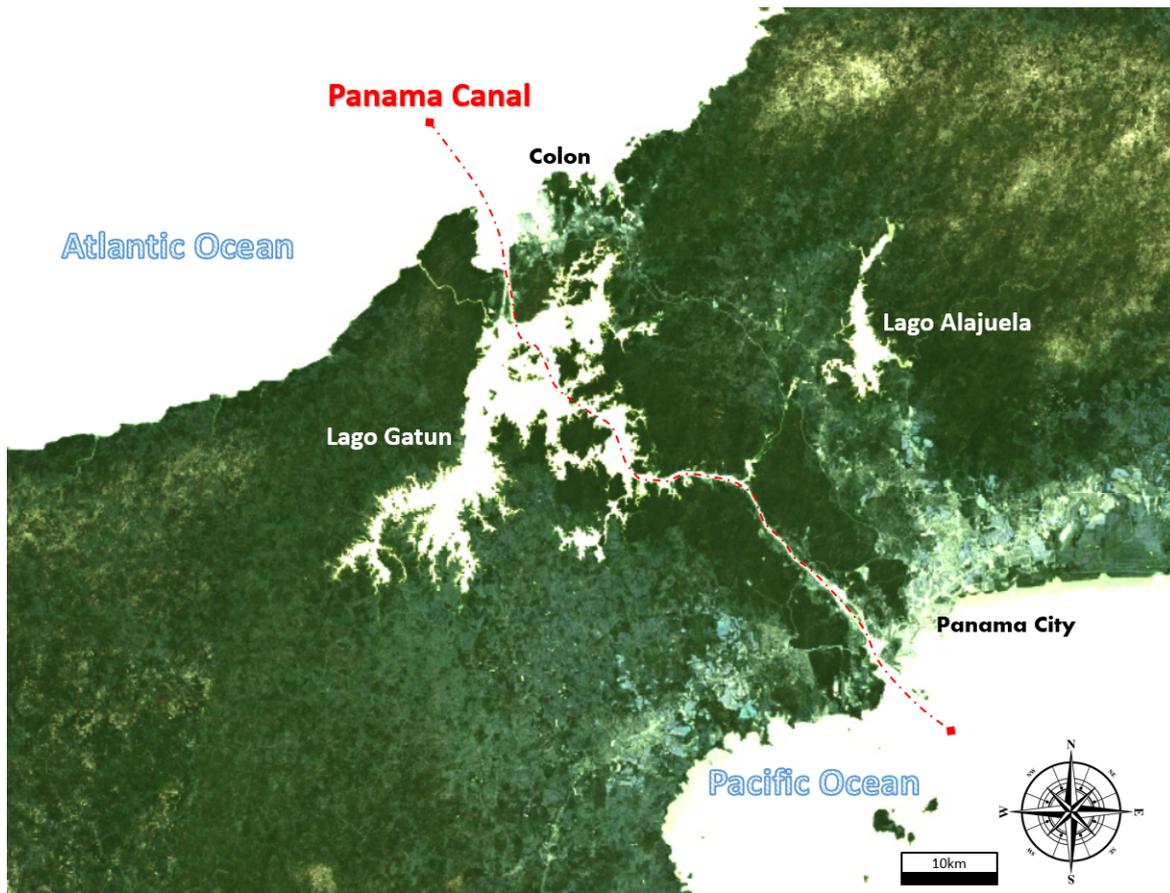


Figure 1.1: Satellite image in false colour of the Panama Canal, all water visualized as white

Lake Alajuela (formerly Madden Lake) provides the Panama Canal with water to maintain a minimum water level required for operation. The Panama Canal passes through the Gatun Lake which consists of a large man-made reservoir as well as dug out sections of the canal, nearly stretching from coast to coast. As the Panama Canal is constantly opening and closing its locks to let ships pass from the Atlantic to the Pacific Ocean and vice versa, the fresh water in the Gatun Lake is released into the oceans. Lake Alajuela is a man-made reservoir separated from the Gatun Lake by the Madden Dam, and provides the Gatun Lake with water when this is needed. Water enters Lake Alajuela through the Río Pequeñí and Río Chagres, and along with the water travels sediment which settles on the reservoir bed as it cannot be

flushed out. All entering sediment accumulates in the reservoir and has been decreasing the reservoirs storage capacity for nearly 100 years. An accurate prediction of the sedimentation process provides insight to the problem, for which a timely solution can be engineered.

"We might have trouble forecasting the temperature of the coffee one minute in advance, but we should have little difficulty in forecasting it an hour ahead." (Lorenz, 1963)

The modelling of sedimentation is a complex task, much like predicting the temperature of a cooling cup of coffee. The coffee eventually cools down to room temperature, when left alone for a sufficient amount of time. While it is cooling, thermodynamic vortices occurring in the coffee make the cooling process extremely difficult to predict. Similarly, without intervention and in enough time, any reservoir will fill up with the incoming flux of sediment. To predict the progression of sedimentation years or decades ahead is far more difficult, not due to thermodynamic vortices, but many other factors influencing this process.

Due to the multi dimensional nature of this complex process, current solutions are often simplified 1D or 2D models. These models utilize analytical methods based on physical processes, empirical methods based on scale experiments, and numerical methods in the form of complex computer programs that often require a vast amount of information and extensive tuning. Naturally, these traditional models are not suited for all scenarios and are limited by over-simplification. Hydrological parameters in combination with river cross sections are used in these models to compute morphodynamic behavior Williams et al. (2016) and may require other parameter values which are not available in all scenarios.

For the Lake Alajuela reservoir of the Panama Canal, several predictions have been made for the number of years until the reservoir becomes dysfunctional due to sedimentation (Loewenberg, 1999). After creation of the Madden Dam in 1935 and with it Lake Alajuela, it was estimated that the reservoir would have a 26% decrease of volume capacity in a period of 3900 years. This prediction was done with limited knowledge and information, no computers, and likely a large amount of political influence. A later study by Alvarado (1985) estimated a period of 100 years until a reservoir capacity decrease of 26%, ending in 2035. This broad estimate was done at a time in which deforestation was not controlled, and since this research many measures have been taken to decrease deforestation and consequently decrease the amount of sedimentation. Since sedimentation has decreased, no follow-up research has been done to predict the current rate of sedimentation in Lake Alajuela.

By training a ML model with historical data of the reservoir, local sedimentation can be predicted without requiring excessive amounts of data measurements, time and expertise. A Machine Learning (ML) model replaces the need for formulas, empirical relations, or other algorithms to be explicitly programmed into the model, as the model is self-learning and lays its own correlation by learning from a historical dataset. The historical data utilized are Digital Elevation Model (DEM)s over a range of 20 years, all having a similar resolution and area coverage. Geomorphological parameters like terrain slope, aspect, and curvature are extracted from the DEMs, and using hydrological data a full analysis and prediction can be made.

Machine Learning combined with DEM computations is currently being applied to many different types of problems, however in the field of reservoir sedimentation modelling there is still room for improvement (Alzaghoul et al., 2021; Costache and Bui, 2019; Avand et al., 2022). As information available per reservoir varies, current researched methodologies are often applicable to a single or specific type of reservoir, greatly limiting the usability of developed methods in other scenarios. This research is focused on combining ML and DEMs in order to construct a pipeline for predicting morpohological behavior without restrictions encountered by traditional methods such as the requiring of extensive calibration and hydrological datasets.

1.1 Objective

To fill the knowledge gap existing in the junction between machine learning and morphological models, the goal of this thesis is to develop a complete pipeline for predicting sedimentation with machine learning. For this purpose, the main research question is:

How to accurately predict sedimentation levels in the Lake Alajuela reservoir using a Machine Learning method?

The main objective of this research is the creation of a pipeline, providing predicted sedimentation levels of Lake Alajuela for future years including the periods around 2026, which is when the next planned bathymetric survey will take place. This requires results from a ML algorithm trained with features extracted from DEMs, which are analysed using hydrological data of the watershed and the dam like yearly precipitation levels, river discharges, and water passed through the dam. The predictions should be of sufficient accuracy to help engineers plan potential dredging operations by indicating the location and magnitude of sedimentation. To reach these objectives, the following sub-questions will be answered:

- *Which sedimentation related features can be extracted from the DEM?*
- *Which ML model best predicts sedimentation in a reservoir?*
- *What is the best set of geomorphological and hydrological features to train a ML model for prediction of sedimentation?*
- *What accuracy can be obtained predicting sedimentation in the Lake Alajuela reservoir?*

1.2 Scope

- *Primary area of study is the entrance of Río Chagres*
The Río Chagres is the largest source of water and sediment for Lake Alajuela, with strong morphological changes occurring in the river mouth.
- *Secondary area of study is the entrance of the Río Pequení into the main basin of Lake Alajuela*
The Río Pequení is the second large source of water and sediment into Lake Alajuela. The upper basin of the reservoir into which the Río Pequení flows will not be regarded in the research due to the fact that it lies nearly dry for part of the training data years, showing purely erosive morphological processes.
- *No deep learning algorithm will be included in the comparison between machine learning algorithms*
- *Data ranging from 1997 to 2018 will be used*
Data ranging back further than this are of poor quality and cannot be used for similar analysis to recent datasets.
- *No hydrodynamic or morphologic equations will be included in the model*
The Machine Learning Algorithm (MLA) will be used to lay connections between the features provided, not requiring the same data required for the traditional methods, nor necessarily using the same relations.

1.3 Data

In the research done during this thesis, data used was provided by the Department of Cartography of the Panama Canal. DEMs of the bathymetry of Lake Alajuela for the years 1997, 2008, 2012, and 2018 were used. These DEMs are point clouds of which the accuracy varies slightly between the years. Additionally historical local rainfall data as well as discharge data of water passing through the Madden Dam and river discharges of the Río Chagres and Río Pequení was made available.

The image in Figure 1.2 below shows an overview of the reservoir with different distinguishable sections. Bathymetric data of the entire lake was available for the provided years, with exceptions of some bays and shallow areas which are not present in all datasets.



Figure 1.2: Satellite image of the Lake Alajuela watershed in false color, all water visualized as white

The main areas of study in this thesis are the river mouth of the Río Pequení in the main reservoir and the Río Chagres river mouth as highlighted in Figure 1.2. In the North of the map in the figure the Río Pequení is shown which initially feeds water to the upper basin of the Lake Alajuela reservoir, and this water eventually ends up in the main basin. From the East of the watershed the Río Chagres enters the main basin, and this is the main supplier of water for the reservoir. The Madden Dam is located in the Southwest corner of the map and this dam is what created Lake Alajuela in 1935.

1.4 Thesis outline

This thesis is divided into 5 main chapters. Chapter 2 consists of the background and related work around the thesis subject. A short description is given of the working of morphological processes in rivers, and the models used to predict these processes. A background is given on machine learning models, and related work is provided of machine learning techniques being used for the purposes of predicting morphological processes.

Chapter 3 gives an overview of the data used in the research, the methodology and the implementation thereof. The full pipeline is explained and the details of the implementation in the use-case for Lake Alajuela are given. The results of the research are presented and discussed in chapter 4. Lastly, in chapter 5 the conclusions of the thesis are stated, and recommendations for future work given. Additionally, more elaborate ranges of test results and other information can be found in the Appendices.

Chapter 2

Background and Related work

2.1 Morphology

In all rivers and creeks eroded soil enters the water and flows downstream with the current. In places of reduced water flow velocity, this sediment then sinks to the river bed, causing the bed level to rise. This is a problem for all man made reservoirs, and thus many studies have been done on predicting the morphological response of lakes and rivers. Numerous variables influence the sedimentation process like the flow, relative pool height, sediment supply from upstream, and sediment size and distribution. External factors also have large influence on the sedimentation in watersheds, being factors such as the land use history, fires, and climatic cycles (Minear and Kondolf, 2009). These variables cause the process of sedimentation to be very complex and difficult to predict.



Figure 2.1: Copernicus satellite images of the Río Chagres basin in the dry season (left) and the rainy season (right) - (Buchhorn et al., 2020)

During the different seasons of the year, the amount of rainfall determines the flow of water and sediment entering the reservoir, as well as the amount of water leaving the reservoir. As a result the water level in the reservoir undergoes large changes, and the magnitude of these changes is visible in Figure 2.1. In the figure the same river mouth between the seasons at different water levels is practically unrecognizable, and thus the morphological behavior in these two scenarios is equally different.

Rivers are constantly undergoing morphological processes to obtain a balance between the supplied water discharge and the natural slope, width, depth of the river and many other factors. This balance can be used to analyse and predict morphological processes in the river.

Water flow is a determining factor in morphological processes, and flow velocity of the water in a river is influenced by a large number of dependent and independent variables. One of the main influences is the water depth which is influenced by what is called the backwater curve. The length of a backwater curve is determined by the depth and the slope of a channel, which in turn affects the flow velocity (Samuels, 1989). When looking at a section of a river to analyse morphology, the upstream and downstream boundaries must be defined. The water discharge as well as the in-flowing sediment flux are the upstream boundary parameters, and cannot be influenced by what happens downstream. The water level is determined by the downstream water depth, and can be predicted higher up the reach using backwater curves. (Blom, 2021; Mendoza et al., 2017)

Backwater curves determine the water level moving upstream which can be an increase or decrease in water level with respect to the water level at the downstream boundary. A backwater curve is created by the changing of a downstream boundary: the local water level, channel width, or change in channel depth. Such a change will then thus induce the backwater curve, changing the water level upstream and with that other flow characteristics and morphological processes that come along with it.

With an increase in depth the flow velocity tends to decrease resulting in sedimentation, and similarly when flow depth decreases flow speed will increase resulting in erosion thus increasing the water depth once again. These morphological responses bring the river back to an equilibrium, however a complete equilibrium will rarely be present due to the constant changing of parameters (Blom, 2021).

When a river ends in a reservoir, the backwater curve moves up from the mouth of the river. The water level in the reservoir is thus a deciding factor on erosion and sedimentation in the section of the river where the backwater curve stretches starting at the river mouth. The water level in a reservoir is constantly changing by nature, and in times of little rainfall the reservoir will be decreasing in water level as the inflow of water will be less than the outflow and at the same time lowering the water level at the mouth of the river. In months of high precipitation the water level in the reservoir will stay around the same point, being the maximum water level for the dam. These boundaries are thus changing constantly year round, and the different seasons will have a different effect on the morphological processes in the river and the river mouth.

If the water level of the reservoir decreases more than that of the river, an M2 backwater curve will occur as seen in Figure 2.2 as the equilibrium water depth d_e , being the depth of the reservoir at the river mouth, is above the actual water depth d , being the depth of the upstream boundary of the river. The M2 curve will then cause a lowered water level upstream and an increase in flow velocity over a large stretch of the river. With the higher flow velocities, the river bed will be eroded and this eroded sediment gets carried by the water until the flow speed once again decreases. As the river flows into the reservoir, the flow speed eventually comes to a full stop resulting in all sediment to be deposited.

When the reservoir water level increases more than the water level of the river, an M1 backwater curve is created, also seen in Figure 2.2 as d_e is below d . This curve increases the depth along an upstream stretch in the river decreasing the flow velocity, and thus causing the sediment carried by the water to be deposited on the river bed.

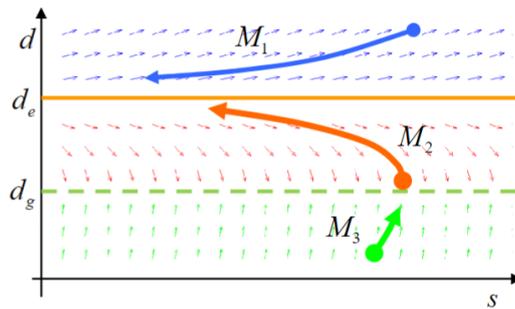


Figure 2.2: Common Backwater curves (Blom, 2021)

All in all, the seasons and the changing water level trigger constant changes in the morphology of the river mouths. River mouth processes are influenced by a large number of factors both from the water supplying river and the basin the river feeds into, river mouths extremely varied coastal accumulation forms. (Wright, 1977) The constant changing of reservoir depth thus causes the sediment to be deposited in the river mouth at different times. Extreme weather scenarios can cause greater amounts of sediment to be transported or eroded from the river. such event will then not only affect the sediment transported at that time, but will also influence the sediment supplied to the reservoir in future years.

2.1.1 Sedimentation in Lake Alajuela

Since the completion of the Madden Dam in 1935, the Lake Alajuela reservoir provides the Panama Canal with water to keep the water level sufficiently high for ships to pass through the most shallow points of the route, being the locks. During the dry season, there is insufficient rain to compensate for the water used by the canals locks, and the Lake Alajuela reservoir provides the required water for the continuation of operations in the canal. The amount of water passing from the Lake Alajuela reservoir to the Panama Canal is regulated by the Madden Dam, and the reservoir receives water from 2 main rivers, being the Río Chagres and the Río Pequeñí.



Figure 2.3: Old photograph of the creation of the Madden Dam taken in 1933

The Río Chagres is the largest river providing feeding into Lake Alajuela and provides the highest water discharge, along with the largest amount of sediment. The majority of sediment transported to the reservoir is deposited in the mouth of the river due to the sudden change of width of the channel. Especially during the rainy season large amounts of sediment are deposited near the mouth of the Río Chagres, and during the dry season with lower water levels, the sediment is carried further into the reservoir.

The Río Pequeñí first flows into the upper basin of the reservoir which is separated from the main basin of the reservoir by a narrow passing. This results in the majority of sedimentation first ending up in this upper segment, and then gradually being transported downstream into the lower reaches of the reservoir. Besides these two main rivers, several small streams and slopes are susceptible to landslides and other events that may also result in sediment deposits in the reservoir.

2.1.2 Current modelling methods

In present day, with powerful Geographic Information System (GIS) software and computational power, accurate and extensive predictions can be made using large numbers of parameters and data sets. In the past, empirical models were used to predict morphological behavior in rivers and channels, such as the sediment transport relations developed by Engelund and Hansen (1967), Meyer-Peter and Müller (1948) and Exner (1920). Such empirical methods will generally underestimate the sedimentation due to oversimplification (Idrees et al., 2021). Table 2.1 gives an overview of a range of parameters. commonly used in existing modelling methods along with the typical morphological response that comes with an increase of this parameter.

Morphological modelling parameters		
Parameter	Unit	Effect
Channel width	[m]	A larger channel width will cause a decrease in flow velocity resulting in sedimentation
Bed level height	[m]	An increase in bed level height may cause a decrease in water depth resulting in erosion
Chezy coefficient	[$m^{1/2}/s$]	A higher Chezy coefficient signifies more bed friction resulting in a decrease in flow velocity and sedimentation
Sediment flux	[kg/m^3]	more sediment flux results in water not needing sediment from the bed, and too much sediment supplied from upstream means the river deposits to match sediment transport capacity
Sediment particle diameter	[m]	Smaller particles require less flow velocity to be carried by the river, and thus smaller particles will travel further than the large grains
Porosity of reservoir bed	-	The porosity determines the amount of fine sediment that can settle in the bed material
Channel slope	[deg]	A larger slope results in higher flow velocities and bed erosion
Downstream water depth	[m]	The downstream water depth is a boundary condition determining the water depth upstream in a channel as well as the flow velocity, and thus influencing sedimentation or erosion

Table 2.1: Overview of Morphological modelling parameters often used

These parameters always contain margins of error and are not always available for reservoirs as they either have to be obtained with specialised equipment, or estimated through iterative processes. Additionally, the inaccuracy that comes with the methods of acquirement for the necessary data these methods use affects the accuracy of tasks depending on these results (Stefanyshyn et al., 2021). Although these analytical models are easy to interpret as they are based on physical processes, the models cannot always be applied due to lack of data or knowledge.

2.2 Machine learning

Machine Learning (ML) models are self teaching algorithms that can be used to find correlation in data without interventions necessary. Their popularity is growing as computers become faster, algorithms improve, and information is more available than ever. ML models can, in some cases, provide high accuracy predictions for natural processes that are otherwise complex to model. There are two types of ML models; regression models and classification models. When using numerical values and requiring a numerical outcome, regression models are used. A Random Forest (RF) model consists of tree-structured predictors called regression

trees, each of these constructed with random selection and order of features. The Random Forest Regressor (RFR) builds a K number of regression trees averages the result (Segal, 2004; Rodriguez-Galiano et al., 2015).

The RF regression model is a popular candidate due to its simplicity and the fact that it required limited effort to tune the model. Mitchell et al. (2021) compared the RFR model to linear interpolation for spatial Sediment Accumulation Rate (SAR) predictions. In spatial predictions the RF provides far more accurate predictions, however Mitchell et al. (2021) mentions that the RF may not be the best suited regressor for spatial prediction with SAR data as the averaging of data between trees results in predicted values converging within the range of observed values. This study by Mitchell et al. (2021) provides an insight into the possibilities using ML models, specifically the RFR model, with sedimentation data, although the predictions made are on a spatial scale and not a temporal scale. Mitchell et al. (2021) was able to predict the SAR to relatively high standards, however the restriction to the random forest regression model and the lack of filtering and selection of features hampered the results. In this thesis the aim is to predict sedimentation on a temporal scale, and multiple MLAs are tested as well as a wide range of features tested, filtered and selected to gain optimal accuracy.

Artificial Neural Networks (ANN) like Convolutional Neural Networks or Recurrent Neural Networks often provide good replacement for traditional models when used for the right purpose with the right parameters. Compared to other ML algorithms, the Artificial Neural Network (ANN) provides strong predictions using hydrological parameters and sediment inflow data (Idrees et al., 2021). EL Bilali et al. (2020) compares the ANN to a modified Universal Soil Loss Equation coupled with a multiple linear regression (MUSLE-MLR) model for predicting reservoir sedimentation trained with sediment yield data and physical characteristics of the watershed, then validated with hydrological data. The ANN provides higher accuracy on reservoir sedimentation predictions compared to the MUSLE-MLR. One of the main advantages of the ANN in this comparison is that it does not require the extensive calibration as is required for the MUSLE-MLR method. The comparison made by EL Bilali et al. (2020) shows the room for improvement in this field of research, however the predictions made are on a general scale computing for total loss of capacity, using only the water inflow and outflow as well as the initial reservoir area as parameters for the ANN thus not taking into account the local characteristics of the terrain. This thesis aims to predict the morphological behavior on a far smaller scale being local predictions, and thus the comparison made by EL Bilali et al. (2020) does not carry over entirely.

2.2.1 Machine learning and morphological parameters

By using geomorphological as well as geo-environmental parameters, Rahmati et al. (2017) found that the RF and Support Vector Machine (SVM) models showing best performance for predicting erosion on for small man-made reservoirs. These models were found to give the best performance and provide robust predictions under a change of the sample data set, showing the models capability of predicting the effects of morphological processes using geomorphological parameters. Rahmati et al. (2017) also states that the RF and SVM models provide sufficiently accurate predictions for the assessment of gully erosion. This thesis aims to complete a very similar task, however the problem will be a regression problem and the main morphological process to be assessed is underwater sedimentation, where in the case researched by Rahmati et al. (2017) this was erosion.

Asadi et al. (2021) used geomorphological and river discharge information as parameters for six different ML algorithms to predict suspended sediment load in rivers. Principal Component Analysis (PCA) is used to select optimal independent features. Gaussian Process (GP)

and Evolutionary Support Vector Machine (ESVM) showed the highest accuracy for prediction of suspended sediment load on a basin scale. The geomorphological parameters used in the prediction were profile curvature, LS factor (Slope Length and Steepness factor (Panagos et al., 2015)), longitudinal curvature, flow accumulation parameters, stream power index (a measure of the erosive power of flowing water), Strahler order (mathematical system for ordering streams (Melton, 1959)), aspect, and vertical distance to channel network. These parameters can be seen in one of the studied sub-basins in figures 2.4 and 2.5.

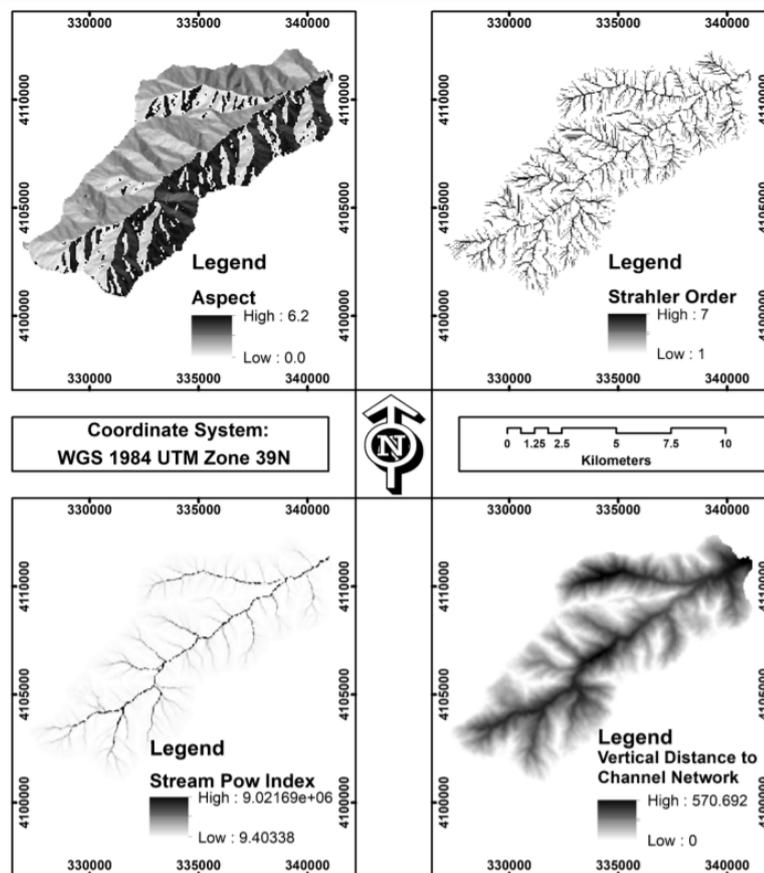


Figure 2.4: Parameters extracted from DEM by Asadi et al. (2021)

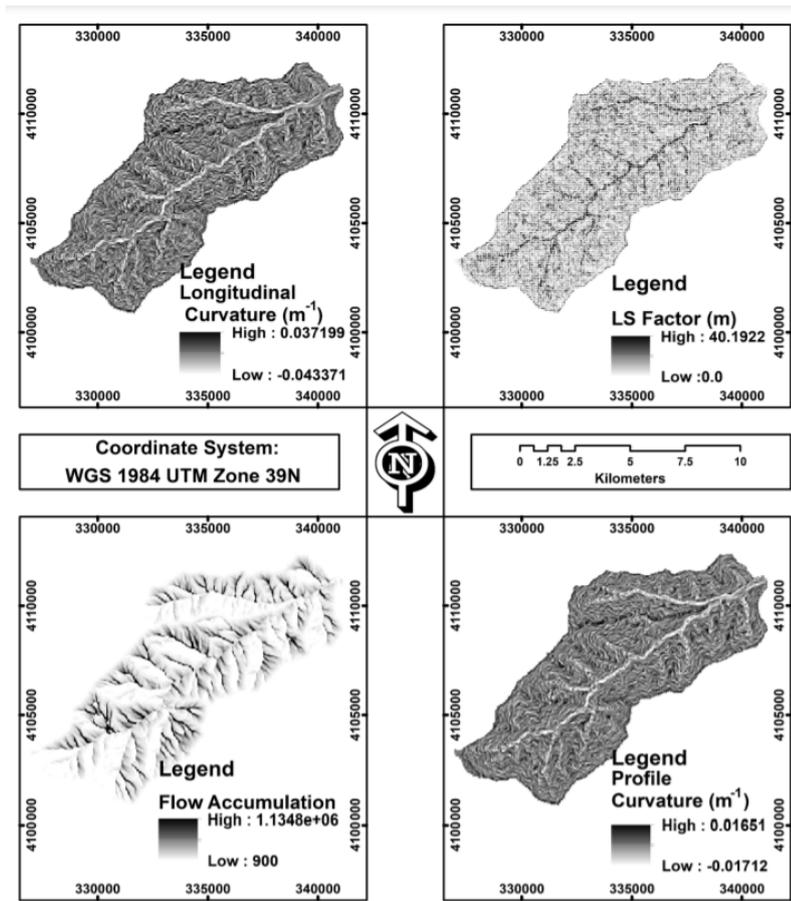


Figure 2.5: Parameters extracted from DEM by Asadi et al. (2021)

As seen in figures 2.4 and 2.5, there are features that correlate to the distance from the valley bottom (e.g. Stream Power Index, Vertical Distance to Channel Network) while other features (Curvature) show correlation to the local geographic properties. The geomorphological features used by Asadi et al. (2021) will be trialed in this research along with other features, however Asadi et al. (2021) studied the relation of these features to erosion, a morphological process different than that taking place under the water surface in reservoirs.

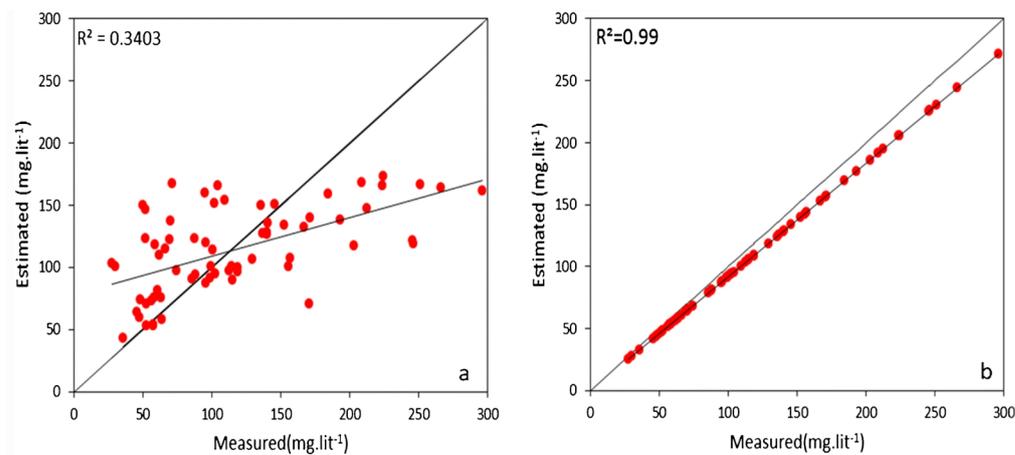


Figure 2.6: Improvement of predictions by Asadi et al. (2021)

The Figure 2.6 shows the improvement of the predictions of suspended sediment load in various locations in a watershed. The left plot shows the predictions made by solely correlating the stream discharge to the suspended sediment load, while the accuracy in the right figure shows the prediction accuracy obtained using the geo-morphological features shown in figures 2.4 and 2.5. This thesis aims to perform a prediction using geo-morphological parameters where most existing methods use mainly hydrological parameters. Asadi et al. (2021) has succeeded in a very similar field, showing there is plenty room for improvement when it comes to understanding and predicting morphological behavior and the use of machine learning models for this purpose.

Jagers (2001) has stated that for the purpose of morphological predictions, the positive and negative aspects of using neural networks are opposite of using 2D or 3D mathematical models. This is mainly regarding the easy-of-use versus the required data, and demonstrates that ML algorithms have the potential to compensate where the mathematical models are systematically lacking. A study on predicting sediment density with RF by Graw et al. (2021) describes the advantage of binning observations together in grid cells of pre-defined size. The training procedure will return inflated correlation coefficients if the observations are not binned properly.

Additionally, Graw et al. (2021) places each selected feature on a individual predictor grid. The median prediction error is then recorded per grid to validate the feature on that grid. The median predictor errors per grid are shown in Figure 2.7. The figure shows the correlation of the individual grids to the prediction error, along with the ensemble grids. As some individual grids obtain higher error than others, the ensemble grids generally decrease in error as the amount of predictors increases. Above a certain number of predictors however, the error does not decrease significantly as is seen in the circled area in Figure 2.7. Random noise grids are added to establish a maximum median error for the predictor grids.

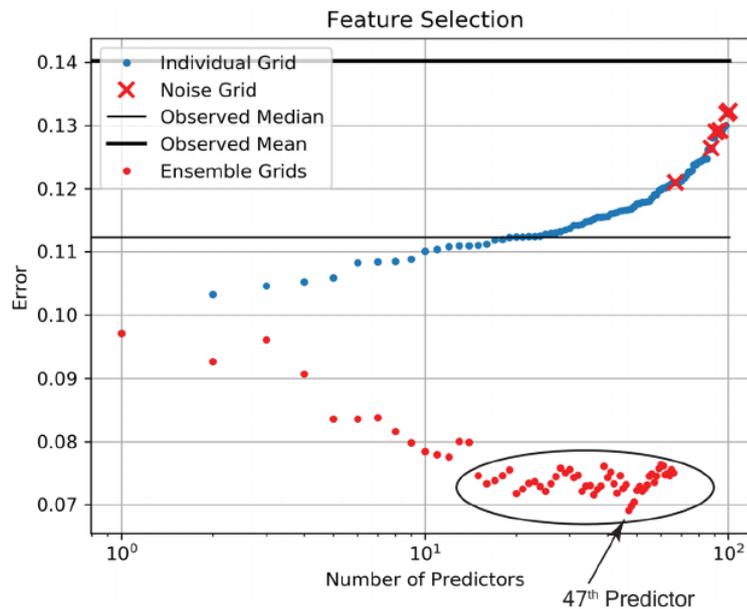


Figure 2.7: Predictor grid median prediction error (Graw et al., 2021)

Selecting the most relevant features in this manner is especially useful when using RF Regression models due to the averaging between trees (Mitchell et al., 2021), as well as Support Vector Machine models (Sahoo et al., 2021). In the case of machine learning problems like morphology models where many parameters affect the dynamics of the situation in sometimes

unknown ways, the selection of the most effective feature set is crucial as the most important features are unknown beforehand.

2.2.2 Support Vector Regression

Support Vector Regressor (SVR) is a machine learning regression methodology in which a symmetrical loss function is used for penalizing all wrong estimations. A function is created to match data points, and an area of margin with a minimal radius is created around the function. When any data points fall outside this radius of margin, these points are penalized, and points within the margin are not. The error is absolute, meaning the location of the points relative to the function does not matter. One of the main advantages of the SVR is its ability to generalize problems, while being able to predict with high accuracy. Additionally, the computational complexity of the algorithm is independent of the amount of dimensions provided with the input space.

As where the Support Vector Classifier (SVC) solves a classification problem providing an output from a predefined set, the SVR is a generalization thereof, returning a continuous valued output. Awad and Khanna (2015) In the Figure 2.8 below, a one dimensional SVR example is shown.

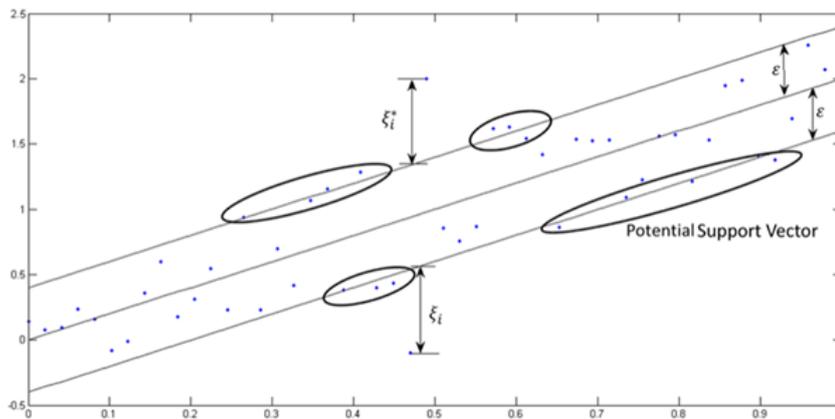


Figure 2.8: Support vector regression one dimension (Awad and Khanna, 2015)

The continuous valued output is computed by a function approximation, produced by the SVR as an optimization problem where the model searches for the function fitting the data points with margin on the smallest radius, and at the same time minimizing the prediction error. The image below demonstrates the effect of under and overfitting a dataset using orders of polynomials. Awad and Khanna (2015)

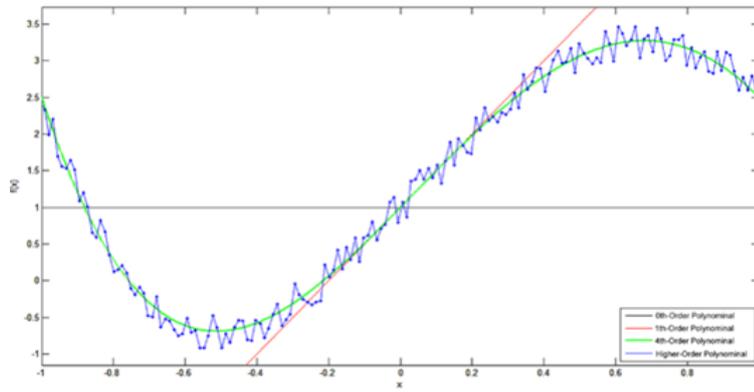


Figure 2.9: Under and overfitting lines (Awad and Khanna, 2015)

The Figure 2.9 shows three lines of various order polynomials. The 0th order polynomial is a function of y having only one value, and thus being a horizontal line. The 1st order polynomial is a straight line with an angle, thus already being able to adapt somewhat to obtain the lowest error relative to the data points. The 4th order polynomial produces a highly accurate prediction while maintaining a smooth profile. The higher order polynomial passes through all data points and thus does not contain any error. When the pattern of the dataset is continued and new points are added however, the 4th order polynomial will have a high chance of predicting with a small error while the higher order polynomial will likely exclusively work on the original dataset, thus overfitting the data Awad and Khanna (2015).

To compute smooth solutions and prevent overfitting, weights are added to the vectors in the SVR which then act as regularizing terms controlling this smoothness, and essentially the order of the solution. To minimize the prediction error, the SVR uses an ϵ -insensitive loss function which penalizes the larger errors. The radius for the margin is determined by the value of ϵ . A smaller radius means more support vectors as the support vectors are those that fall outside of the margin. The smaller value of ϵ with the smaller margin thus also gives a smaller tolerance for error. When working multidimensional data with nonlinear functions, the data is mapped into kernel space. Awad and Khanna (2015) The kernel of an SVR model is the function which transforms the space if the data the model is trained with. The new transformed space contains cores which are then used to fit the best function to fit the model. There are different types of kernels, each with different coefficients, and thus different kernels will perform stronger on different datasets. (Huang et al., 2021)

To evaluate forecasting models, four types of error metrics can be distinguished: scale dependent metrics, percentage-error metrics, relative error metrics, and scale-free error metrics (Hyndman et al., 2006). Depending on the data and the model used to make the forecast, an error metrics must be chosen to best fit the required situation. Hyndman et al. (2006) suggests that the Mean Absolute Scaled Error (MASE) is the best accuracy metric since it is the only accuracy measurement that can be used for all forecast methods and types of series.

Current mathematical models for predicting sedimentation are heavily dependent on flow and sediment related hydrological parameters. To prevent the necessity of this information and the required expertise needed to adapt and analyse such models, a ML model can be used to find these relations from the supplied data to make these predictions. To date, the majority of sedimentation research done using ML adopts hydrological and geological parameters such as soil types, suspended sediment values, and flow velocities. Predictions of morphological processes using ML techniques are currently not implementing time steps. The use of time steps in combination with ML is something that will be trialed in this research, to attempt to expand the range of possibilities of the ML model for morphological predictions.

There is a lack of knowledge on the use of geomorphological parameters with ML algorithms as a lot of computational effort is required for such solutions at the moment, as the state of art is developing. This thesis is focussed on specifically that, appending to the current knowledge available. As previous estimations of sedimentation levels in Lake Alajuela are very global and done without the available DEMs and computational possibilities presently available, this thesis provides new insight into the sedimentation occurring in Lake Alajuela, as well as predictions of local morphological changes.

Chapter 3

Methodology and Data

This section will provide an overview of the methodology required to predict morphological changes in rivers or lakes using a machine learning algorithm and historical bathymetry data. In Figure 3.1 the pipeline is presented, and details per step will be provided in the sections of this chapter

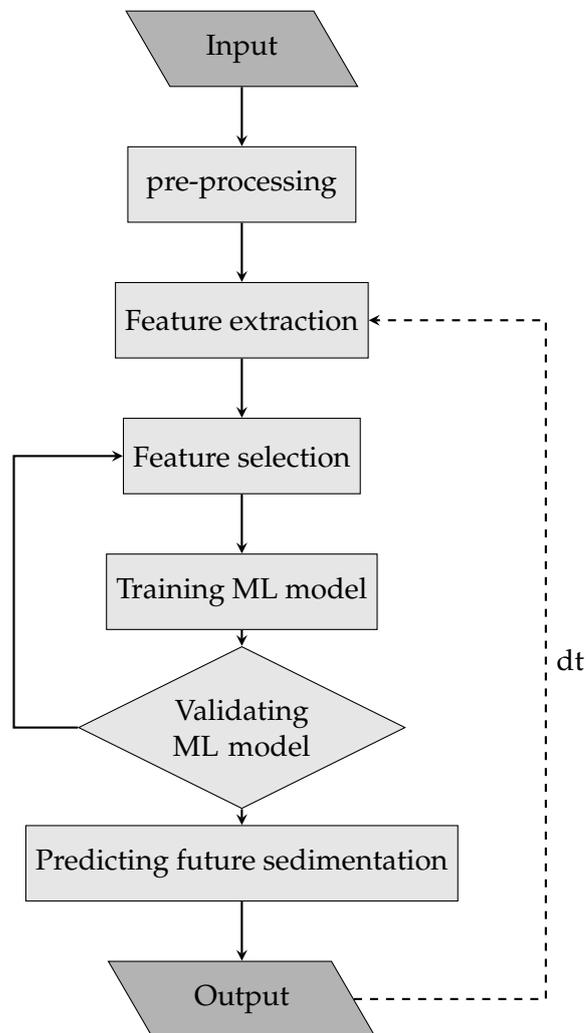


Figure 3.1: Research Pipeline

3.1 Data used

A range of years in historical bathymetric models and hydrological datasets were used for this research. These were point clouds of various resolutions, with the time of measurement ranging between 1928 and 2018. The decision was made to utilize the datasets of the years 1997, 2008, 2012, and 2018 since the earlier datasets were of much poorer quality than the later pointclouds. The different datasets were also gathered with different equipment, and processed with the use of different software and pipelines. The earliest model was made of the terrain when there was no water yet in the lake, and thus measured by hand having a large margin of error. The 1983 dataset was made with single beam sonar equipment, but of very low processing quality. The 1997 and 2008 datasets were also measured with single beam echo sounding, however these had to be backwards engineered from triangulations which have resulted in the estimated margin of error to be between zero and 0.3 meters. The 2012 and 2018 datasets contained the bathymetric measurements made with high precision multi beam sonar equipment, and this data could directly be implemented into the pipeline giving the models a much smaller estimated margin of error. All datasets were shifted vertically slightly due to realignment, however this is believed to have left a maximum error of 0.1 meters. The table below provides an overview of the datasets that were available:

year	type	resolution [m]	Level of Detail	Margin of Error [m]
1928	TIN	unknown	Very Poor	unknown
1983	TIN	unknown	Poor	unknown
1997	TIN	1.0	High	0.3
2008	TIN	1.0	High	0.3
2012	Bathymetry	3.0	High	0.1
2018	Bathymetry/LiDAR	3.0	High	0.1

Table 3.1: Overview of datasets

The sediment front entering the basin is moving inward and can be clearly identified in the images of Figure 3.2, at the top of every image. Even though the start of the sediment front is visible in the 1983 dataset as well, the level of detail is not sufficient to distinguish the scale of morphological features that can be seen in the later datasets. The resolution of the 1928 and 1983 datasets are unknown, however as can be seen in Figure 3.2 the difference between 1983 and the later years is substantial. As the bed of the lake is covered in humps, pits and gulleys in the data from 1997 to 2018, the same area in the 1983 dataset contains visually distinguishable triangles. The resolution of the datasets in 1997 and 2008 are higher than that of 2012 and 2018. The difference in this resolution however is not of importance as the morphological are of a larger scale and are equally distinguishable in both resolutions, thus meaning these are of the same level of detail.

The two most recent datasets were provided with the original bathymetric readings, whereas the 1997 and 2008 datasets had already been triangulated. The triangulated data has the advantage of being pre-filtered, processed and ready to use, however it is not possible to trace back artifacts or locations of interest in the original measurements.

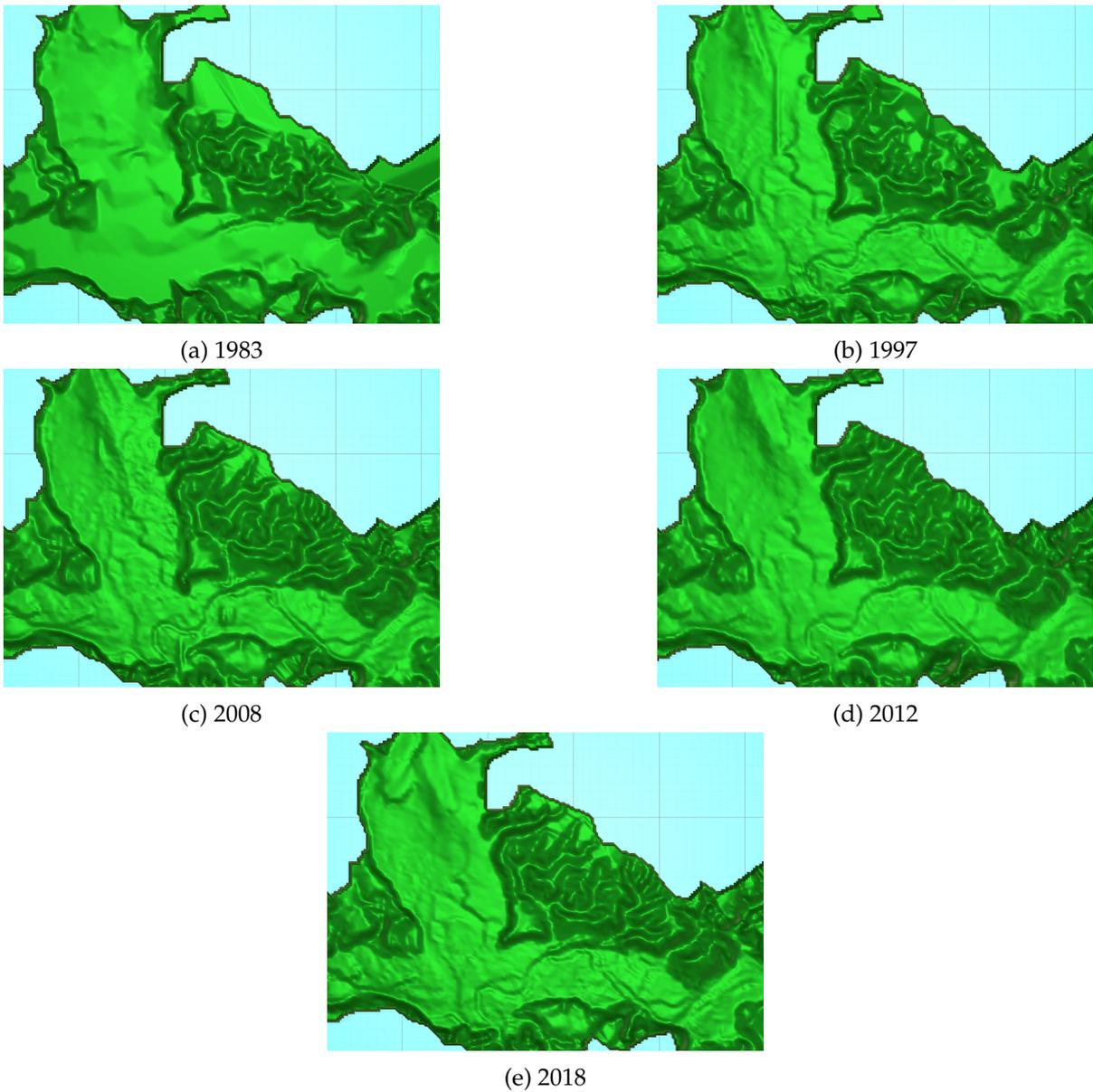


Figure 3.2: Chagres river mouth in different datasets of different quality

Average yearly water discharge in the inflowing rivers as well as the water discharge passing through the dam, being the outflow of water from the lake were used for analysis and prediction of sedimentation. In the boxplots below, the collected values of morphological changes occurring in different areas of the reservoir are shown, as well as that of the total reservoir. The data points used for these boxplots are the interpolated points from the DEMs of the various years. The interpolated points all fit onto the same grid, and thus the differences between the years show the morphological changes.

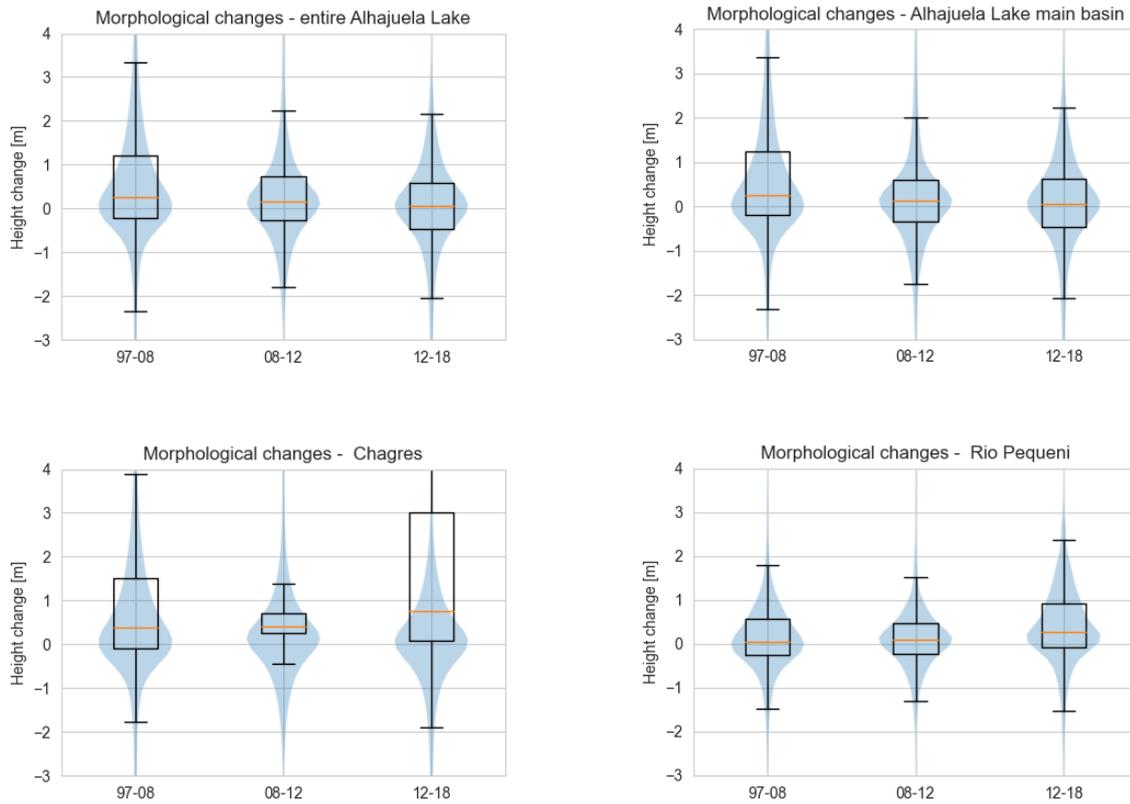


Figure 3.3: Sedimentation levels of datasets

As seen in Figure 3.3, around the river mouths the majority of morphological activity is sedimentation, as the average height is increasing. The morphological activity in the total reservoir is centered around zero which should be expected. The lake is sufficiently large to prevent any sedimentation from raise the average morphological activity of the lake visible in a plot showing the condensed data on a small scale. The violin plot for the Río Chagres shows a second large peak in sedimentation values round 1.5 meters. The changes in between these datasets have passed over 11 years, and the uneven distribution of the sedimentation in the is likely due to several heavy erosive events occurring upstream causing waves of sediment to flow downstream in different magnitudes. The rest of the violin plot, centered around zero are the more mild processes occurring in all recorded years.

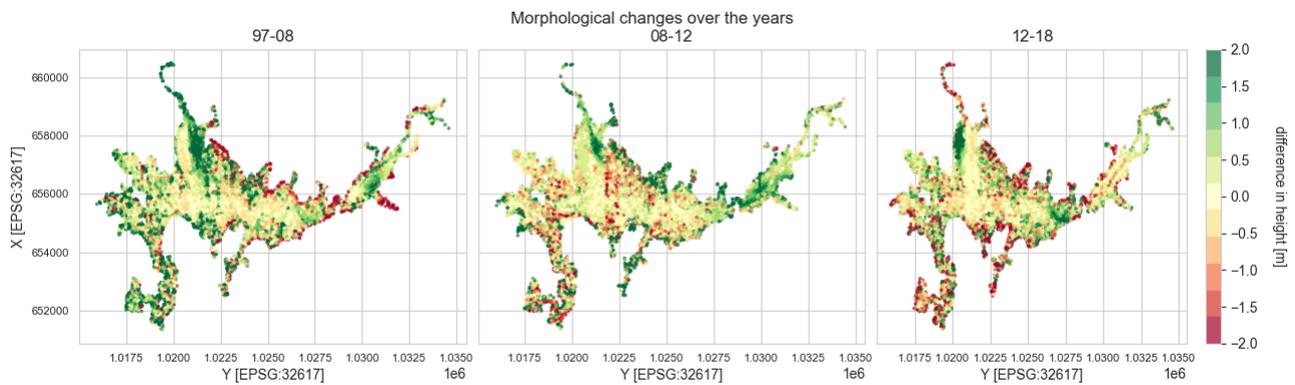


Figure 3.4: Visual representation of morphological changes in Lake Alajuela in the years 1997 to 2018

As shown in Figure 3.4 the edges of the lake contain a large amount of strong morphodynamic behavior. These edges are prone to mudslides and erosion, but are also the areas more likely to contain measurement errors due to shallow and muddy water and vegetation. The 2018 dataset additionally contains Lidar measurements along these shallow edges, which the other datasets do not have. This may be a factor causing differences between the years of 2012 and 2018. The deepest area of the lake, being its center, has little to no changes in height, and thus little to no morphological activity. This area is removed far from all gullies, creeks, and other activity near the shores, and has a relatively stable height with very little water flow.

3.2 Programming details

This research was done in three different programs. All code for the Python machine learning pipeline and the C# feature extraction pipeline are available at: github.com/LMarinusL/MorphologyPredictor. This section will give an overview of the tools used and how the tools were utilised.

3.2.1 FME

FME was used for data preparation and interpolation. The data sets of all years were clipped and interpolated to a grid of fixed dimensions. This pipeline in FME essentially removes all small noise from the grid due to the Triangulated Irregular Network (TIN) that is created with a certain ambiguity. All files are then sent to C# in this fixed grid and format, allowing the C# code to work for all data sets without needing altered parameters when changing data set.

3.2.2 Unity

Unity works with C# code with which the grids from FME could be imported.

Mesh

Unity contains a Mesh class which is very useful for visualising results for the purpose of debugging, but also analysis of intermediate results. When a mesh is initialized, the vertices used in the mesh are automatically assigned normal vectors. These normal vectors can then be used to compute the slope and aspect of the vertices.

3.2.3 Python

The machine learning phase in the pipeline is all done using python, with various libraries. For all machine learning functionalities, the scikit-learn library was used. This includes the training, testing, and validating of various machine learning algorithms. Additionally, scikit-learn contains functionality for analysing results, hyper-parameter tuning, pre-processing, pipeline construction, error metrics, and feature importance computations. Pandas was used to read the imported file and transform the original text file into a directly usable database with column indices. The data was then organized with pandas, and with the indices specific columns are easily extracted and used in different prediction, training, and visualisation functions in the python code. The Matplotlib library allows for custom plots and graphs for visualising results. For on the spot computations and corrections to the data, Numpy library was used.

3.3 Data preparation

The supplied data has already been processed thoroughly by the cartographic department of the Panama Canal. The provided pointclouds are thus absent of noise, and could directly be used for further steps in the data preparation process. The TIN datasets came in the form of dense pointclouds, which needed to be thinned and interpolated to TINs again, backwards engineering the data to the FME TIN structure. The further preparation of data was done in FME. All datasets undergo the same process of calibration, thinning, triangulation, clipping and interpolating to a grid. This way all datasets contain points on the same locations, allowing for direct comparison. These steps will now be explained in further detail.

3.3.1 Calibration

Using data sets that date back more than 20 years in time, there are misalignment issues can cause issues especially when analysing the morphological changes that have taken place. On the horizontal plane, all datasets are accurately alligned, however there is a large difference in vertical allignment between the individual datasets as well to the actual height in the area. To calibrate the datasets, different methods were trialed and compared.

Points of reference

First, multiple points were chosen in the lake of which the suspected difference in elevation over time was estimated to be minimal. Such points were located near docks, where boats must always be able to pass with a minimal draft, and also these points need to be far away from the river mouths feeding water and sediment into the lake.

Center of lake

The methodology used by the Cartography Department of the Panama Canal suggests that the least amount of sedimentation occurs at the center of the lake. Here the depth is greatest, and the change of sediment sinking and settling here is the lowest due to the distance from the mouths of the rivers.

Satelite Lidar reference

An attempt was made to calibrate the datasets using Lidar measurements from the open source ICESat database. This dataset contained relatively little points along the lake however,

and since solely the 2018 dataset contains Lidar points of the surrounding area, the other years could not be properly calibrated using the ICESat reference points.

Method used

To obtain the morphological changes being the change in height between the datasets, the calibration of the datasets to the real world is not relevant. The datasets were ultimately calibrated using areas in the center of the lake, however solely areas that were seen to undergo very little change in morphology. That means that humps, pits, and ridges have not changed in these areas in 20 years. All cells within these areas were then compared between the years, and the datasets were calibrated to result in minimal change of height. This means different points were chosen where little morphological change is expected, and all cells in a radius of those points are compared between the years and then calibrated to result in the smallest possible difference.

To validate the changed calibration, the morphological changes in the terrain along with the hydrological data were analysed, and the conclusion was made that the new calibration provided far more realistic results in both the areas of study as the other distinct areas of the reservoir such as the basin in front of the dam and the Salamanca area containing a harbour.

3.3.2 Thinning and triangulation

Depending on the dataset, a certain amount of thinning was done to make the dataset workable while staying as true as possible to the original values. The datasets that came in the form of TIN's were extremely dense point clouds with faces of points representing the faces of the triangulation. These datasets underwent a thinning process, after which a TIN was made using FME, of which the result contained the same shape as the original pointcloud, but with a far more usable size and format. The datasets containing the bathymetric single-beam measurements did not undergo the thinning process, since these point clouds did not contain added points like the TIN pointclouds. These bathymetric measurements were then triangulated as well.

3.3.3 Clipping

All datasets are clipped to the same area. This ensures that between every year of data available, the same point in another year can be compared. The final clipped area is the intersection of all horizontal planes of the different DEMs.

3.3.4 Grid projection

To be able to compare points between years, and the properties that will be assigned to them in feature extraction, all datasets were projected on the same grid. This ensures the same amount of data points in the same locations for all data sets, and providing clear visualization of the computed features. The even spread of cells prevents lack of data points where the original triangulation may have larger triangles.

3.4 Feature computation

Features are added to the cells of the grid per year. These features are all normalized before training the MLA. This section will give an overview off the different types of features extracted from the models and added to the grid cells.

3.4.1 Hydrological features

The first type of features explained are the features extracted from the DEM's that are connected to the hydrological characteristics of the lake.

Runoff score

The runoff score per cell is a hydrological property that is computed using the model, but is regarded as a hydrological feature. The runoff score is an important feature in erosion models, as it shows where water will flow on a terrain. Gravity leads water down the path of the largest slope, and in more complex or smaller slope situations the path of least resistance. To mimic this in the DEM, water drops are simulated at every cell of the grid, which then flow a specified number of times to the cell that is positioned lowest of its neighbours relative to the distance between the two cells, thus following the steepest slope.

This model represents an underwater terrain, thus the runoff score will not have the same effects on erosion on the terrain as on dry land since water will not flow in the same directions. Loose sediment moves up and down the water column by forces of downward gravity and upward forces due to eddies in the water. Assuming the sediment will eventually move down the slopes of the underwater terrain, the runoff model can still provide strong correlation to the morphological processes. The runoff is modelled per pixel, resulting in narrow one pixel flow lines. The runoff score in the terrain is computed with single directional flow using the cells 8 neighbors. The result of this are runoff patterns that are concentrated strongly along certain pixel lines of cells. To compute the runoff feature layer for the grid, a runoff drop is initiated from every cell of the grid to get a complete runoff layer for the terrain. In addition to this, runoff drops are initiated at the locations of the different river mouths, the amount of drops dependent on the discharge that flows from this river.

Average runoff score

The runoff scores at every cell is averaged amongst the neighbouring cells in order to compensate for the single cell lines that builds up the original runoff model. The sum of the runoff scores of neighbouring cells within a distance r is divided by the amount of neighbouring cells. The Average runoff score within the distance r is then added as a feature to the cells of the grid for various values of r . In Figure 3.5 below, the averaged runoff scores are demonstrated on different scales of averaging.

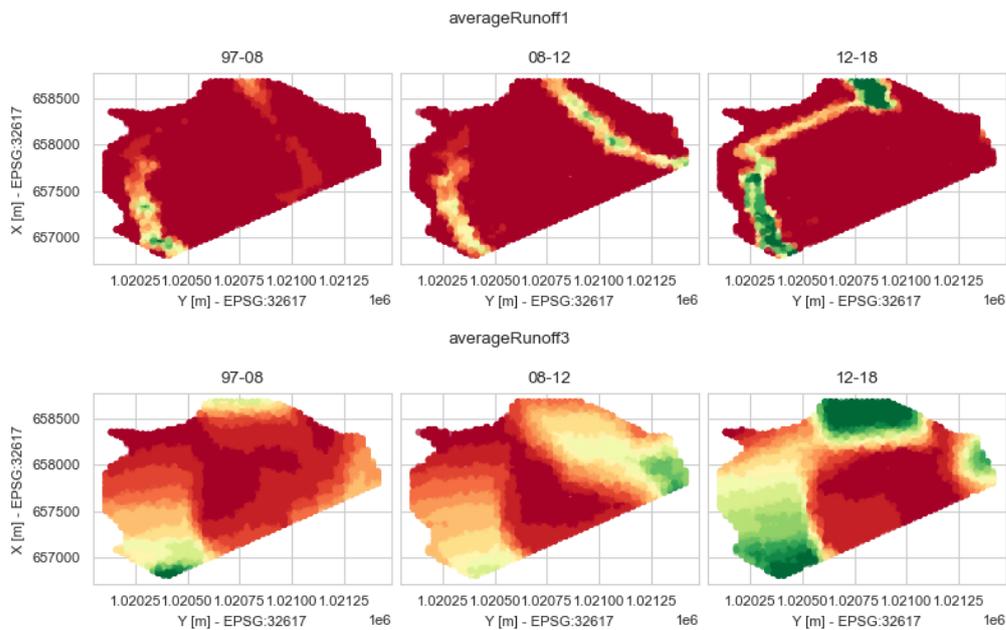


Figure 3.5: Averaged runoff scores

Flow path related features

Using the same principle as is done with the runoff score, the flow path of the rivers is simulated as if it was above water surface of the reservoir. This provides us with a likely flow direction under water, since water flow will encounter the least resistance in the deepest part of the river bed where the runoff score will be highest. Using this river flow channel, several features are computed including the the distance from a cell to the river channel, the river length at that point on the river channel, a feature that combines the river length and distance to the channel which is the weighted Manhattan distance to the river mouth, and the aspect of the slope of that cell relative to the direction of the river channel.

- Channel length - When the channel is constructed following the runoff pattern, the distance between every newly added cell and the previous cell is added to the total length of the channel. this provides every vertex of the channel the cumulative length it has up until that point. For all cells in the grid, the channel vertices are iterated to compute which vertex is closest to that cell, and then the channel length at this vertex is added to this grid cell as a feature. The visual representation of the feature values for the Río Chagres site is shown is Figure 3.6.

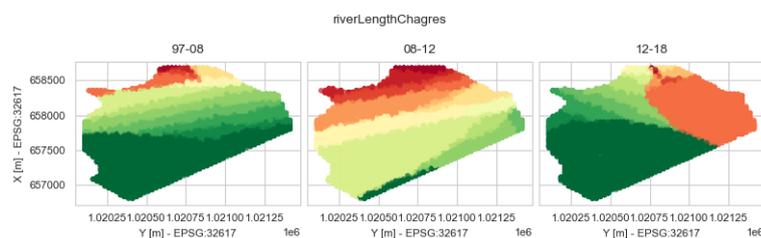


Figure 3.6: Length of flow path for dataset years

- Distance to channel - Similarly, for every grid cell the channel vertices are iterated for the closest vertex to that cell, and the distance to the channel (the closest vertex) is added as

a feature for that cell. The result is shown in Figure 3.7.

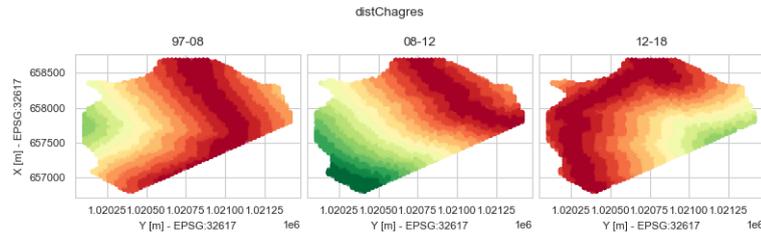


Figure 3.7: Distance to flow path for dataset years

- Weighted Manhattan distance to river mouth - Combining these two features, the length of the river is added to distance to the river channel to the power of 1.5. This power is added to magnify the effect of the cells being further away from the flow channel, therefore amplifying the importance of the cells being closer to the channel. The values of the Weighted Manhattan distance to the river mouth for the Río Chagres are visualized in the image 3.8.

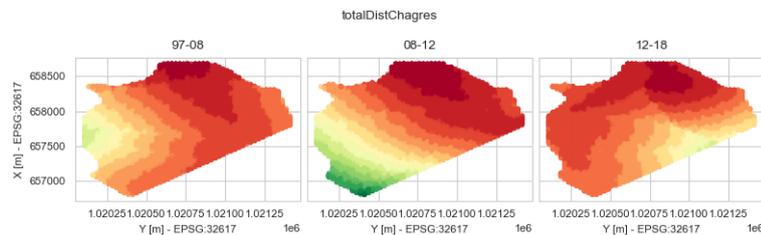


Figure 3.8: Weighted Manhattan distance to river mouth for dataset years

- Aspect relative to channel - The aspect of slopes in the terrain is relevant to the direction in which water flows past it. This is why a feature added to the cells is the cells own aspect minus the aspect of the closest segment of the channel, which is the segment connected to the vertex of the channel closest to the current cell. The feature values for the relative aspect are visualized in Figure 3.9.

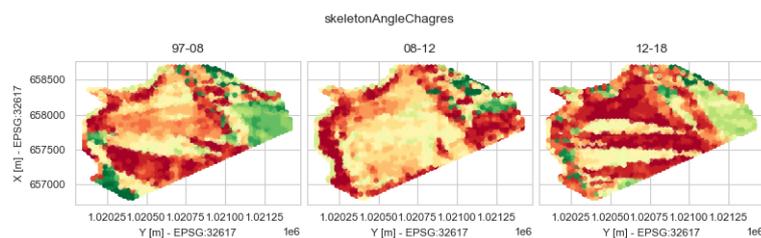


Figure 3.9: Relative aspect angle to flow path for dataset years

Depth

The local depth is an important feature, with a strong effect on the hydrological and morphological processes occurring in the river mouth. The flow velocity is directly correlated to the sedimentation process, as flow determines the amount of sediment that the water can transport. No exact water depth is used here, since this fluctuates strongly during the year depending on the amount of rainfall, as well as the operations of the canal and the amount

of water passed through the dam. The rule of thumb maintained is that all height values in the grid should fall below the vertical depth value, in order to prevent positive and negative depth values.

3.4.2 Geometric features

Geometric features are extracted from the models, being more standard properties which can influence morphological processes.

Slope and aspect

The slope and aspect of the cells in the terrain mesh are extracted from the mesh class in Unity. These are then extracted a second time from the same terrain after smoothing the terrain on several degrees. This allows small and large features to be recognized in the terrain. Additionally, the relative aspect is computed, being the local aspect value relative to that of the surrounding cells. when in a stream, there is an area with an entirely different aspect of slope compared to the rest of the area, morphological changes are likely to occur.

Curvature

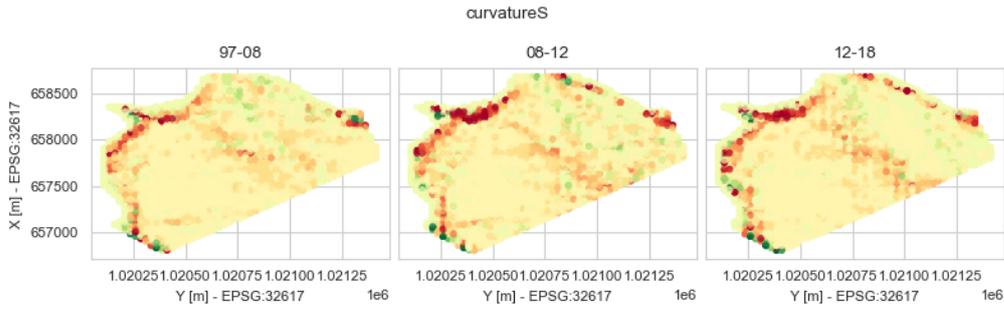
To compute the curvature of the terrain, the second derivative value of the surface using the 8 neighbouring cells. The formula shown in 3.3 is used, being the same formula used for the curvature tool in ArcGIS (Esri).

$$\alpha = [(A2.z + C2.z)/2 - I.z]/L^2 \quad (3.1)$$

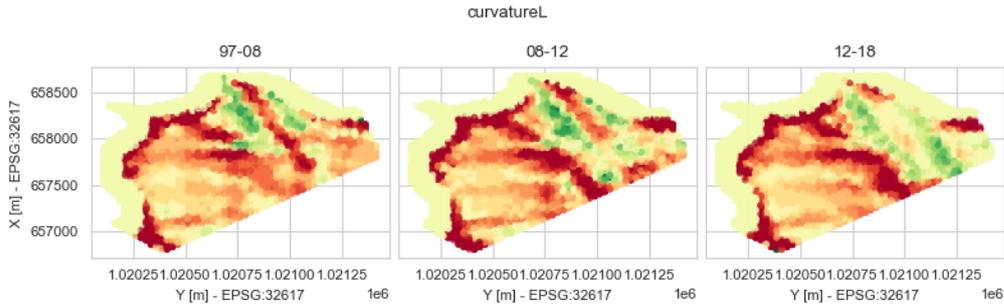
$$\beta = [(B1.z + B3.z)/2 - I.z]/L^2 \quad (3.2)$$

$$Curvature = -2 * (\alpha + \beta) * 100 \quad (3.3)$$

The curvature is then computed on several scales, larger scales meaning the curvature is computed using height values from cells spaced further away from one another. The different scales of the curvature are demonstrated in the Figure 3.10 below. The border of the terrain contains a constant value, due to the curvature taking the values outside of the DEM's borders as zero.



(a) Curvature on small scale



(b) Curvature on large scale

Figure 3.10: Curvature on scales

Relative height

To allow the algorithm to recognize local humps, pits, and other small morphological features in the terrain, the relative height is computed at each cell. This is done by taking the height of the current cell and dividing it by the weighted average height of the 8 neighbouring cells. The weight used to average the neighbouring cell height is the distance to the current cell. This results in the formula 3.4 where $I.z$ is the height of the current cell, $C_n.z_n = 1..8$ are the height of the surrounding cells, and the $dist(A, B)$ function is used to compute the distance between two cells.

$$RelativeHeight = \frac{I.z}{\sum_{n=1}^8 C_n.z_n / dist(C_n, I)} \quad (3.4)$$

3.4.3 Temporal features

When the data available provides coverage over a large variety in a range of years, temporal features can be added to the data. Such features can include rainfall, river discharge, dam discharge, and sediment flux data. In the case of this thesis, and the data available for Lake Alajuela, there is an insufficient variety of temporal data available, and these features were thus not used for training the model.

3.5 Model and feature selection

To select the combination of model and feature selection, several methods were used. In this section these methods will be explained.

3.5.1 Feature selection criteria

Feature importance based on mean decrease in impurity

The random forest regression method can be used to rank the features based on the amount of times these appear in the nodes of the random forest regression trees. This then not only allows for defining the most important features for the RFR, but also provides a benchmark which can be used to set as a minimum level of performance. This minimal level of performance is determined by introducing a random feature, being a feature with a purely random value at every cell. Any feature showing a lower importance than the feature with these randomized values is thus clearly not correlated strongly enough to the morphological changes occurring, and can be disregarded.

With the features that show high importance for the Random Forest Regression model, different tests can then be performed with this baseline of features to use for the different models, to see the effect of the absence and presence of these features. The effect of the features is tested on all models starting with all features, computing the impurity, and removing the least important feature. With this new set of features the predictions is then made once again and impurity computed. This process is iterated until only one feature is left.

Analysing feature selections

To analyse the effect of certain types of features; curvature, river path, or runoff score related features, one or more of these groups of features was removed. The resulting prediction can then be analysed with other feature selections to see the type of inaccuracies are caused by the lack of a specific feature type.

Besides excluding specific clusters of features, individual features are also incrementally excluded, with an analysis of the result at every step. Performing this process several times, allows for a view of what the best performing feature selection may be, for the different models tested. In the figures below, an overview is presented of the SVR, Multi-Layer Perceptron Regression (MLPR), and RFR models performing with different numbers of features, where only the features with the highest importance are kept.

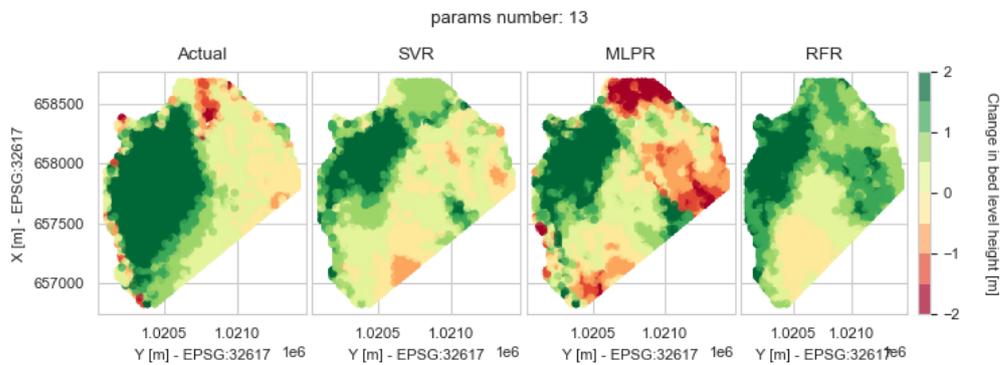
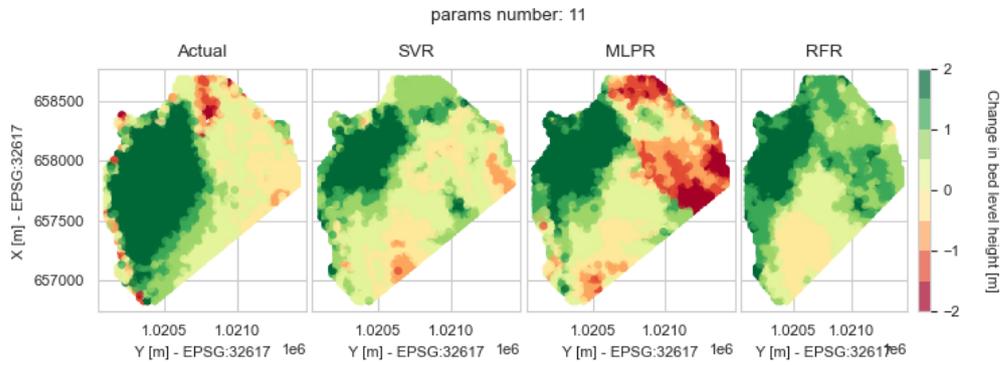
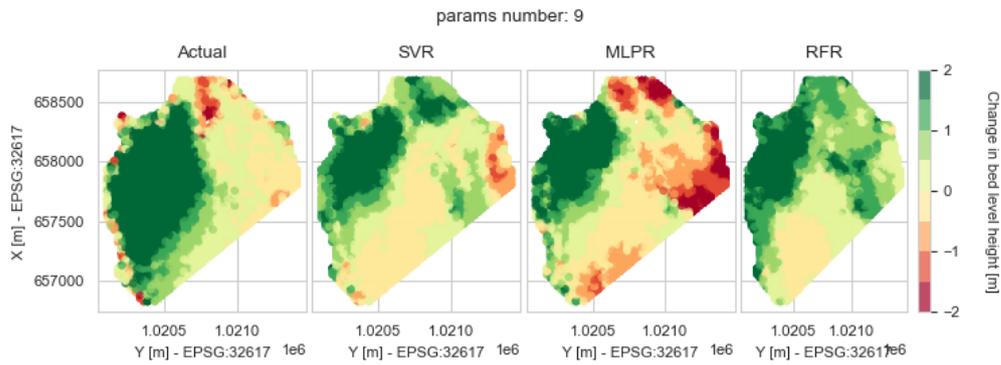
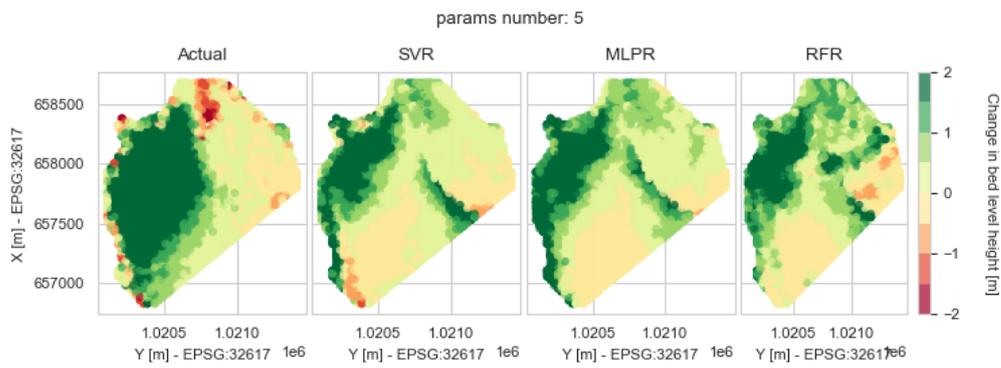
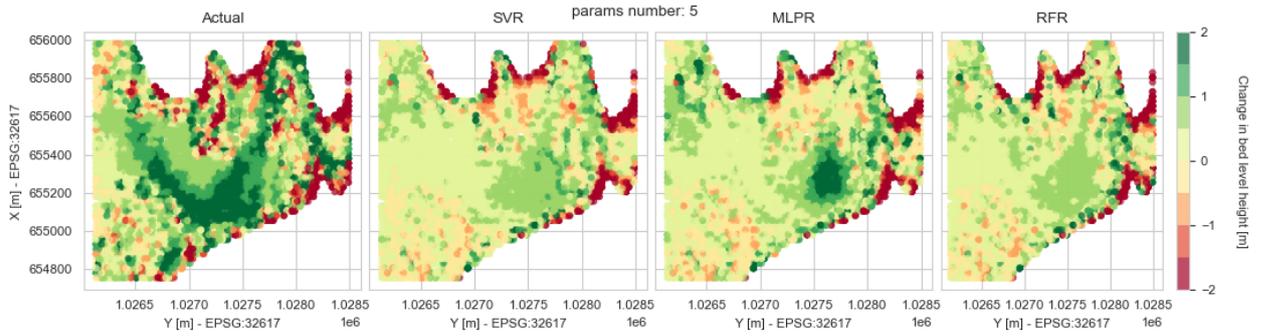
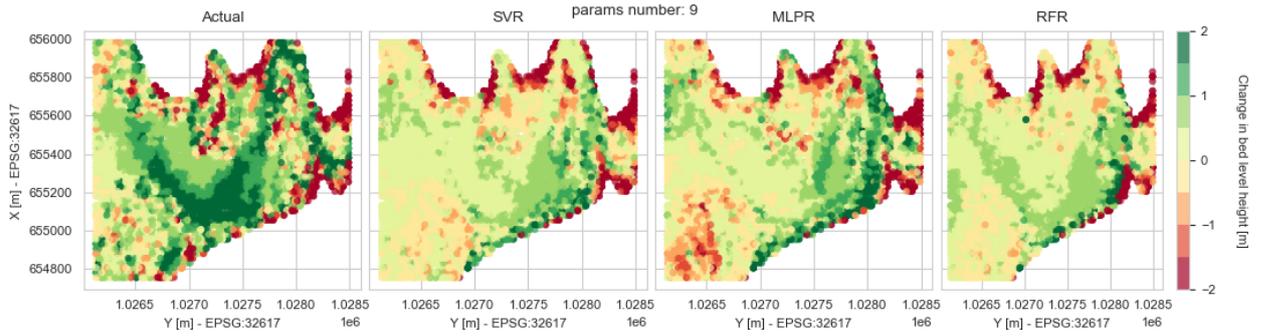


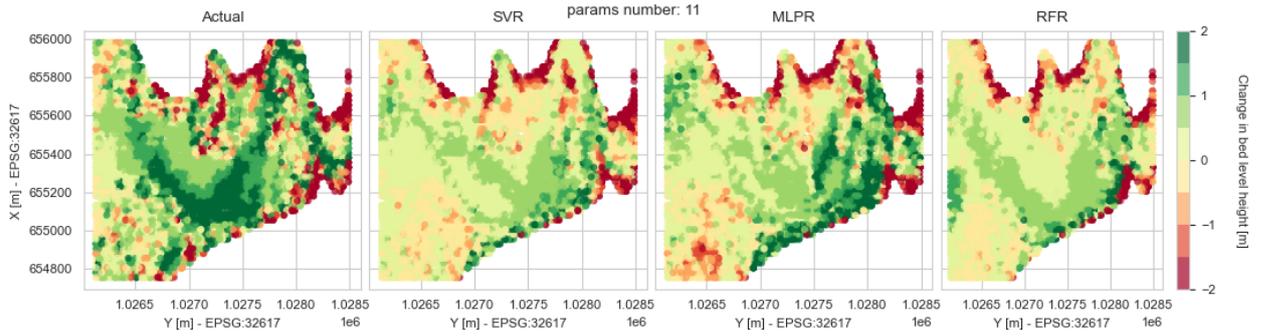
Figure 3.11: Predictions for range of features at Río Chagres



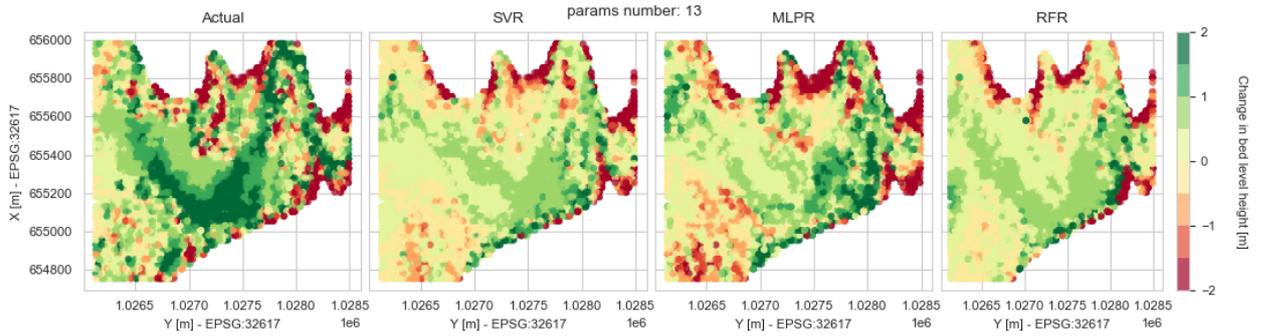
(a) Predictions with 5 features



(b) Predictions with 9 features



(c) Predictions with 11 features



(d) Predictions with 13 features

Figure 3.12: Predictions for range of features at Río Pequení

From the visual comparison over ranges of different sizes of feature collections, and using different ML models, strong and weak points can be found for the different predictions. For a smaller number of features, the predictions remain relatively simple and the effects of the individual features can often be recognized. Using 11 features, a distinct advantage of the

SVR can be recognized in the Río Chagres basin, as the SVR predicts a large sediment hump around the correct area and does not predict any activity in the rest of the basin, where the MLPR predicts erosion and the RFR predicts more sedimentation in other locations.

In the Río Pequení basin tests, all models predict the general area of sedimentation correctly when using more than 7 features, and the mayor difference in performance here is determined by the amount of errors made in the prediction. Adding features can cause the models to predict dynamic morphological behavior incorrectly, and this is what occurs for the MLPR model where the shape of the sediment location is no longer predicted due to too many features affecting the prediction. To add to this visual analysis of the predictions made, the RMSE values of the predictions are shown in the Figure 3.13.

no. of features	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
SVR	1.63	1.77	1.74	1.73	1.52	1.53	1.52	1.52	1.53	1.51	1.51	1.5	1.49	1.51	1.62	1.63	1.63	1.7
MLPR	2.16	1.83	1.81	1.88	1.53	1.54	1.53	1.54	1.51	1.56	1.46	1.42	1.43	1.45	1.56	1.58	1.58	1.7
RFR	1.59	1.64	1.65	1.64	1.53	1.58	1.59	1.59	1.6	1.59	1.59	1.59	1.64	1.59	1.62	1.57	1.68	1.8

(a) RMSE values in meters of the predictions for the Río Chagres Mouth

no. of features	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
SVR	0.877	0.822	0.822	0.822	0.827	0.812	0.800	0.802	0.807	0.806	0.798	0.794	0.798	0.808	0.817	0.835	0.824	0.800
MLPR	0.937	0.867	0.867	0.872	0.846	0.826	0.805	0.816	0.839	0.802	0.781	0.783	0.782	0.786	0.797	0.796	0.786	0.803
RFR	0.819	0.814	0.814	0.811	0.815	0.807	0.810	0.820	0.826	0.833	0.882	0.919	0.937	1.274	1.254	1.074	1.413	1.723

(b) RMSE values in meters of the predictions for the Río Pequení Mouth

Figure 3.13: RMSE values [m]

Although the two areas of study and the morphological processes taking place in these individual locations are very different in nature, the RMSE range of the predictions show similar patterns. In both cases the MLPR with 5 to 8 features seems to be presenting a relatively accurate prediction. The SVR gets the relatively low RMSE for a wider range of the amount of features used, while the RFR performs best with a much larger amount of features. Depending on the model chosen, the number of features used in the final prediction varies. The flow parameters are the most insignificant features in both cases, and thus these parameters are disregarded. To take the river and dam discharge into consideration, these parameters are added to the prediction analysis in a later stage. The remaining features are not as equally ranked in the two cases.

Some of the features are however very similar such as the curvature and the average runoff, and these features are present on various scales. To reduce the total number of features, only the most important scales of these features are kept in the prediction. All features that are associated with the river path are deemed as crucial to the mobility and adaptability of the prediction, and are thus kept in both predictions.

3.5.2 Model selection

Three different machine learning algorithms were tested in the course of this research. Before testing the models however, the type of problem had to be defined as either a classification or a regression problem. Due to the numeric nature of predictor and prediction values, the morphological prediction is a regression problem, and thus regression algorithms were tested. The methods tested were the Support Vector Regression (SVR), Multi-layer Perceptron Regressor (MLPR), and the Random Forrest Regression (RFR). The Gaussian Process Regression (GPR) was also initially trialed, however this model performed far worse than the other models and was excluded from the further research.

Criteria

Comparing the different regression algorithms was done using various criteria.

- *Numeric accuracy*

To compare the accuracy between predictions, the RMSE computed between the predicted height difference and the actual height difference of the 2012 and 2018 DEM's, while training the models with the data between 1997, 2008, and 2012. This way the testing of the models is essentially mimicking the future sedimentation predictions the pipeline is designed to do. This is done for the two river mouths of the reservoir, which have the most significant morphological changes. The RSME for the river mouths is computed using all grid cells allocated within a specified area that is considered part of the river mouth.

- *Probability accuracy*

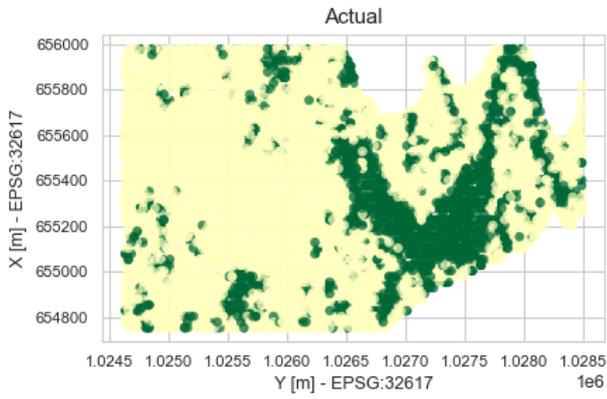
To test the algorithms capability of predicting the location of high sedimentation levels, a probabilistic approach was taken. The model is trained several times over, each time predicting the sedimentation for the same region. The areas of highest probability can then be compared to the actual sedimentation that occurred, and whether this coincides with the real locations at which the sedimentation of more than a specified amount occurred. The amount of cells for which the algorithm either correctly predicts a high or low sedimentation value, divided by the total number of cells is the accuracy. The formula for the accuracy in this analysis is given in the equation 3.5 below.

$$Accuracy = \frac{\sum CorrectHigh + \sum CorrectLow}{\sum AllCells} * 100 \quad (3.5)$$

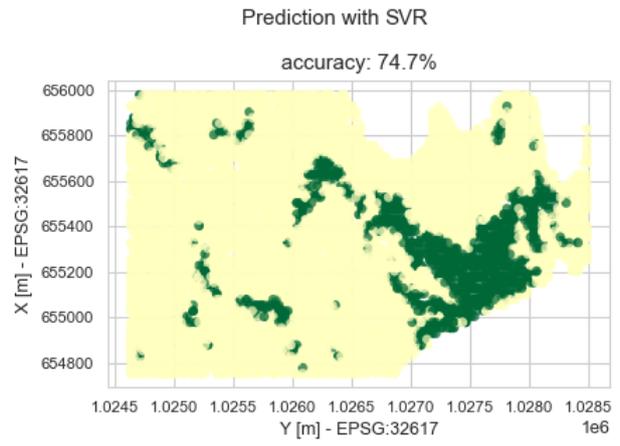
*CorrectHigh = Predicted cell value AND actual cell value are above threshold
for at least an x amount out of all iterations*

*CorrectLow = Predicted cell value AND actual cell value are below threshold
for at least an x amount out of all iterations*

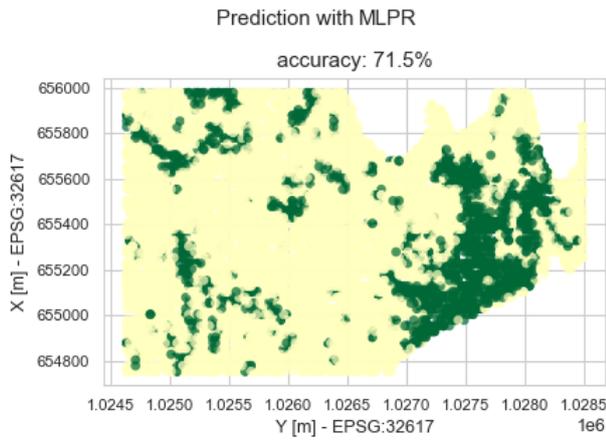
The likeliness of the model to predict relatively high levels of sedimentation or erosion in a specific region regardless of the magnitude thereof is an valuable aspect of a prediction as this highlights the areas of interest. Natural events have a strong influence on the magnitude of sedimentation levels, but the location and the total amount of sediment likely to end up in a location can be predicted and shown in the figures 3.15 and 3.14 below.



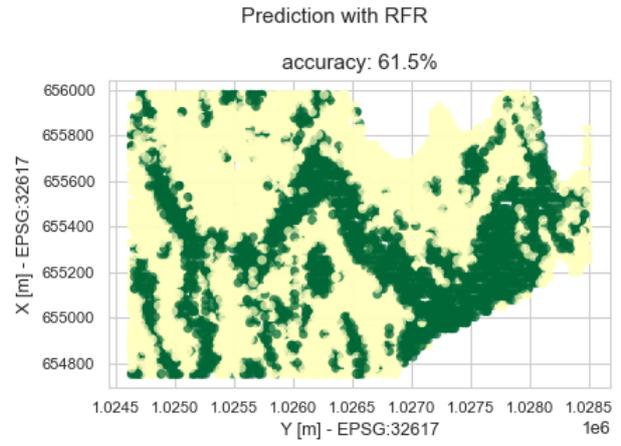
(a) Actual changes on 2 meter threshold



(b) Probability with SVR



(c) Probability with MLPR



(d) Probability with RFR

Figure 3.14: Prediction probabilities for Río Pequeñi

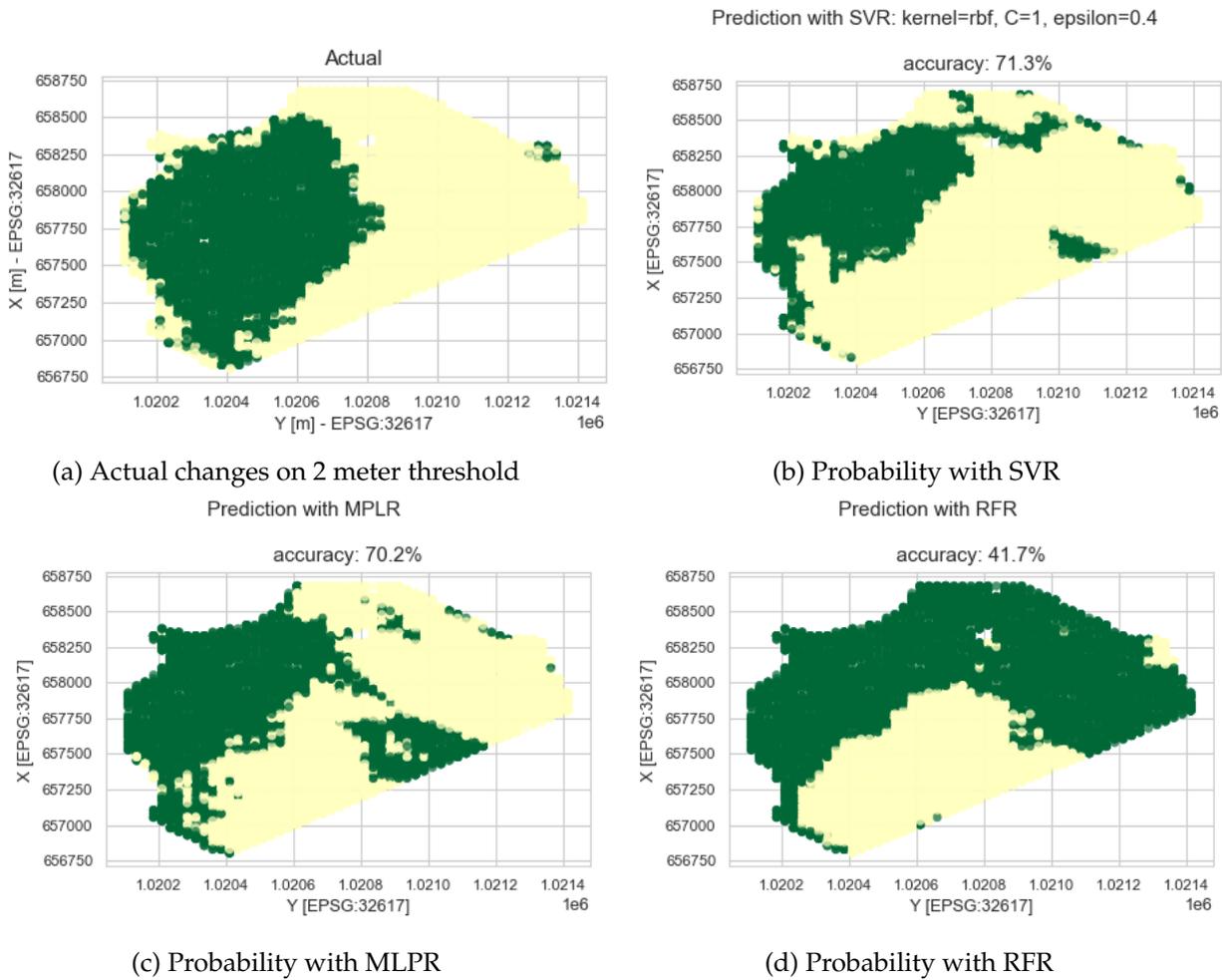


Figure 3.15: Prediction probabilities for Río Chagres

As seen in the figures, the SVR obtains the highest accuracy for both study areas, and tends to refrain from under or overpredicting as done by the RFR and MLPR. The actual percentage obtained is influenced by the benchmark set for the high sedimentation classification as well as the bounds of the area of study. As can be seen in the actual representation of the sedimentation in both figures, the benchmark chosen contains a clear area of interest, with the benchmark sufficiently high to focus the performance trial on the area where the new sediment enters the reservoir from the inflowing river.

- *Sensitivity/robustness*

To test the sensitivity and robustness of the different models, 2 tests were performed. First off, the predictions were made using the data to which an increasing amount of noise was added. Secondly, besides studying the Río Chagres, the models were tested to perform in mouth of the Río Pequení. This river is much smaller than the Río Chagres, carries less sediment, and the morphological characteristics of the lake section the river flows into is very different to that of the Río Chagres.

Selected model and features

Analysing both the local and the global numeric predictions, the model is first chosen. Even though the lowest RMSE value is obtained by the MLPR as seen in Figure 3.13, a visual analysis of the local predictions in figures 3.11 and 3.12 shows that the SVR provides a

more stable prediction. The SVR gets a low RMSE for a wide range of amount of features, and the local predictions show the extreme sedimentation values in the correct locations while the MLPR brings high levels of erosion and sedimentation where there should be no morphological changes. The RFR does not obtain a sufficiently low RMSE, and the local predictions along with the probabilities of the predictions are not as accurate as the SVR or the MLPR. The probabilistic accuracy of the predictions as shown in the figures 3.15 and 3.14 all show the SVR to provide the highest percentage of accuracy for predicting the high classification of sedimentation under a benchmark of 2 meters.

The SVR model is thus chosen as the model to perform the further predictions with for both the Río Chagres and the Río Pequeñí. For the SVR, a very similar RMSE is obtained with 5 to 14 features, and thus the local and probabilistic accuracy is used to select the amount of features used as well as the final set of features used. This will combine the visual analysis of the tests with a concrete numeric percentile accuracy value

For the Río Pequeñí and the Río Chagres cases, different sets of features are used. This is due to the fact that the magnitude of the these two rivers is different, as well as the geological characteristics. All factors of these two rivers are on different scales, including the water flow, sediment flux, width and depth. These influence the hydro dynamics of the water flowing through the terrain, as well as the morphodynamics changing the terrain. Per river, the optimal set of features is thus defined. The final set of features for the mouth of the Río Chagres is the following list of 14 features:

- depth
- flow path length
- Manhattan distance to river mouth
- distance to flow path
- runoff score - smoothed on 10 cells
- runoff score - smoothed on 5 cells
- aspect
- smoothed slope
- Angle respective to flow path
- runoff score - smoothed on 2 cells
- relative height - radius 4 cells
- relative height - radius 2 cells
- curvature - radius 4 cells
- slope

These features are chosen as they provide the best probabilistic local prediction, as well as a relatively low RMSE.

The final set of features for the mouth of the Río Pequení is the following list of 12 features:

- slope
- Manhattan distance to river mouth
- depth
- flow path length
- runoff score - smoothed on 10 cells
- smoothed slope
- relative height - radius 1 cells
- runoff score - smoothed on 5 cells
- distance to flow path
- curvature - radius 4 cells
- relative height - radius 2 cells
- relative height - radius 4 cells

This set of features is chosen based on the low RMSE value obtained, as well as the probabilistic accuracy of the local prediction. The local predictions vary only slightly for the Río Pequení prediction. The prediction contains the best local probabilistic accuracy for 12 features, as more features in the training of the model provide for a more stable prediction in this case. An overview of the features per study area with the related Gini importance in percentage is shown in the Figure 3.16 below.

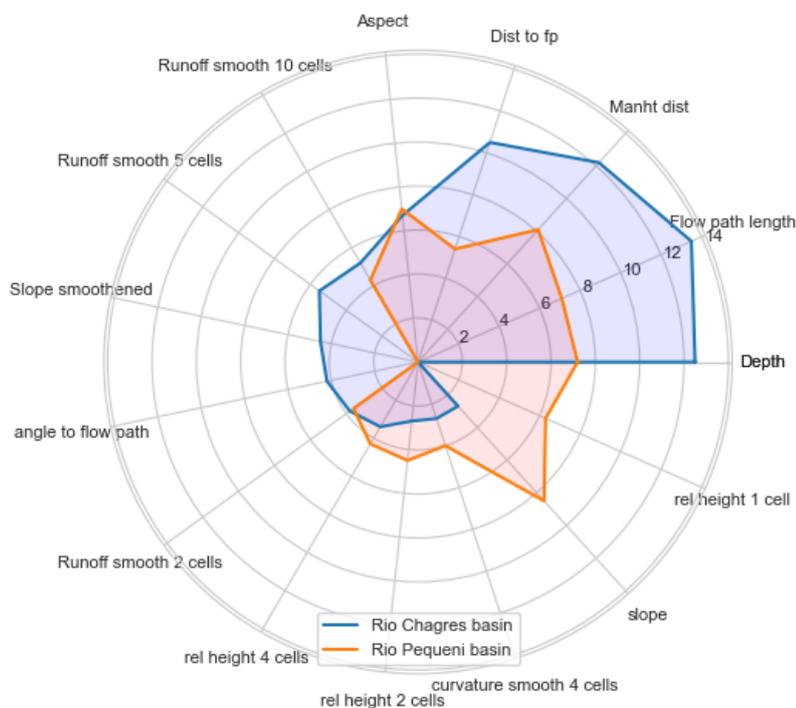


Figure 3.16: Radar graph of feature importances in percentage for both river basins

As seen in the Figure 3.16, the Río Chagres case leans far more on certain features than the Río Pequeí case. This can be due to the fact that the morphological behavior in the Río Chagres basin is more dynamic, and thus the parameters that change with the computed flow path are of much higher importance in this study case. The Río Pequení case has a more spread out use of the parameters, and does not make use of certain parameters that the Río Chagres model does. In both cases most of the features related to the runoff model and flow path seem to be of relatively high importance, magnifying the importance of a correct flow path computation for the iterations of the predictions.

3.6 Hyper-parameter tuning

In order to improve the SVR model performance, the models predictions were compared for ranges of values for these parameters. The parameters tested are:

- *kernel*: The kernel selected is the methodology by which the algorithm fits the data.
- *C*: C is a regularization parameter used to penalize in the fitting process.
- *epsilon*: Epsilon determines the radius of the area in which no penalty is given to wrongly predicted points.
- *degree*: This parameter is only used in combination with the poly kernel. The parameter determines the degree of polynomials used when fitting the data. As mentioned in Section 2.2.2, the degree can result in under or over-fitting when not chosen correctly.

First, a broad orientation was done testing all kernels for a selection of commonly used parameter values for the C and epsilon parameter values. A visual comparison of the predictions made by the different kernels is shown in Figure 3.17.

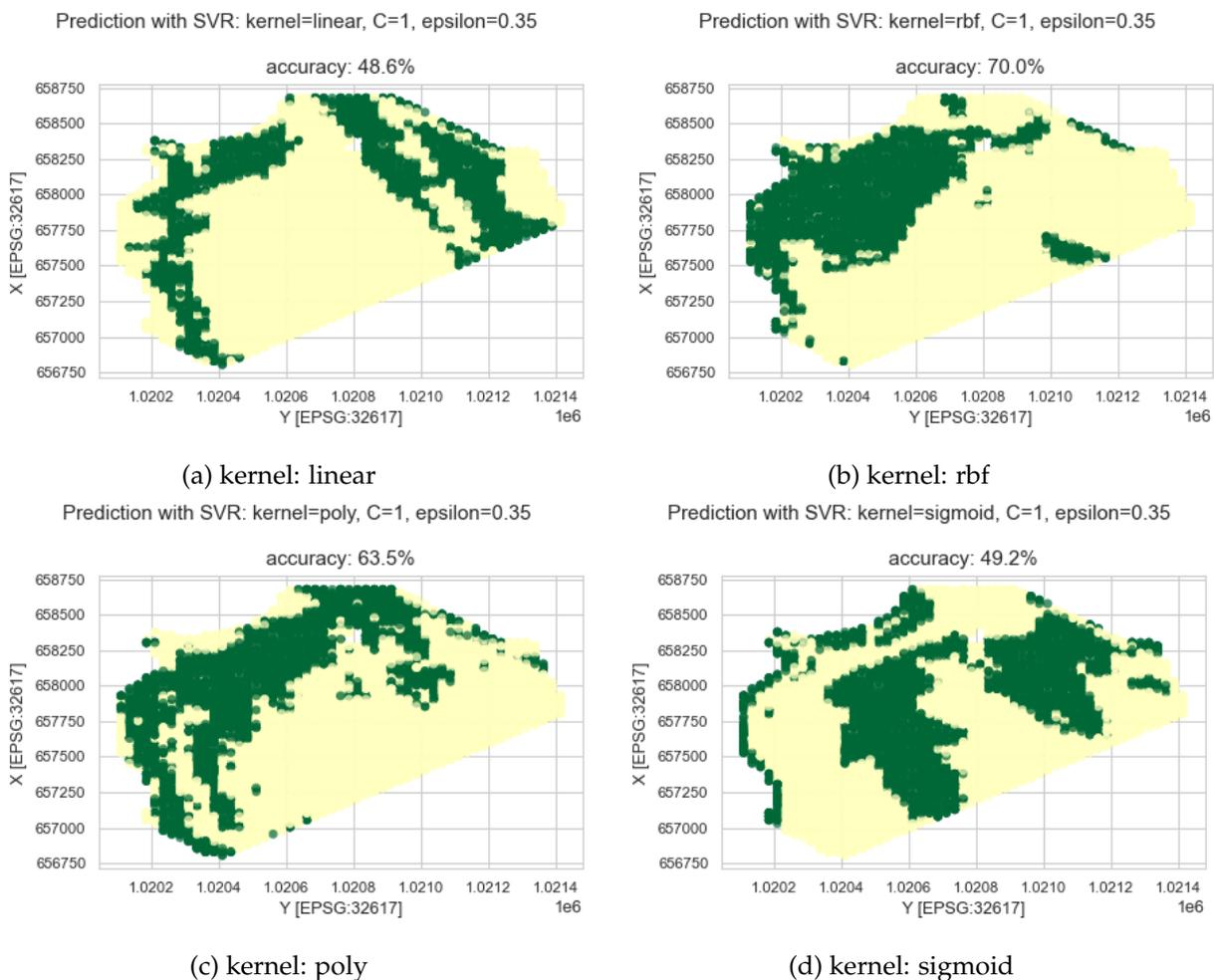


Figure 3.17: Prediction probabilities for Río Chagres using various kernels

The poly and rbf kernels are the only two kernels showing results within an acceptable accuracy, and thus these kernels are then further tested to find the strongest parameter values.

The largest influence on the results using the rbf kernel come from the C parameter. The range of epsilon values show very little variety in the predictions, apart from extreme values resulting in lower accuracy but the difference is practically negligible. The highest average accuracy of the predictions made with the rbf kernel is obtained with parameters: C = 1 and epsilon = 0.3.

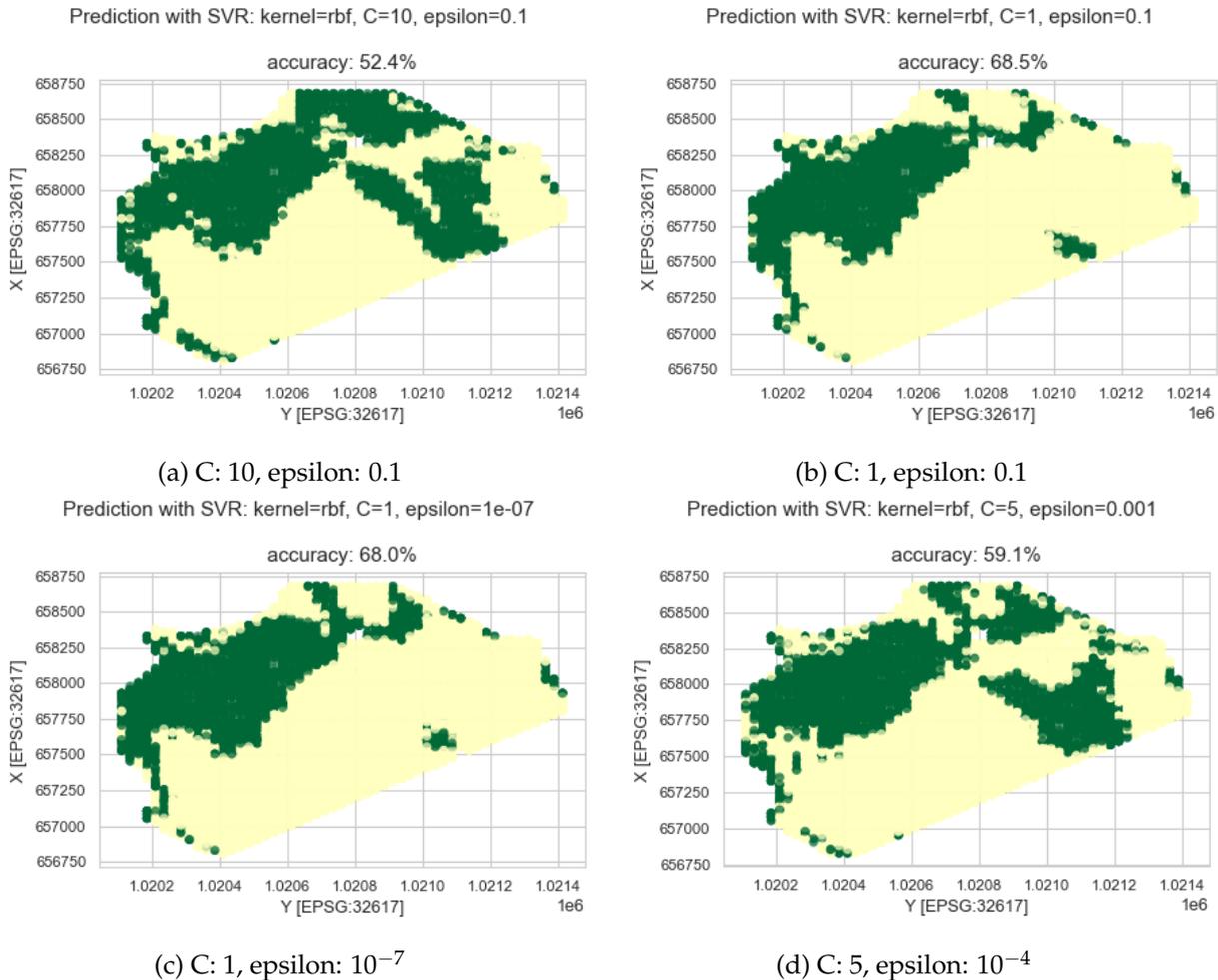


Figure 3.18: Prediction probabilities for Río Chagres with rbf kernel and range of different parameters

Using the poly kernel, the best accuracy is obtained with a degree of 3. This is the default value used, and in this case also the best fitting type of polynomials for the algorithm. The highest accuracy obtained with the poly kernel is around 65%, lower than the accuracy obtained with the rbf kernel. The results of the tests done with the poly kernel are shown in the Figure 3.19 below.

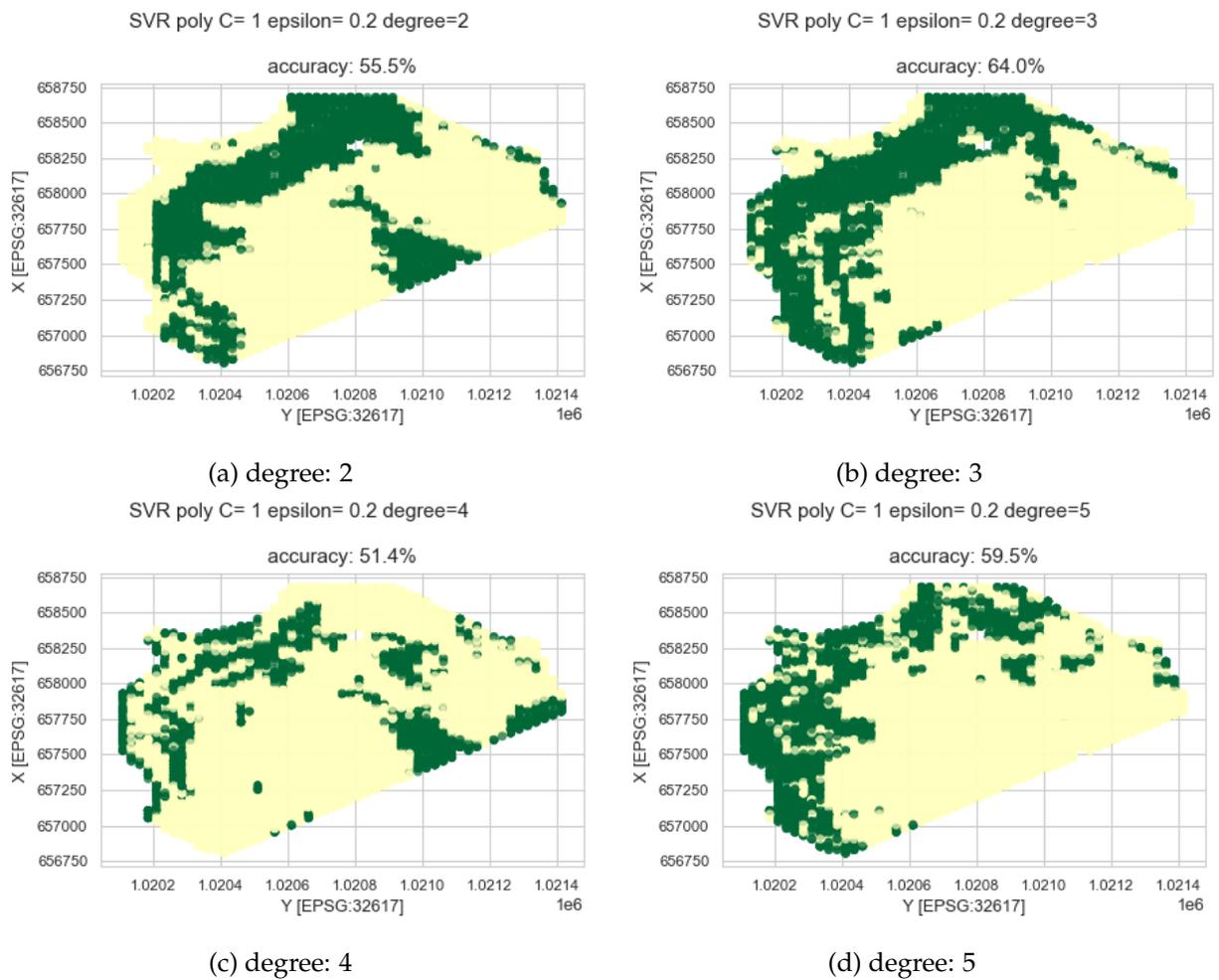


Figure 3.19: Prediction probabilities for Río Chagres with poly kernel and range of different degrees

The highest average accuracy for the Río Chagres predictions is thus obtained with the rbf kernel, a C of 1 and the epsilon value of 0.3. To validate the hyperparameters that were chosen as the best set for the Río Chagres study area, a hyperparameter grid search is also done for the Río Pequeñí study area. The results for the different kernels are shown in the Figure 3.20 below.

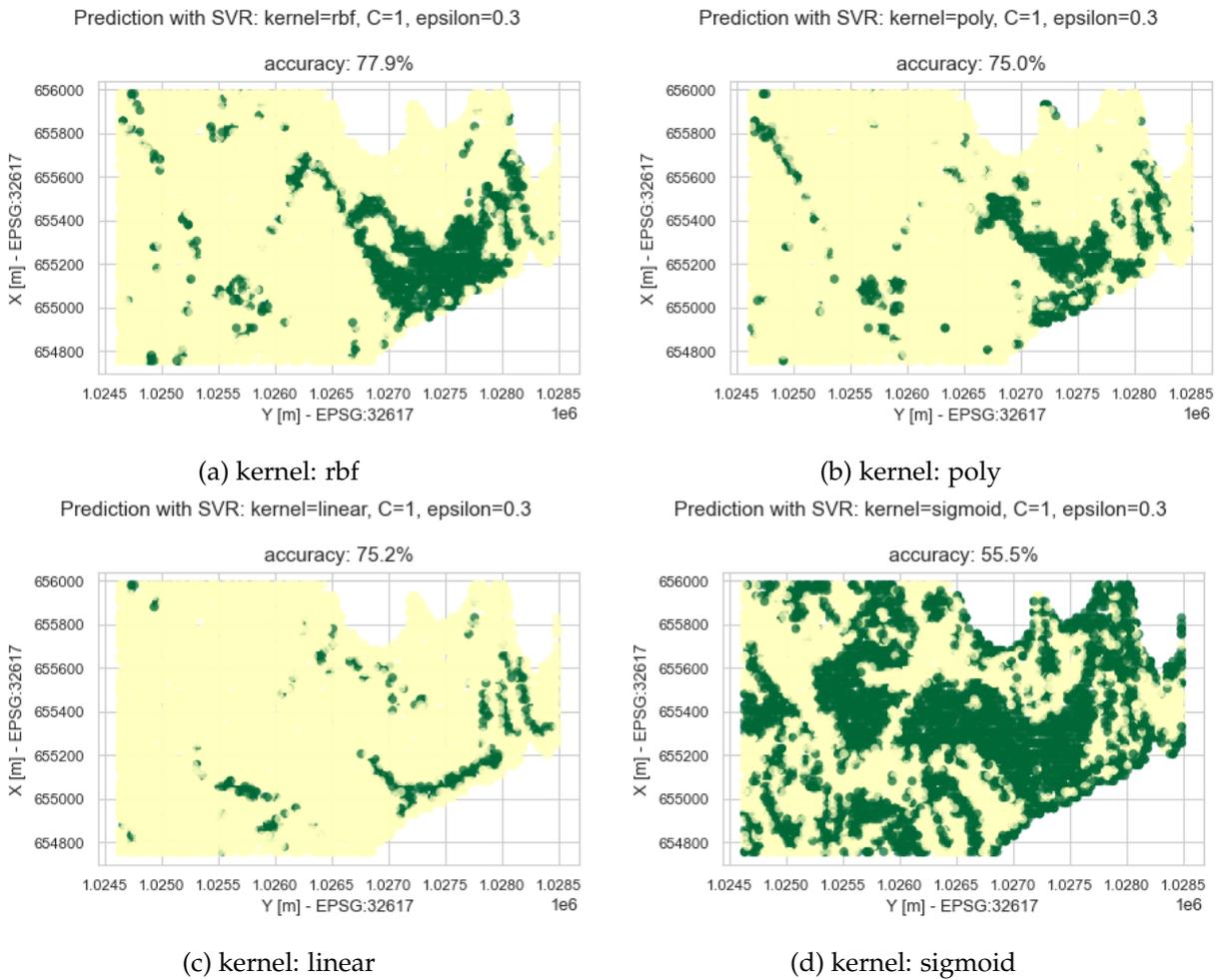
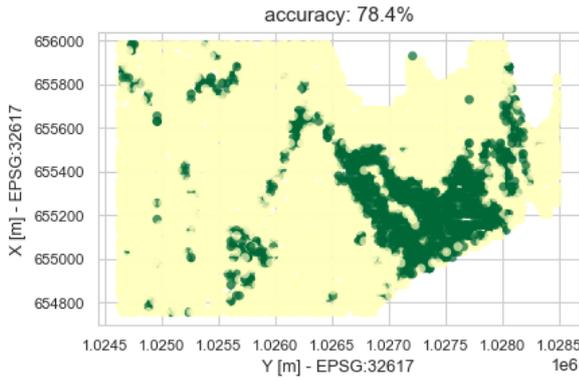


Figure 3.20: Prediction probabilities for Río Pequení with different kernels

Due to the more stable nature of the morphology in the mouth of the Río Pequení, the kernels in turn behave differently to the data. The linear and sigmoid kernels predict poorly once again, although the linear kernel appears to obtain a high accuracy in the prediction. This is not the case, as the large low value sedimentation area in this study area rewards the low sedimentation prediction biased. The rbf kernel predicts significantly better than the poly kernel in this case, whereas in the Río Chagres prediction the two kernels initially obtained more equal results. The sigmoid kernel does not achieve a high accuracy, but seems to predict following a very different end strong pattern compared to the other kernels.

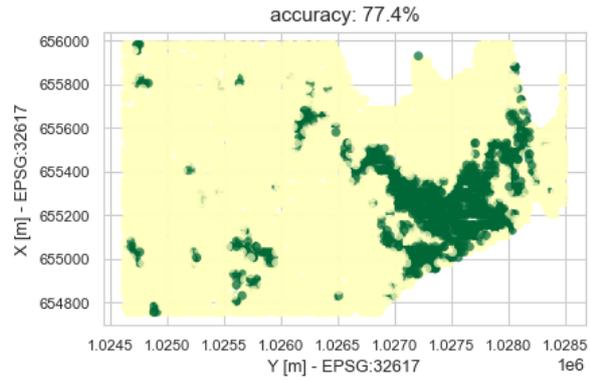
Although the prediction is too high, this results shows the sigmoid kernel has the potential to be the best candidate in a slightly different type of prediction. Due to the critically bad prediction made with the sigmoid kernel for the Río Chagres however, the kernel is not a candidate for this research. All kernels follow the general flow of the river in the Río Pequení basin, however the rbf kernel gets the highest accuracy. For the rbf and poly kernels a grid of hyperparameters was tested to determine the best hyperparameter value set. A selection of the results from these tests are shown in figures 3.21 and 3.22.

Prediction with SVR: kernel=rbf, C=1, epsilon=0.5



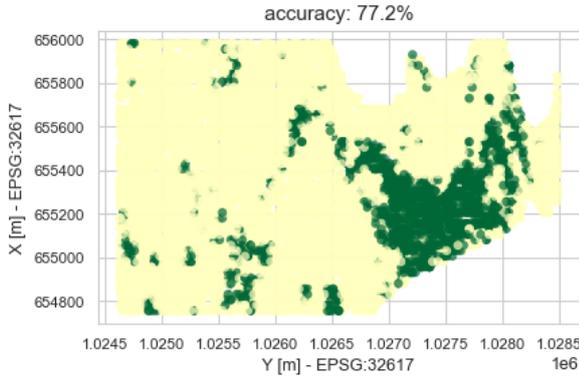
(a) C: 1, epsilon: 0.5

Prediction with SVR: kernel=rbf, C=1, epsilon=0.1



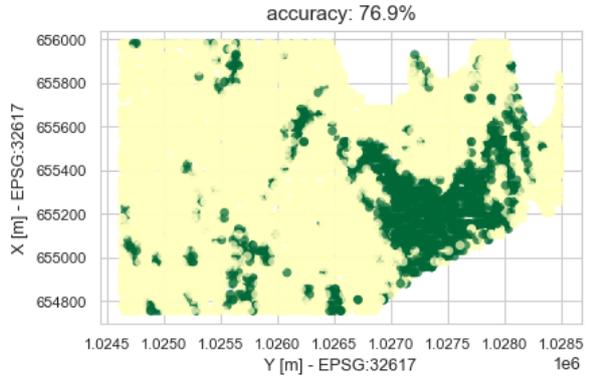
(b) C: 1, epsilon: 0.1

Prediction with SVR: kernel=rbf, C=5, epsilon=0.1



(c) C: 5, epsilon: 0.1

Prediction with SVR: kernel=rbf, C=10, epsilon=0.1



(d) C: 10, epsilon: 0.1

Figure 3.21: Prediction probabilities for Ríó Pequení with rbf kernel and range of different parameters

The model with the rbf kernel obtains a relatively high accuracy in all cases. The insensitivity to the changing of the hyperparameters is likely due to the more simplistic approach necessary to predict the sedimentation in the Ríó Pequení river mouth as the morphological processes are less dynamic.

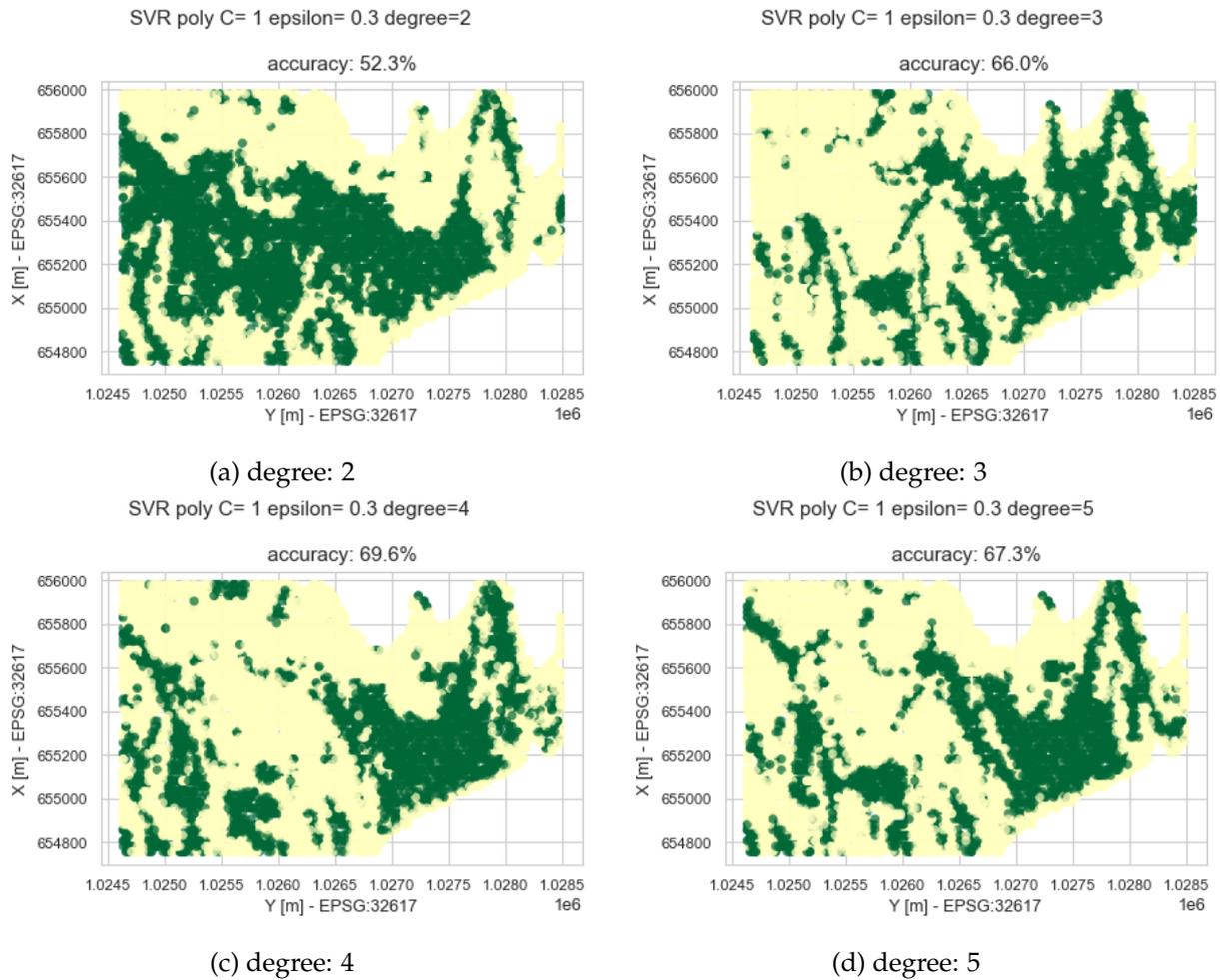


Figure 3.22: Prediction probabilities for Río Pequeñí with poly kernel and range of different degrees

As seen in Figure 3.22, the model using the poly kernel is largely dependent on the degree as also found in the case of the Río Chagres. A 4th degree polynomial obtains the highest prediction accuracy for the Río Pequeñí mouth, however the overall result is lower than that of the predictions made with the rbf. In the case of the Río Pequeñí, the highest accuracy prediction is thus obtained with hyperparameters very close to those used for the Río Chagres, and thus the hyperparameters chosen are:

- kernel: rbf
- C: 1
- epsilon: 0.3

3.7 Future prediction in time steps

Morphological processes do not occur in linear patterns. In rivers, an erosion or sedimentation front will move up or down stream to adapt the slope of the river to the sediment discharge, the sediment carrying capacity of the flow, and other parameters that may have changed, altering this sediment carrying capacity. The location of the highest sediment deposition rate is thus constantly changing, and the location depending on the terrain influencing the flow.

Since the variety in temporal range of the training data is rather small, we resort specifying the water discharge per time step instead of the years. This eventually results in the same format, since a prediction period of 10 years will require a specification for this period being a large or small amount of rainfall. By analysing the training data, the amount of discharge assumed by the model in each time step can be estimated, and thus the amount of steps needed to be taken for an approximation for a period with a certain amount of total discharge can be computed. To predict these time steps using terrain features of the previous years, the data flow was designed as shown in Figure 3.23.

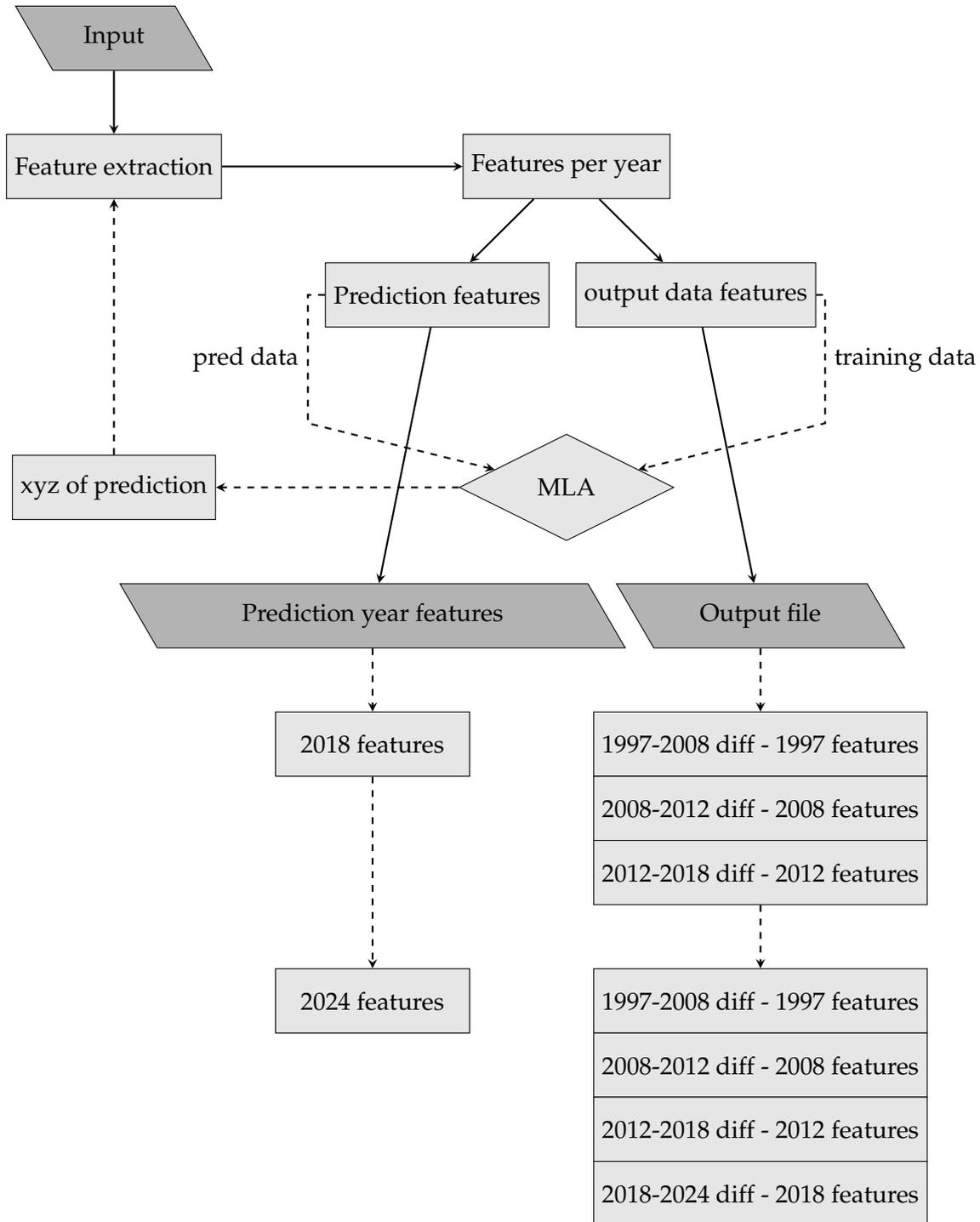


Figure 3.23: Data flow

At the start of the pipeline, the original input datasets go through the feature extraction process. Three data years can be used to train the machine learning algorithm. The features of the year 1997 are used to predict the difference in terrain elevation between 1997 and 2008. The same applies for 2008 and 2012, leaving the 2018 data with solely features, and no value for the difference in elevation. The 2018 features are used for the prediction to be made for the changes in elevation between 2018 and the next predicted year, in this case 2024. The predicted terrain for 2024 is then sent as an .xyz file to the feature extraction, from which the new layer of features can be used to make the next prediction and with that completing the first time step in the prediction.

This process can then be iterated, basing every consecutive prediction on the morphological characteristics of the terrain affected by the previous prediction. This enables the algorithm to predict the moving sedimentation and erosion fronts. In situations with large amounts of moving sediment containing extreme values in the training data, predictions can be done in intermittent steps by scaling the predictions. Scaling predictions and splitting the prediction into smaller steps prevents extreme values from blocking parts of the channel and creating unrealistic humps or artifacts in the terrain. In this matter the quantities of predicted sediment are not only determined by the prediction of the machine learning model in one iteration, but multiple iterations in which the model will recognize the point at which a certain location should not receive any more sedimentation. This methodology is essentially requiring the model to re-evaluate the situation in various time steps, and at every step a decision is made for the amount of sediment that will arrive in that location.

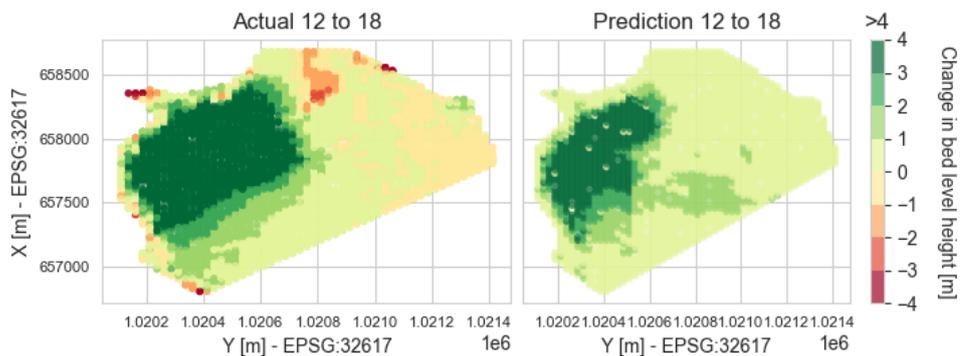


Figure 3.24: Río Chagres prediction for 2018 using time steps

To test this mechanism in a short time frame, the prediction for the Río Chagres basin was predicted over three time steps, all three intermittent predictions scaled and summed together. The result of this prediction is shown in the Figure 3.24, where the accuracy is slightly lower than the single step predictions made for this period as the predicted sediment layer is more spread out over the basin. The magnitude is around the same value as the single step prediction, as the predictions are scaled down and summed to the same factor. Due to the flow path computation getting stuck for this prediction, the amount of steps is limited when starting from the 2012 dataset due to a large sediment hump blocking the flow path right at the river mouth. As the 2018 prediction with multiple steps predicted the large growing sediment hump, this methodology is shown to work once again in diverging conditions modelling morphological movement over time.

Chapter 4

Results and Analysis

4.1 Accuracy analysis

Testing the predictions with the sedimentation values between 2012 and 2018 limits the training of the model since it can only be trained with 3 years of data instead of 4. Nevertheless the model can successfully make predictions, and these predictions are tested and analysed below.

4.1.1 Numeric

Although the predictions made are on a temporal scale, the Mean Absolute Scaled Error (MASE) metric for forecasting models as advised by Hyndman et al. (2006) is not suitable in this scenario due to the short variety in years of data. The Root Mean Squared Error (RMSE) is thus used for accuracy analyses in the testing of the model.

Cumulative error

To analyse the errors made by the models, the errors and absolute errors made in the predictions are first demonstrated in cumulative histograms as seen in Figure 4.1.

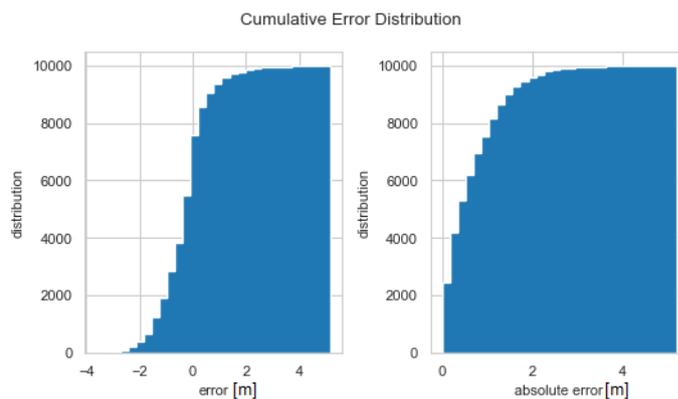


Figure 4.1: Histogram of errors made in prediction for Río Pequeni

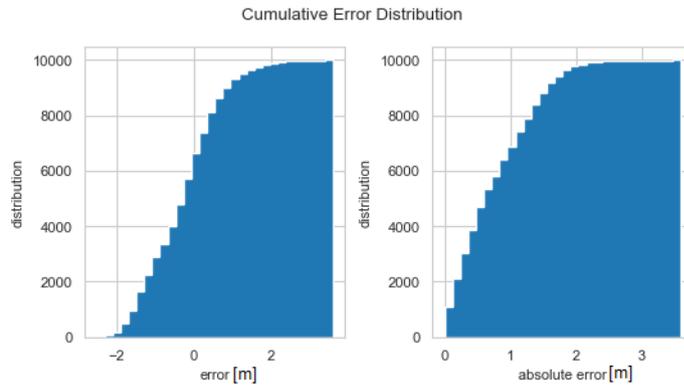


Figure 4.2: Histogram of errors made in prediction for Río Chagres

The majority of errors occurring are between the magnitudes of 0 and 0.5 meters. From the figure we see that for the Río Pequení the amount of over and under-predictions are approximately the same. This does not refer to the amount of sedimentation and erosion falsely predicted, but the locations at which sedimentation was either under- or over-predicted, as there is relatively no erosion occurring.

In the figures 4.3 and 4.4 the distribution of the errors is visualized over the range of values of sedimentation both given by the prediction, and the range of actual values. In an ideal prediction, the values in the plot would form a horizontal line on the zero axis. In the left plot a line can be seen at the bottom the graph, which is due to all values below -2 meters being due to all outliers cut out of the dataset. These outliers are around the borders of the DEM and the upper part of the river and are considered artifacts disturbing the model.

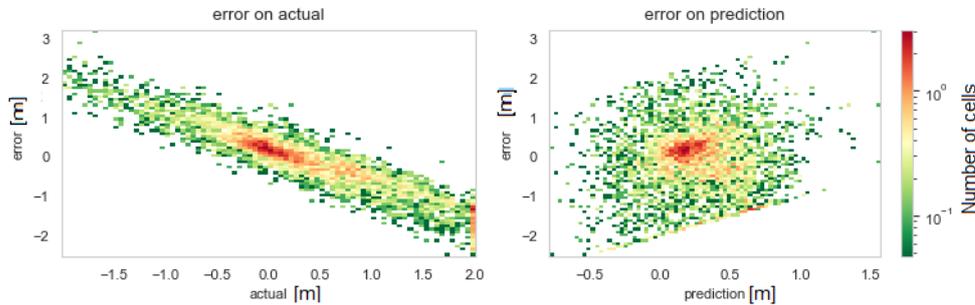


Figure 4.3: 2D Histogram of errors made in prediction for Río Pequení

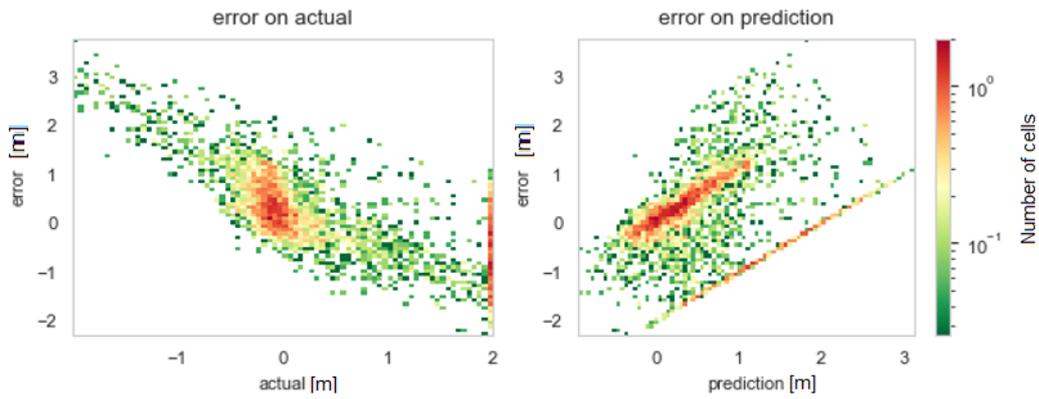


Figure 4.4: 2D Histogram of errors made in prediction for Río Chagres

From the plots above, the difference in spread of values is clear as the Río Chagres basin contains one area of high sediment values, and the rest of the area is relatively stable. The Río Pequení basin's morphological changes are more spread out, as well as the errors made on the range of the prediction. Where in the Río Pequení prediction the most errors are made at much smaller values, the Río Chagres with larger sedimentation values also contains a larger range of higher errors.

Gross error

The gross errors occurring in the prediction will be analysed in this section. Errors that are greater than the maximum predicted value of sedimentation are considered gross errors, the value of which depends on the area of study. For both the Río Pequení and the Río Chagres, any error greater than 1.8 meters is considered a gross error. There are three types of gross errors occurring in the predictions which are highlighted in the Figure 4.5.

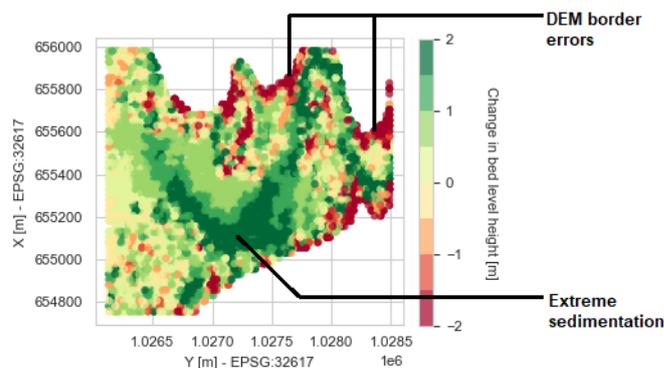


Figure 4.5: Indications of different gross errors

The first error we can identify are errors occurring due misreadings, strong morphological changes on small scale, and artifacts in the DEM around the borders of the lake. These errors are largely filtered out but will still occur and are easily recognizable, however these do not compromise the analysis of the prediction.

The second type of error are underestimations of the morphological processes occurring. This is especially evident in the year used for testing the data, in the period between 2012 and

2018 the amount of sedimentation occurring in the river mouths has been much larger than the years prior.

The final error type is the error that occurs when the location of the sedimentation is wrongly predicted, and the actual sedimentation levels are extremely high as mentioned before. These errors are generally the largest errors, as in the case of the period 2012 to 2018 these errors can be up to 6 meters. For this reason the gross errors can greatly increase the RMSE of a prediction, and the local prediction probabilistic accuracy should always be analyzed. The errors occurring due to the location of sedimentation wrongly predicted are not considered gross errors as these are within the range of the average error values. The predicted sedimentation values do not get as high as the extreme actual sedimentation values, and are thus of a smaller magnitude.

4.1.2 Probabilistic

Besides the numeric accuracy, the accuracy of the probability of local predicted sedimentation is an important factor to be analysed. For each time the model is trained and sedimentation predicted, the amount of runs the sedimentation is predicted to be above a certain threshold in cells of the study area divided by the number of total predicted cells is the accuracy of the model. There is a threshold here for the minimum height to be predicted, and the minimum amount of times a high or low value is predicted out of the runs made. Naturally, testing this with a large amount of total runs will provide the most representative result.

In Figure 4.6 the probabilistic accuracy of the prediction for the Río Pequení mouth is shown. The average probabilistic prediction accuracy is between 70 and 80%, depending on the threshold and the boundaries defining the training and testing area.

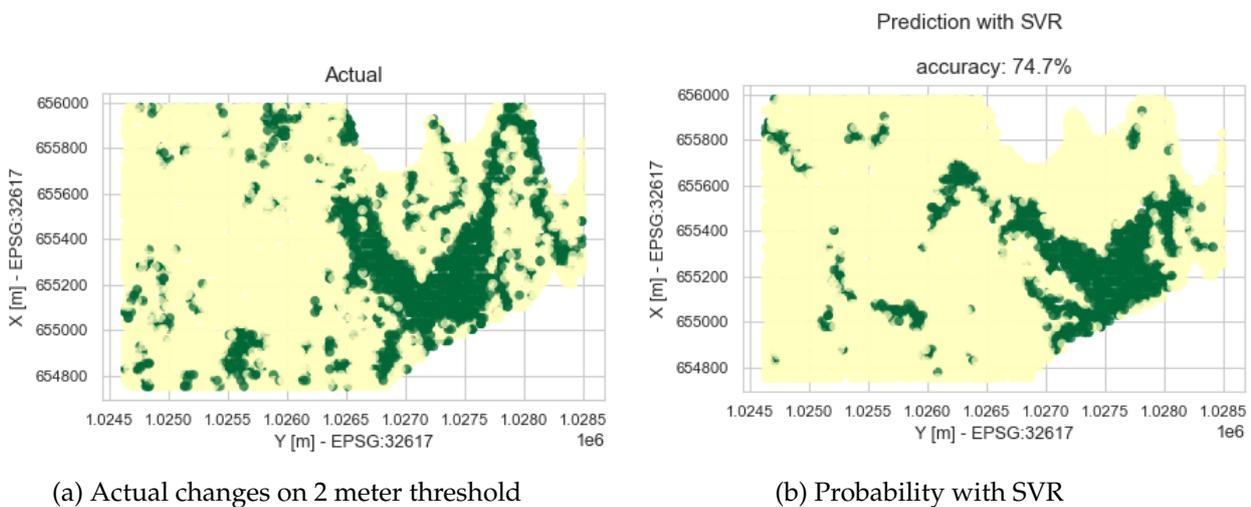


Figure 4.6: Prediction probabilities for Río Pequení

In the Figure 4.7 the probabilistic accuracy of the prediction for the Río Chagres is given. The average accuracy for the Río Chagres is also between the 70 and 80%, which is again dependent on the area of the river mouth selected and the chosen threshold. If a larger area is selected of which less is susceptible to strong morphological changes, the accuracy of the prediction will increase strongly due to the relative amount of morphology needing to be predicted being much smaller. For this analyses, in both the case of the Río Pequení and the Río Chagres the bounding box is chosen to contain as much morphologically dynamic area as pos-

sible.

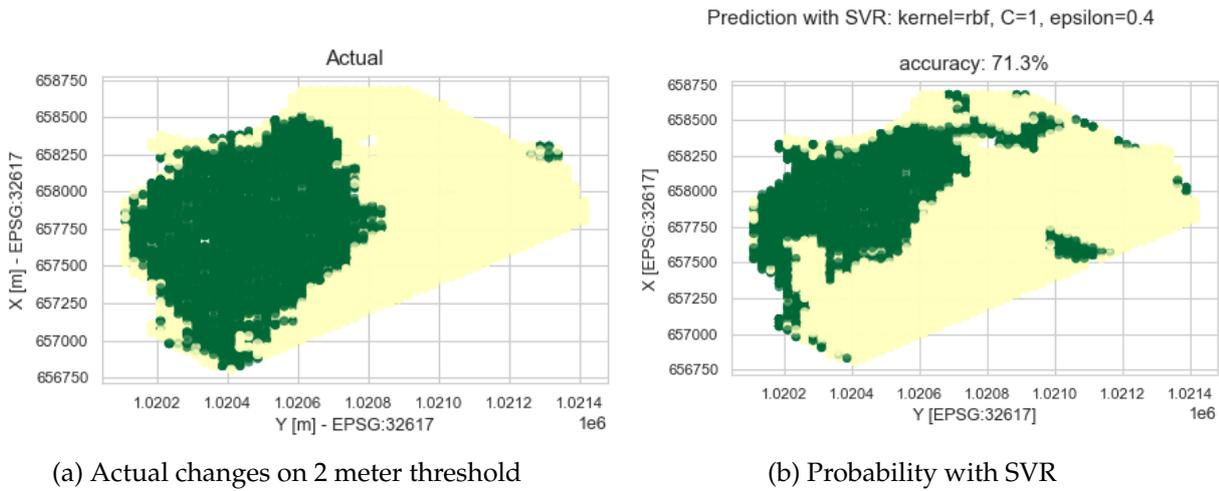


Figure 4.7: Prediction probabilities for Río Chagres

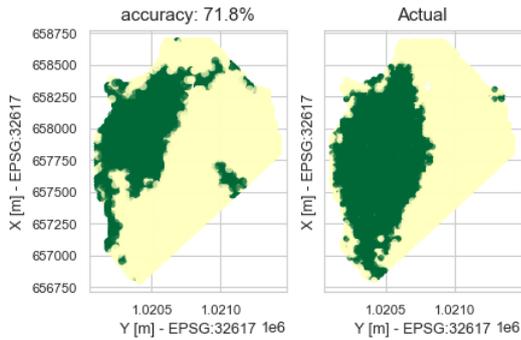
The capability to accurately predict the location of future sedimentation is thus strong, and adaptable to different areas as shown above.

4.1.3 Robustness to noise

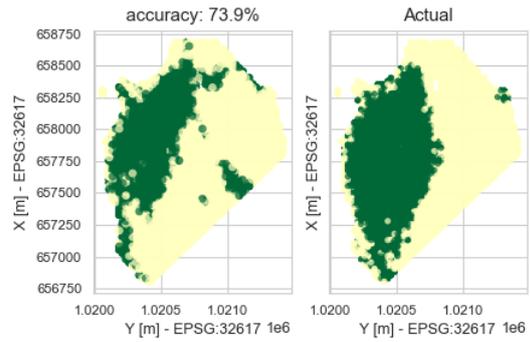
To test the performance of the prediction to noise, several tests were performed. First off, to check the effectiveness of the features used, a feature was included with a random value per cell. Any feature then showing a lower importance than this random feature in the impurity score can be deemed as invalid.

Secondly, noise was added to the starting data, thus allowing the noise to carry through into the features used in the prediction and/or training of the model. The noise is added in the form of random added values in random intervals. This allows for the testing of the effect of noise in the data at various steps in the program.

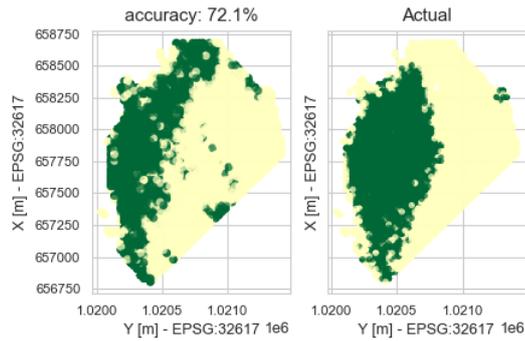
Noise is added at random intervals adding an extra height value in points of the DEM before extracting features. In the Figure 4.8a the control prediction is shown, made with the SVR model with the original data and standard settings. In Figure 4.8b a height of 1.5 meters is added at random intervals of approximately 1 in 50 cells in the training data only to see the effect of this noise in the predictions of the test. In Figure 4.8c a height of 1.5 meters is added at random intervals of approximately 1 in 20 cells in the training data. In Figure 4.8d a height of 1.5 meters is added at random intervals of approximately 1 in 20 cells in the test dataset, showing the effect of the algorithm trained with regular data and the prediction done with a dataset containing noise. The final noise test shown in 4.8e is done with noise in both the test and training datasets.



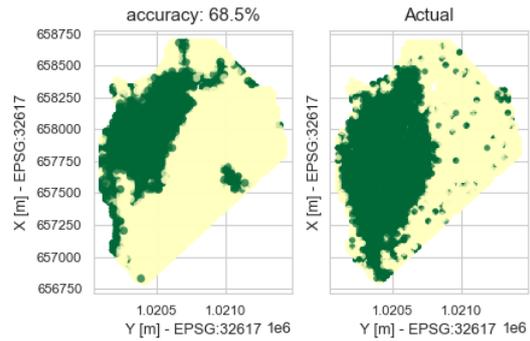
(a) No Added noise



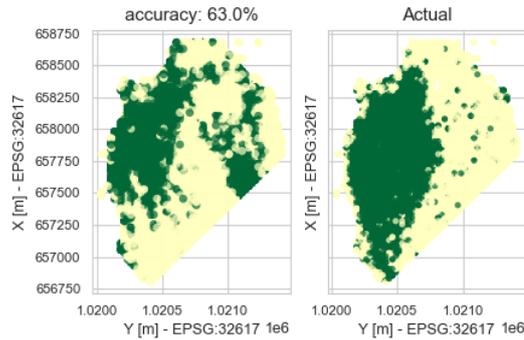
(b) Test 1: 1/50 frequency of noise in training data



(c) Test 2: 1/20 frequency of noise in training data



(d) Test 3: 1/20 frequency of noise in testing data



(e) Test 4: 1/20 frequency of noise in all data

Figure 4.8: Overview of tests done with added noise

In table below, the accuracy scores of the probabilities for the models predictions and the RMSE for the predictions with the different noise tests performed are given.

Prediction	Accuracy	RMSE [m]
Control	72%	1.57
Test 1	74%	1.58
Test 2	72%	1.77
Test 3	69%	1.64
Test 4	63%	1.66

Table 4.1: Noise test results

The small differences in accuracy between tests can be due to the individual test run having a stronger prediction than others. More clear however is that the RMSE of all tests is higher

than the control run, which is as would be expected. The increase in RMSE is however far from substantial considering on average 1 in 20 cells was given a strong vertical shift. The accuracy is most strongly affected in the final two tests, with the accuracy of probabilistic prediction falling below the 70%, and in test 4 even reaching 63%. Taking into consideration the amount of noise added to the datasets for test 4, the model seems to keep the predictions relatively stable and shows strong robustness to the noise that was added.

4.1.4 Adaptability

The actual performance of the model highly depends on the type of morphological changes occurring in that area. Sediment moves all throughout the reservoir, however the most systematic and largest changes occur near the mouth of the two entering rivers, the largest being Río Chagres. To compare the accuracy in a different area of the lake, the mouth of the Río Pequení is analysed. The largest predictive challenge was the Río Chagres 2012 to 2018 sedimentation, as a scenario occurred for which the model was not trained, and still managed to predict the change in the location of sedimentation.

The model is applicable and able to predict morphological changes in the different areas of the lake by going through the pipeline individually for each use case. This is due to the fact that the two river mouths are of different scales, and thus require the features for the model to be of a slightly different scale and have a different distribution in the feature types.

To predict morphological behavior in a different area, the location of the river's entrance into the reservoir needs to be added in order to let the program compute the flow paths per year of data available. The flow path then needs to be checked per year, since any error in this initial flow path computation will cause the training data features to be invalid.

4.2 Limitations of data availability

4.2.1 Years of data

For a MLA to be successfully trained with a feature, there needs to be a correlation between the feature and the result, and a wide range of values for this feature in the data should be available. Features dependent on morphological characteristics of the terrain are naturally provided in a large range. Features connected to specific years however, are solely provided in the amount of years of data provided. For DEM's the number of available years is the limiting factor, and thus features connected to a specific year are less likely to work. In the case of Lake Alajuela datasets, there were 4 different years of datasets. The rainfall and discharge data were available per year, however to train these with the model effectively would require a larger range of different years. To effectively include the rainfall and discharge data in the analysis, another method is required since these cannot be included in the features for the algorithm due to the lack of range.

4.2.2 Accuracy of DEM's

Certain morphological features extracted from the models are more sensitive to details in the terrain than other features. To provide an equivalent and realistic coverage of the features over all years of available data, the level of detail between the years should be as close as possible.

4.3 Comparison to common methods

4.3.1 Numerical models

Modelling river morphodynamics is an extremely complex challenge with too many factors of influence to simulate with 100% accuracy. To get an understanding of the processes and create stronger predictions, simplifications have to be made. At the core of river morphodynamics are the so called 'Open Channel models' in which an open, straight, uniform channel is assumed. By splitting any river into segments of straight channels, analysis can be made and predictions formed on more complex systems.

When bends and bifurcations are introduced, a layer of complexity is added to the system as the behavior of the water in such locations is highly dynamic, complex, and difficult to predict. A number of theories stand however, explaining the transportation of sediment within bends and bifurcations. No matter the complexity of the model, upstream and downstream boundary conditions must be set. These conditions are not influenced by what occurs within the area included in the models prediction, but will influence the outcome.

The model used in this thesis does not regard any of the formulas or principles used in traditional models. Instead of these formulas the MLA computed its own correlations between features and the resulting morphological response. The same boundary conditions are valid as in any model however, and in order to define a prediction the discharge must be predetermined in order to properly analyse the prediction result. In most traditional models however, one of the upstream boundary conditions set is the sediment flux. This is not a boundary condition set in the model of this thesis, as the value for this is not known and the MLA compensated for this by basing the prediction on historical data, assuming a similar sediment flux.

A major difference is the area of study. Traditional models will consider river sections, while in this thesis the focus is laid on the mouths of the rivers flowing into the reservoir. Since the width of the river essentially multiplies by a factor 10 or more as the river mouth opens up, the direction of flow and sediment deposition can no longer be predicted with traditional methods. The ML model uses the terrain as its input, and therefore does not work with a bounding area the way traditional models do.

The more complex computer models available nowadays for predicting morphology are using grids that follow an axis along the center line of a river. Such a grid will not work effectively in a river mouth, as with the extreme widening of the river the grid cells are stretched completely out of proportion.

4.3.2 Morphology ML prediction models

The existing models are often one of the two types; erosion prediction models or spatial prediction models.

Many of the Morphological predictors using MLA's are used to predict erosion on a terrain outside of large bodies of water such as lakes and reservoirs. As in such terrain, water flows with shape the terrain, the water flow patterns and velocities are much easier to predict than flow under the surface of the water. The velocity along with other characteristics of the flow are determining of its erosive capacity, and thus features such as the curvature of the terrain provide strong correlation with the erosion that can occur in an area. This is a main difference with the prediction of morphological changes below the surface, since the underwater flow is far more difficult, if not impossible to predict. Similar features were trialed and used in this thesis, and not all features proven successful for above water level erosion helped the morphological changed below surface.

More extensive research has been done towards spatial predictions of sedimentation using MLA's. Even though these predictions are not on a temporal scale, many of the features

generally used in spatial predictions can be relevant in temporal predictions as well, since the temporal prediction in this thesis includes a spatial prediction. The difference between the spatial prediction done in research by (Mitchell et al., 2021) for example and the research in this thesis, is that (Mitchell et al., 2021) predicts the total level of sediment at the same time-frame as the data measurements were taken, while in this research the prediction is the additional sediment that will accumulate in a location after a specified interval, starting from a specific year. In this thesis, underwater reservoir sedimentation is predicted on a temporal scale, and is thus of a different category than the types of models explained above.

4.4 Model application

The sedimentation in Lake Alajuela is concentrated in two regions where the inflowing rivers enter the reservoir. These are the regions in which the large majority of morphological activity occurs, and thus these areas are of interest for the analysis of sediment entering the reservoir. The upper basin of the lake is disregarded due to the morphological behavior occurring there at low water level, which is not relatable to the behavior in the other sections of study. The two areas of study are distinctively different, containing different behavior in different magnitudes. Nevertheless, the same pipeline is used for both use cases to predict the morphological changes that will occur in these regions. An analysis of the hydrology and morphology is done in Section 4.4.1 in order to validate and argue the predictions produced by the model.

4.4.1 Hydrological analysis

There are three main external factors determining the morphological behavior in any reservoir: the incoming water discharge, outgoing discharge, and the incoming sediment flux. The incoming and outgoing water discharges determine the water level of the reservoir. The sediment flux enters the reservoir, but does not flow out of the reservoir (with the exception of smaller reservoirs containing a flushing mechanism). In Lake Alajuela, water discharge and sediment flux enters the reservoir through the Río Chagres and the Río Pequeñí.

The Río Chagres is the main source of water entering Lake Alajuela, and carries the largest amount of sediment into the reservoir. As this river enters the reservoir, the water depth of the flow channel increases and the channel naturally gets wider. This results in a decrease in flow velocity resulting in sedimentation. As sedimentation is constantly occurring in the river mouth, the flow depth and width of the path is constantly changing. As the incoming discharge of water is determined by upstream boundaries, this does not change by a change in the channel geometry, and thus the flow of water will either increase in depth, or in this case flow through a path of less resistance, being a deeper channel. The flow in the mouth of the Río Chagres is thus constantly adapting and changing, resulting in the location of the sediment deposition changing with it. Figure 4.9 shows the location of sedimentation visibly changing to the other side of the river mouth which is a kilometer away.

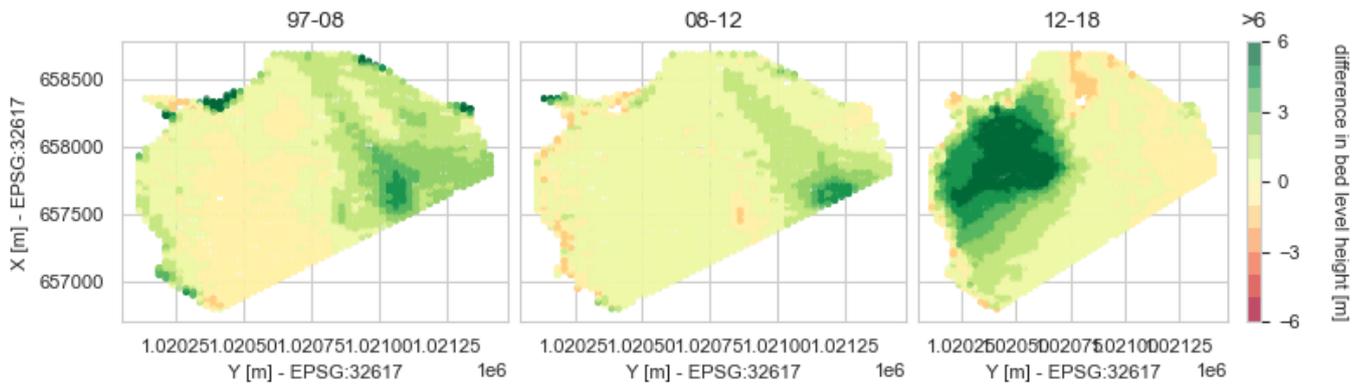


Figure 4.9: Sedimentation in Río Chagres Mouth in periods 1997-2008, 2008-2012, 2012-2018

The Río Pequení is the second river entering Lake Alajuela and is smaller carrying less water and sediment. This river enters a secluded upper basin of the reservoir located on the north side of the lake. This section of the lake shows a strong effect of the different morphological processes occurring in times of high and low water level and rainfall. Sediment is deposited in almost the entire width of the basin when water level is high, raising the bed slightly. Then when the water level is low, the river cuts a path through the bed of the basin, eroding a narrow channel through the lake bed. This is clearly visible in the image 4.10.

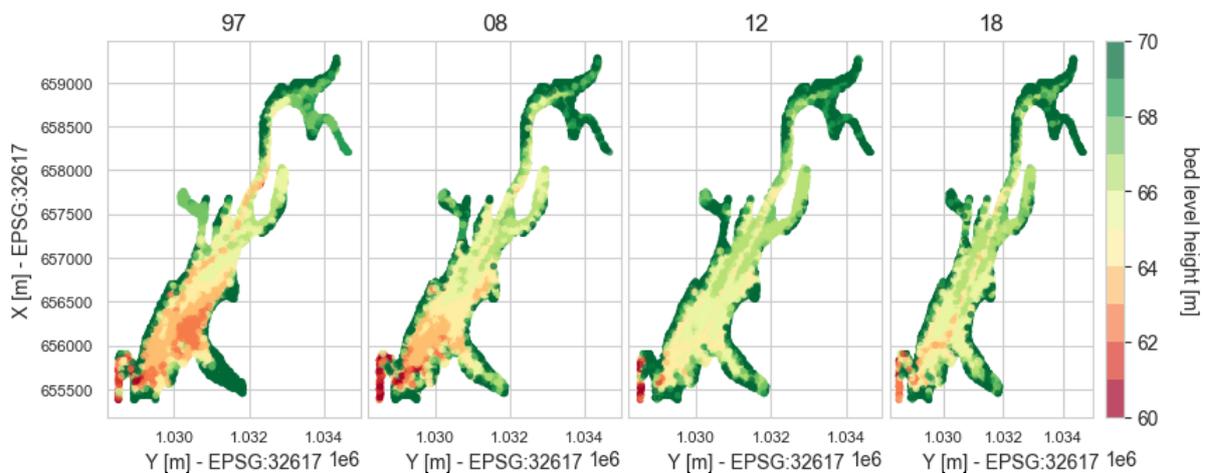


Figure 4.10: Heights of Río Pequení Mouth in different years

From 2012 to 2018, there was much less rainfall than years before, showing very clearly in the morphological changes in upper part basin of the reservoir. The water level during this period was drastically low, resulting in no sedimentation and solely a channel being cut through the lake bed. In the area of the lake at the mouth of the Río Chagres this is less evident, as the water is deep enough not to be affected as much by the lower water level. The differences in height between the years are shown in Figure 4.11 which also show the substantial difference in morphological changes from 2012 to 2018 compared to the rest of the years. The morphological changes occurring in this latest period have little correlation to the changes in the years prior. Lowering of the water level of the main section of the lake result in the upper part of the lake to be partially dry, or close to dry to the point where large parts of the lake do not allow flowing water. In the sections without flowing water, there is no morphological changes, as is seen in the yellow regions in Figure 4.11. The only morphological

changes occurring in these low water level periods will be the erosion of the river carving a route through the land. No sufficient data is available to model this process, as DEMs and water levels of this region would need to be available at monthly intervals since the water level can change at different times during the year.

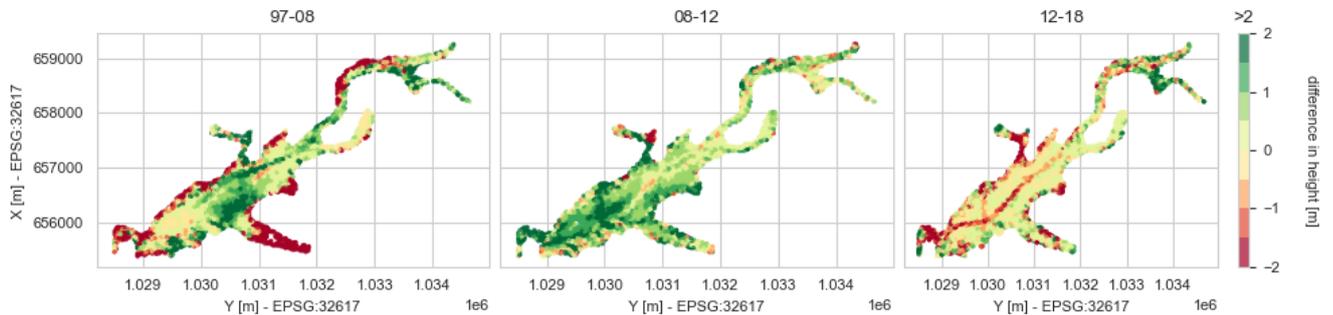


Figure 4.11: Differences in height of Río Pequení Mouth between different years

The connection to the main basin of the lake is then considered the mouth of the Río Pequení to Lake Alajuela. Looking to the entrance of this upper section of the lake to the rest of the lake, a clear sediment discharge is seen which then increases when water level is low as sediment erodes from the northern section as seen in Figure 4.11. The differences between the years are shown in Figure 4.12.

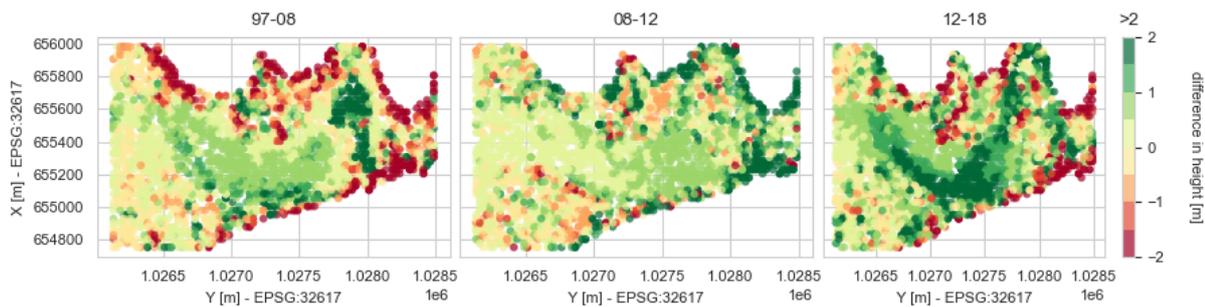


Figure 4.12: Differences in height of Río Pequení extended into main reservoir between different years

As seen in the Figure 4.12 the flow of the channel here largely determines the locations at which the sediment is deposited, and do not seem to differ in extreme amounts. The last year contains a much stronger amount of sediment deposited, which is difficult for the MLA to predict as this does not occur in the training data. The variance of the location of sediment deposition is low in this area, as the flow has not changed path in great extents like it has in the mouth of the Río Chagres. The algorithm can thus predict the location of the sediment relatively accurately as was shown in Section 4.1.2, however the large fluctuation in the amount of sediment deposited in the most recent years was not predicted.

4.4.2 Río Chagres

To successfully compute future predictions for the Río Chagres, multiple attempts had to be made due to the flow path created with the runoff pattern getting stuck in the start of the river mouth. These faulty flow path computations lead to strange predictions with

extreme values and thus the model must be tuned and iterations be done to prevent this from happening. Additionally, the morphological events in the mouth of the Río Chagres are quite severe, as large amounts of water and sediment pass through the region, more so than the Río Pequeñí. The model is trained based on the data with such large events and thus predicts more extreme values as well.

The prediction was finally made in several steps, leading up to a future prediction for the year 2024, assuming regular rainfall patterns. The final result of the prediction is shown in the Figure 4.13.

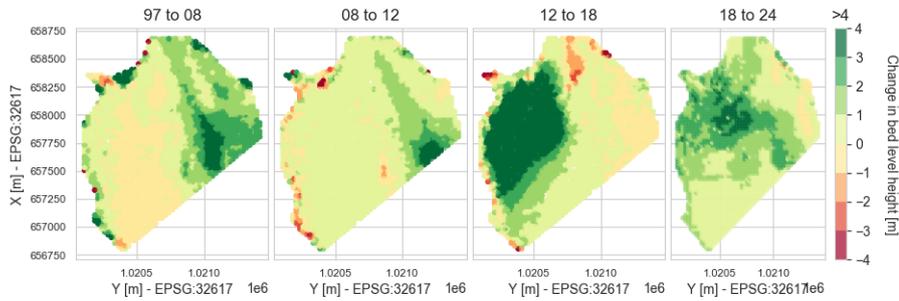


Figure 4.13: The morphological changes that have occurred in the Río Chagres mouth and the predicted changes for 6 years after the last measured data

In the Figure 4.13 the dark green areas are the predicted zones of high sedimentation levels which clearly shift location throughout the years, and again throughout the prediction. We see that after 2018, the model predicts the sedimentation front to move slightly towards the center and across the river mouth basin, whereas between 2012 and 2018 the sedimentation was almost exclusively on the left side of the basin. To analyse the movement of the predicted sedimentation in further detail, the intermediate steps are visualised in Figure 4.14.

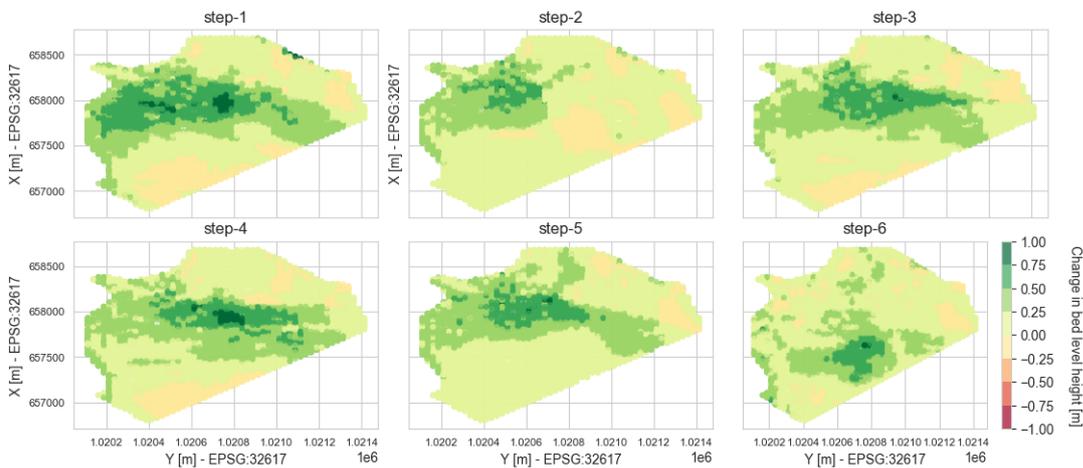


Figure 4.14: The scaled steps of that make up the Río Chagres prediction

In the steps made to predict the sedimentation that will occur over the next 6 years shown in Figure 4.14, 3 phases can be distinguished. In the first phase, a large part of the sediment arriving in the river mouth still ends up in the left side of the river mouth relative to the entering flow, as it has done in the years from 2012 to 2018. The second phase, starting in step 3, shows the sediment front to have moved to the center of the river mouth, depositing little

sediment on the left side of the river mouth. The last phase starts in step 6 of the prediction, and this step shows the sediment front to have moved downwards in the river mouth. The accumulating sediment in previous years then raised bed level in the river mouth sufficiently to increase the flow speed of the channel to carry the sediment further downstream. In 2026, the sedimentation front is thus 500 meters further downstream than it was in 2018, according to the prediction. The steps following the 6th step could not be utilized as the flow path computation got stuck at the top of the river mouth, a limitation of the methodology.

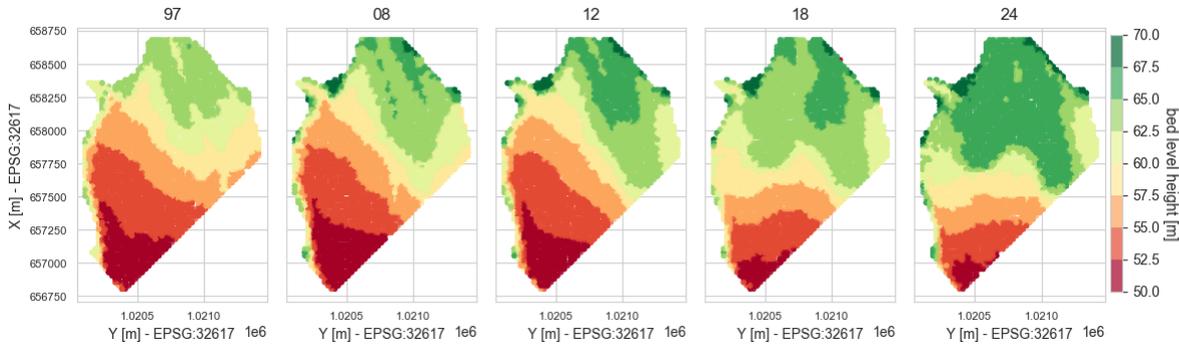


Figure 4.15: Heights of Río Chagres predicted into future years

Looking at the differences in height between the predicted years and the historical years as shown in Figure 4.15, we first analyse the differences occurring between the historical datasets. between 1997 and 2012 we see a clear gradual increase in height around the same region, being the right side of the river mouth. Between 2012 and 2018, the large shift in sedimentation location occurred, and this is visible in the height map as there is a sudden large increase of height on the left side of the delta. The results of the consecutive time steps following the historical height map of 2018 is a gradual increase of height along the fronts of the sedimentation areas, and a large increase at the left side of the delta.

The Table 4.2 below shows the volumes of sediment that have entered the Río Chagres basin between the years of collected data, as well as the predicted volume of sediment to have entered the basin between 2018 and 2024. In the Río Chagres basin large fluctuations of sedimentation are seen, the amount of sedimentation between 2012 and 2018 is more than 3 times the amount it was in the period prior. This is due to the landslide sediment moving down river into the basin, which was previously washed into the river in the rainy season between 2008 and 2012. This shows that natural events can be used to predict these extreme amounts of sediment entering the reservoir, as the sediment requires several months or years to move downstream. The predicted period shows a lower amount of sedimentation, more along the magnitude of the years between 1997 and 2012, which is a realistic result as the years before 2018 have not brought any extreme rainfall or landslides and thus large amounts of sedimentation are not expected.

Period	Volume of sedimentation	
97-08	$4464 * 10^3$	$[m^3]$
08-12	$3074 * 10^3$	$[m^3]$
12-18	$9755 * 10^3$	$[m^3]$
18-24	$3011 * 10^3$	$[m^3]$

Table 4.2: Volumes of sedimentation in Río Chagres basin

A key aspect which can be observed in the height maps is the sediment front which gets

“pushed forwards” every year. Each time the river takes a new flow path, the location of deposition of the sediment is moved. The downstream movement of this front is clearly visible between the years 1997, 2008 and 2012, while between 2012 and 2018 the movement was mainly a sideways movement. The prediction has thus shown an initial sideways movement of the front followed by a downstream movement. In the height map a general increase of bed level height is visible, but more importantly there is a strong increase in height between the two fronts. The 6 years in the prediction show the gradual movement of the front, raising the bed level over a broad area until the 6th step. The step 6 of the prediction then shows the location of what will likely be the next large deposition front.

The use of time steps restricts extreme values in the prediction as the flow path is able to move during the intermittent steps of the prediction along with the changes occurring to other features. This then allows for the sediment front to move during the prediction as it would naturally, for as long as the model allows it and features are able to be computed.

4.4.3 Río Pequení

To compute the future predictions for the Río Pequení, very little work had to be done regarding the setting and shifting of bounds and parameters. In the figures 4.16 and 4.17 the results of the sedimentation prediction for the mouth the Río Pequení are shown. This prediction is done using a SVR model with a set of 12 features.

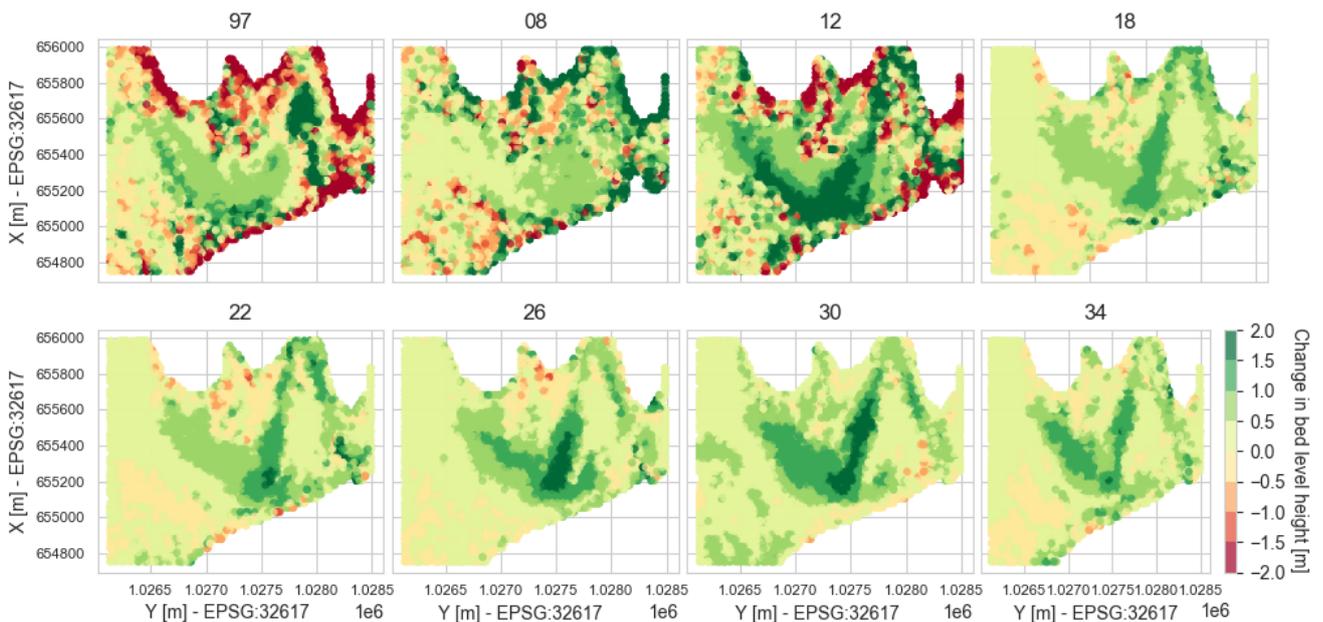


Figure 4.16: Differences in height of Río Pequení predicted for periods into future years

As seen in the Figure 4.16, the location of sedimentation has historically been in one curve of the river entering the reservoir, and remains in this curve for the prediction.

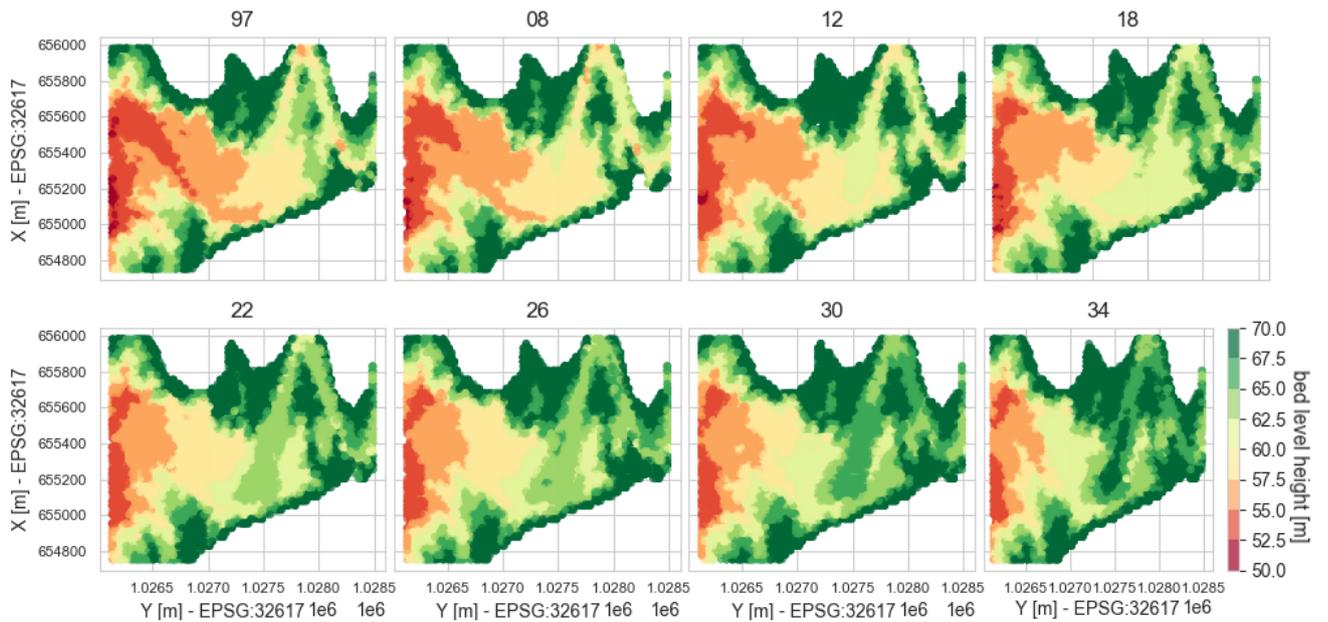


Figure 4.17: Heights of Río Pequení predicted into future years

Note that the location of the sedimentation in the end of the curve slightly shifts, as the runoff pattern moves with the deepest part of the channel. Besides this, the geological characteristics of this area ensure that the sedimentation locations vary relatively little. The shift of the sedimentation location is due to the channel shape moving throughout the years as erosion and sedimentation change the terrain. Whereas historically the largest amount of sedimentation has been occurring in the outer reach of the river bend, the model predicts future sedimentation to be centered around the inside of the river bend as the flow path adjusts to the morphological responses. As the sedimentation is stacked up around the center and outer area of the bend, the flow path is pushed toward the inside of the channel and thus placing future sedimentation in this region.

In the Table 4.3 below the volume changes over the different periods are presented. As the Río Pequení predictions are much more stable, these predictions go much further into future years, and the volume of the predictions as well as the location of the sediment remains around the bounds of the training data. The model thus predicts a stable and consistent continuation of the sediment deposition in the same areas it has been deposited in the previous 20+ years.

Period	Volume of sedimentation	
97-08	$753 * 10^3$	$[m^3]$
08-12	$1752 * 10^3$	$[m^3]$
12-18	$1534 * 10^3$	$[m^3]$
18-22	$1053 * 10^3$	$[m^3]$
22-26	$920 * 10^3$	$[m^3]$
26-30	$1346 * 10^3$	$[m^3]$
30-34	$2096 * 10^3$	$[m^3]$
34-38	$1083 * 10^3$	$[m^3]$

Table 4.3: Volumes of sedimentation in Río Pequení basin

4.5 Reproducibility

The methodology used in this thesis is highly reproducible, granted that there is sufficient historical data available with small enough resolution. The complete code used in the research is available on Github, and the individual functions and algorithms can be used to reconstruct a similar pipeline on a different platform. With basic knowledge of Unity, code can be implemented and improved in its current architecture. As this code and project in itself is highly specialized for reservoir, and even one reservoir in particular, the code cannot out of the box for any scenario.

To apply the code to a reservoir, some time will have to be invested into setting up the code in Unity to the point where the project will successfully compile, and the settings will need to be tuned to ensure correct flow path computations. As reservoir sedimentation is such a large scale problem though, the actual time necessary to invest in order to apply the code of this research is relatively very little compared to the insight it provides. All libraries used are open, and the pre-processing done in FME is a simple interpolation and projection on a grid which can be done in any GIS tool and is not a sensitive or complicated step in the pipeline. Although the data used is not public, any DEM dataset covering multiple years of bathymetry will be able to produce results with the use of this pipeline and provide insightful predictions.

Chapter 5

Conclusion and Discussion

The disciplines of machine learning and river morphology have been connected in this thesis using various aspects within the field of Geomatics. This has produced a visually comprehensible prediction of future sedimentation in Lake Alajuela with far more detail than earlier made predictions, and one that can be verified in next planned bathymetric survey. The issue of sedimentation in Lake Alajuela and the prediction thereof is now more than a rough number of years, but a visible front of sedimentation entering the reservoir from the water supplying river. With this visualized prediction, the future priorities within the organisation of the Panama Canal can be shifted towards finding a solution for this problem. The conclusions of this research are discussed in the sections of this chapter.

5.1 Research overview

The aim of this thesis was to develop a pipeline for predicting sedimentation in the Lake Alajuela reservoir using machine learning techniques. To fill the gap of knowledge on the sedimentation process occurring in the lake, and the exact amount of aggregated sediment present there, this thesis provides insight into this process and the continuation thereof. No accurate prediction on sedimentation occurring in the lake has been made, while this information is required for developments in establishing a future for the Panama Canal. Morphological predictions with the help of ML techniques is an area that is being researched more and more in recent years, however the majority is concentrated in above water surface erosion processes and river channel morphology. These morphological processes are distinctly different than the processes occurring in Lake Alajuela, and thus required a tailored approach. To fulfill the aim of the research, a main research question was defined with four sub-questions, which are reviewed below.

How to accurately predict sedimentation levels in the Lake Alajuela reservoir using a Machine Learning method?

The research in this thesis has provided several factors that have shown to be of high importance. First off, the areas of importance have shown to be exclusively the mouths of the two rivers entering the lake. The large majority of new sediment settles in these regions, and thus these are the areas of interest when researching the sediment entering the reservoir. The behavior between the two river mouths differs due to the magnitude of water and sediment entering the rivers, the geological positions, and properties of the terrain. Due to this difference the machine learning models are best trained individually per region of study to be able to predict this behavior in the future. A range of features are used for the prediction,

and depending on the characteristics of the river mouth, different characteristics are of higher importance.

- *Which sedimentation related features can be extracted from the DEM?*

Using a runoff algorithm to locate the deepest part of the lake, several features can be extracted. The runoff algorithm follows the natural flow direction, from which per step the length of the flow path is computed. Every cell in the grid is then given the closest flow path segment as a feature, with the length of the path at that segment, distance to the path, and the relative angle of the slope of the cell to the angle the flow path.

The runoff score is also added to grid cells, providing the amount of flows passing through that cell when a full grid runoff simulation is performed. Additionally, the relative height parameter aids to recognize bumps and pits in the terrain on various scales. The curvature is extracted on multiple scales, these scales being the curvature computed using cells at a certain radius. Lastly, the more standard geometric features such as slope, aspect, height and depth are added.

- *Which ML model best predicts sedimentation in a reservoir?*

To assess the results of the predictions made by the ML models, two metrics were used: RMSE and a probabilistic accuracy assessment. The lowest RMSE scores were obtained by the MLPR for both the Río Chagres and Río Pequení cases, however these were obtained with a small selection of features, and the visual representation of the prediction was not realistic. The RFR has an overall higher RMSE, and the local probabilistic accuracy is much lower than the SVR and the MLPR. The SVR on the other hand had only a slightly higher RMSE than the lowest RMSE obtained by the MLPR, but maintained the low RMSE value for a larger range of feature selection numbers, making it more robust to changes in feature selections. Additionally, the SVR has a higher local probabilistic accuracy, meaning it is more likely to predict the correct locations for future sedimentation to occur. The robustness and accuracy of the SVR thus make it the best model suited to predict sedimentation in the tested scenarios.

- *What is the best set of geomorphological and hydrological features to train a ML model for prediction of sedimentation?*

It is crucial for the model to know the flow path of the river, since this is a deciding factor for the deposition location of sediment. The features connected to this path are thus crucial in any prediction since it marks the areas that are most susceptible to sedimentation. The relative height will mark features in the terrain that will cause an instant increase or decrease in flow resulting in morphological response. Curvature indicates an acceleration or deceleration of flow, resulting in direct morphological changes above the surface of any body of water. In a reservoir, this effect is dependent on the flow in that particular region, and is thus not as strong of an indicator by itself as for erosion above the surface. These features on different scales can provide the full coverage for both river mouths of the Lake Alajuela reservoir, however each responds more strongly to a different combination of these scaled features due to the difference in terrain size and properties. The two rivers supplying Lake Alajuela of water and sediment are of very different magnitudes, and exist along different geological characteristics. As a result, the sedimentation in these areas is predicted with different sets of features, however both making use of the general features mentioned above but on different scales.

- *What accuracy can be obtained predicting sedimentation in the Lake Alajuela reservoir?*

To assess the accuracy of the predictions made, the probabilistic accuracy and the nu-

meric accuracy are considered. The probabilistic accuracy can be assessed visually, and from the tests done can be seen that the regions at which the sediment will be deposited is accurately provided for purposes such as dredging and local solutions. The accuracy of the probabilistic prediction for locations to have high amounts of sedimentation is up to 80 percent, depending on the location and the curb set for the high sedimentation level classification.

To assess the numeric accuracy, the RMSE metric is used to compare the overall accuracy of the different predictions. In both river mouths, the RMSE score ends relatively high. For the Río Chagres, the average obtained RMSE is around 1.5 meters, while for the Río Pequení this is 0.8 meters. The difference in the RMSE for these rivers is due to the fluctuation in the sediment flux, and thus the sedimentation, occurring between the years of the training and testing data set. The Río Chagres had a much larger amount of sedimentation occurring in the years of the testing data than the training data years. The lack in variety of years in the datasets therefore hindered the numeric accuracy obtainable.

5.2 Contributions

This thesis builds upon previous researches done in the field of morphological process modelling with machine learning methods, and provides insight into the sedimentation occurring in Lake Alajuela with a prediction towards future sedimentation. The main areas in which this thesis contributes to the current state of art and knowledge base are listed below.

- *Time steps in prediction using ML model*

An important contribution of this thesis is the pipeline through which a machine learning model is used to predict morphological changes over time in multiple time steps, or iterations, instead of just a single prediction. This new methodology is also gateway to predicting local quantities with high accuracy, as this approach allows the modelling to depend on entirely different parameters than traditional methods.

The step wise methodology of the prediction limits the predicted values from becoming too large, as larger values will affect the features for the consecutive time steps. This means that the quantities and locations can be predicted to a certain extent, leaving the only unknown the time variable, as the amount of rainfall and sediment discharge in future years is not known. Due to the model re-evaluating the situation at intermediate steps, large errors are prevented as all errors are essentially scaled down and are less likely to affect the entire prediction.

- *Local sedimentation prediction in reservoirs*

The majority of research done on local morphological predictions has been in rivers and channels, where the width and depth parameters are relatively steady. Reservoir sedimentation predictions have for the most part been focused on total added sediment in the reservoir. Local sedimentation in reservoir cannot be predicted with traditional methods as the width and depth parameters present in channels are no longer valid.

In the research of this thesis a pipeline is developed to predict the local sedimentation levels in a reservoir, providing a tested accuracy between the 70 and 80% depending on the location and the individual trained model.

- *Morphological processes in Lake Alajuela*

The state of knowledge of the morphological processes occurring in the Lake Alajuela

was relatively limited as the sedimentation and erosion occurring are often seen as results of extreme natural events. Hurricanes and draughts cause morphological processes to be relatively unpredictable, however this thesis uses a range of data over 20+ years to oversee these events and predict future morphological events.

The research done in this thesis thus provides insight into possible consequence of years of morphological processes occurring in the entire watershed, the result of which ending in the Lake Alajuela reservoir.

- *Morphological predictions using runoff model features*

Morphological predictions with machine learning techniques have been done with a range of different features. In this thesis however, new features have been developed using a runoff model created with the historical DEMs available, and the predicted DEMs consecutively when a prediction is made with time steps.

These features that use the runoff model are developed to help the MLA recognize the possible discharge of river flow that remains in a specific location as the river enters the reservoir. The runoff models are inherently used for predicting erosion above the surface of large bodies of water, however in the case of predicting sedimentation under the surface of the reservoir, the runoff model has proven valuable.

5.3 Limitations

The methodology developed in this thesis contains a number of limitations, listed and elaborated below.

- *Available Data*

The lack of variety in the temporal range of the training data does not allow the model to take a time measure as input. In order to predict a specific amount of time, the discharge between the training data years needs to be used in the analysis and thus the result computed indirectly. A dataset with a more elaborate timeline would provide the opportunity to use the time interval or the actual date as a feature with the training data, and allowing the interval and date to be provided for prediction inputs as well.

- *Runoff model failure*

One of the limitations of the current methodology is that it is highly dependent on the runoff model. This runoff model is initiated at the mouth of the river, and should flow into the center of the lake. When several predictions in time steps have been made and the sedimentation stacks up around the river mouth as a result, the runoff model can essentially get "stuck". This means the runoff model no longer flows into the lake but stays close to the initial position of the river mouth.

- *Data cleaning*

In order to ensure that the datasets could be compared to one another at all points and flawlessly pass through the pipeline, the data was extensively cleaned beforehand. This meant clipping the DEMs on one another and projecting these all on the same grid. In this process, data is lost by clipping, interpolating and rounding. If more of the original data could be kept, better input data could be provided at the start of the pipeline providing perhaps more accurate results.

- *Border of DEM*

For certain features, specifically those using surrounding cells to compute smoothed or

averaged values, the cells around the borders of the DEM will have errors in feature computations. This border around the DEM affects the training of the model as well as the predictions made with the model. This is predominantly visible in the large scale curvature feature where a large margin is available due to this feature not averaging the values but using the cells at further distances. The curvatures of cells within a certain distance of the border have the same value with one the the values used being the value given outside the DEM which is zero.

5.4 Recommendations and future work

For any future work involving the modelling of morphological changes on terrain using machine learning algorithms, several recommendations can be made and proposals for future work have emerged. These will be explained in further detail below.

- *Larger variety in data on temporal range*
A similar methodology can be utilized in an area where there is a data set available with a larger variety of data over time. Where the small amount of different years of data has been a limiting factor in the this research, it would be highly interesting to see the improvement more time steps in the training data will have on the MLA. For a data set containing a larger amount of years available, it would then also be interesting to see the effect of the hydrological data since the correlation could then be better recognized by the MLA.
- *Above-surface erosion application*
Research can be done on the effectiveness of the methodology used in this research on above water surface erosion processes. Since the methodology is using the runoff and curvature parameters to train the MLA, the application for erosion modelling could be of interest.
- *Improved flow path computation*
An improvement can be researched for the creating of the flow path. As the pattern now works with 8 neighbours, the angle range is limited and thus the flow is not taking a natural form. Due to the fact that this parameter plays such an important role in the methodology, an improvement in this area could benefit the algorithm greatly.
- *More elaborate depth measurement data*
The effect of the depth on the morphological processes can be researched. For a dataset with more coverage over the temporal range, and depth measurements per time period, the effect of depth can be much better utilized. In the results from this research, and the analysis of the case of the upper Lake Alajuela basin, it is clear that the depth of the water level plays a deciding role. By expanding the research with the broader range of depth measurements not only can it improve the training of the algorithm, but predictions can be made for the morphological responses in the lake for different levels of water depth over a period of time.
- *Step-wise morphology predictions with machine learning model*
As this research was a start to an experiment in the complex field of morphological

modelling with machine learning, many stones have been left un-turned. Very interesting results have been obtained by combining the ability of machine learning models to interpret and utilise parameters in unique ways with time steps and iterations for prediction morphological landscape evolution. Even though the predictions with machine learning models are often sufficient without time step implementations, an integration of time steps and the automation thereof could bring the best of both worlds, being machine learning and numerical morphological modelling. This thesis has touched on this, however I expect a lot of innovations and improvements can be made in this area.

Appendix A

What did not work

A.1 Planform and profile curvature features

To model erosion processes above water, profile and planform curvature are important factors. The acceleration of flow is determined by the profile curvature, and acceleration of flow is an event that triggers erosive processes. On the contrary, deceleration of flow is likely to result in sedimentation. The convergence and divergence of flow which is represented by planform curvature will indirectly influence the water depth, as well as flow speed, which then influences geomorphological processes. (Panagos et al., 2015; Esri)

A.2 MAT

Initially, it was anticipated that the MAT would be a main component that would be utilized to extract features from the data. The reasoning behind this is that in landscapes with prominent features like rivers and mountains, the MAT will construct faces through the centre of the terrain, following the lowest points in the terrain being the rivers. When applying this to the Lake Alajuela bathymetry model however, it became apparent that the faces of the MAT had their origin along the edges of the lake, not representing the direction of flow. This is due to two factors. First off the bed of the lake is relatively planar, not providing enough variation in height for MAT faces to be constructed along the details of this flat lake bed. Secondly, the accuracy of the data was not high enough for the MAT to have enough points for faces along the detailed features that are of interest in this project.

A.3 Temporal features

Features containing values that are constants per year for the data, such as the date, rainfall data, or river discharge were applied when testing the model. These featured however do not contain sufficient variety within the dataset to be used in the training of the model. Instead of using this for the training the model, this data was used to analyse both the data and the prediction, to obtain a better understanding of the use case.

Appendix B

Feature importances

In the figures below, the computed Gini importance for both study areas is shown for the full collection of features when computing with the Random Forrest Regression model. Although this model is not used in the final methodology, these figures give a good overview of the impact certain features may have in the predictions, and which features are important in all cases versus only a selected few.

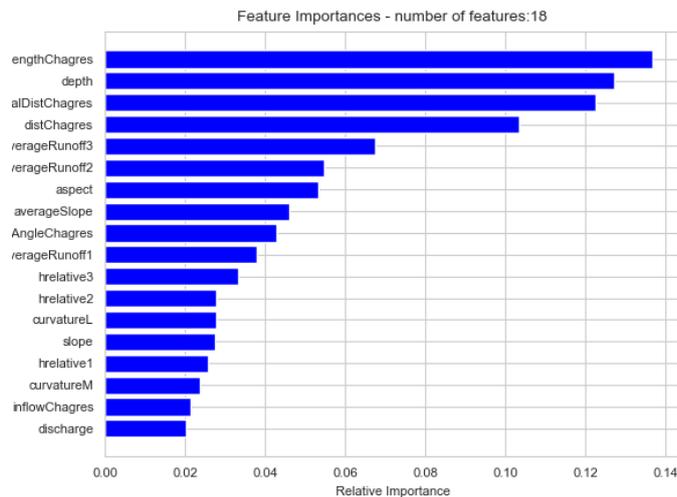


Figure B.1: Features ranked on importance for Río Chagres RFR prediction

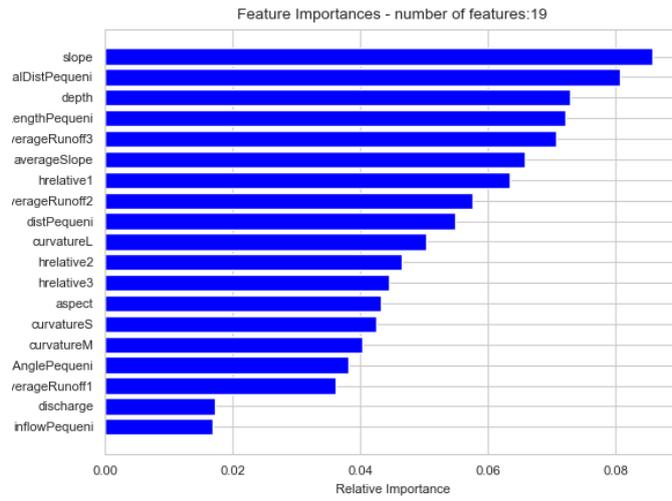


Figure B.2: Features ranked on importance for Río Pequeñí RFR prediction

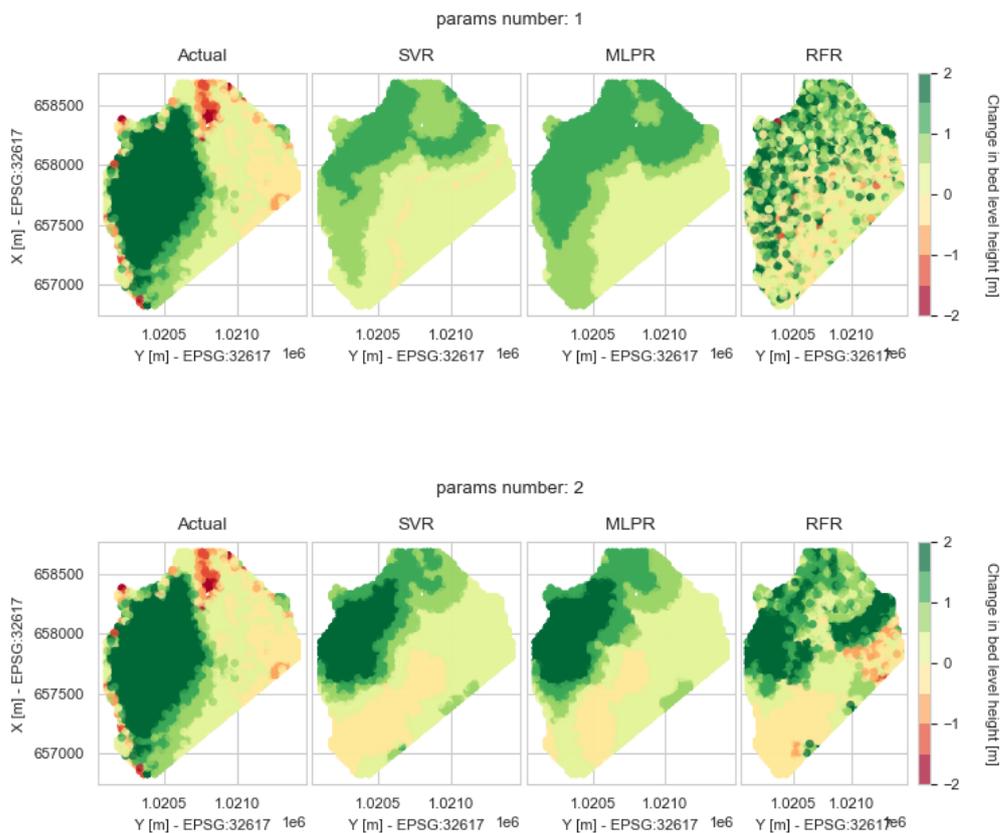
Appendix C

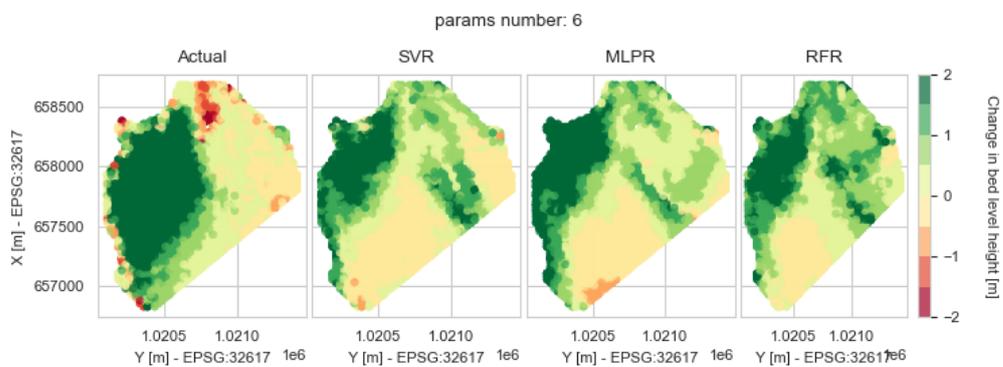
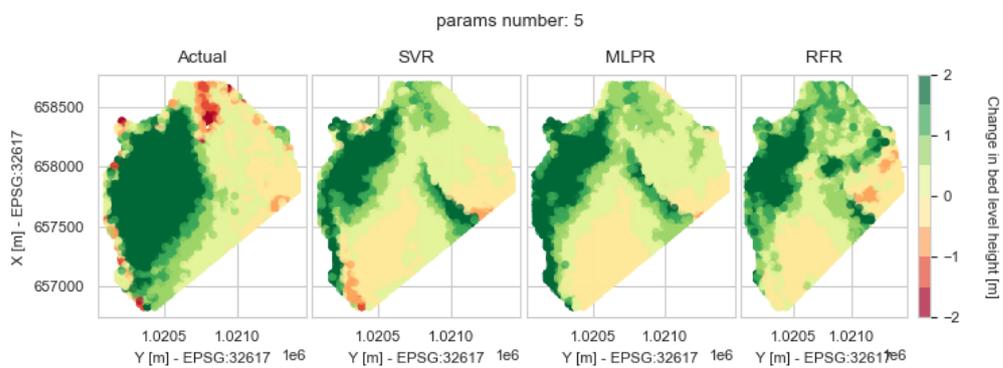
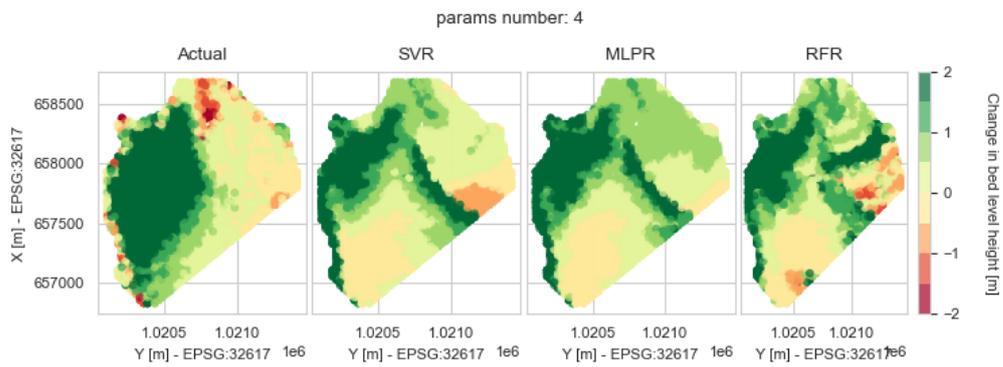
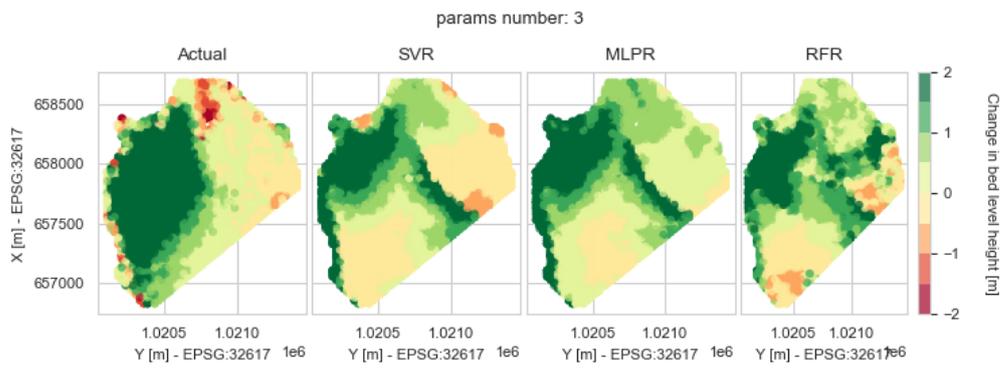
Model testing results

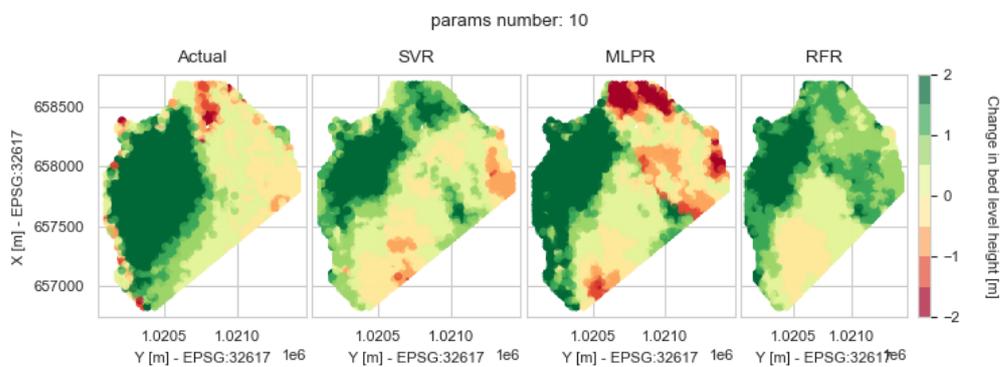
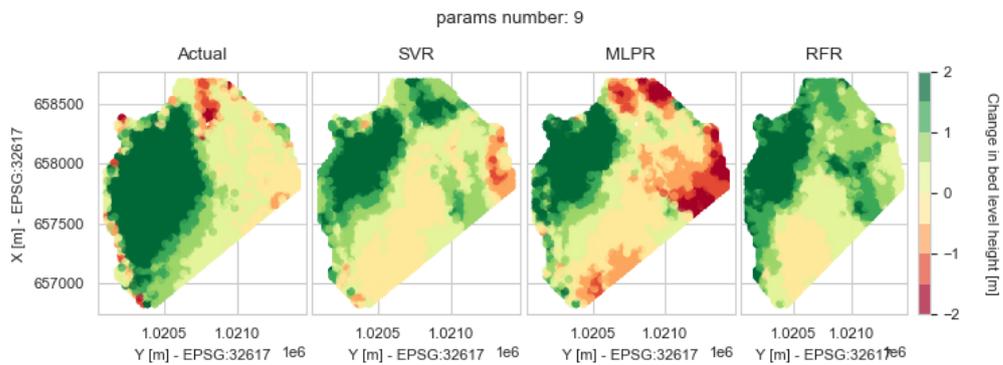
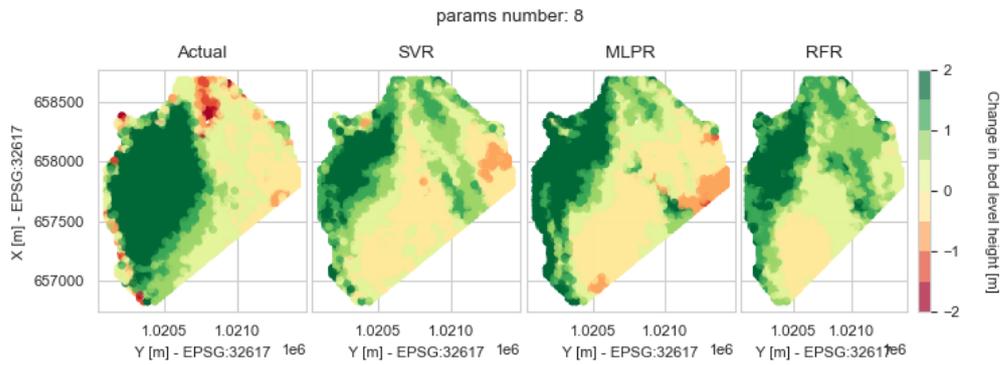
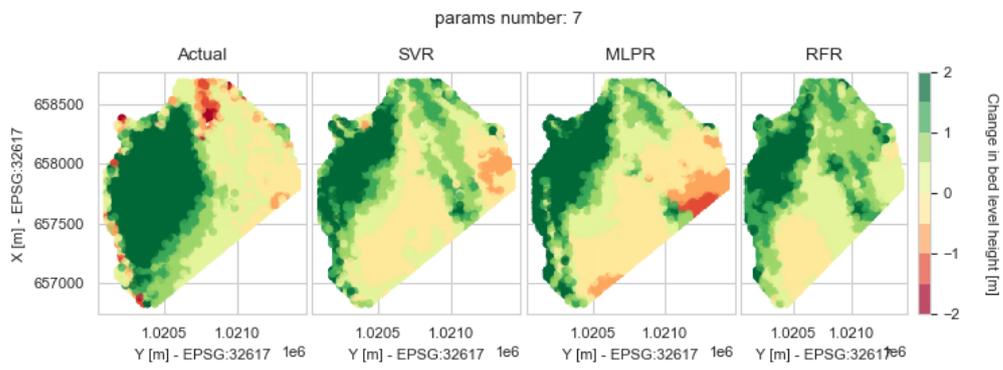
The figures below show a full range of model testing results performed which was used to analyse and decide on the model and parameter selection made. More ranges like this were performed, with slightly different sets of parameters.

C.1 Río Chagres

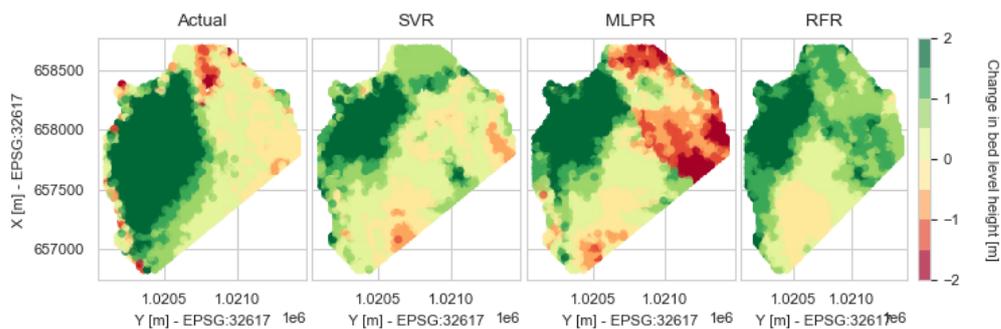
First, the Río Chagres model tests with a range of features ranked on importance is given in the following figures:



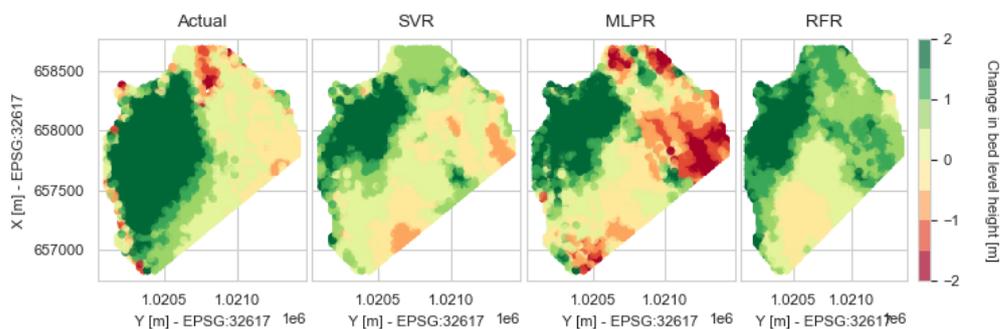




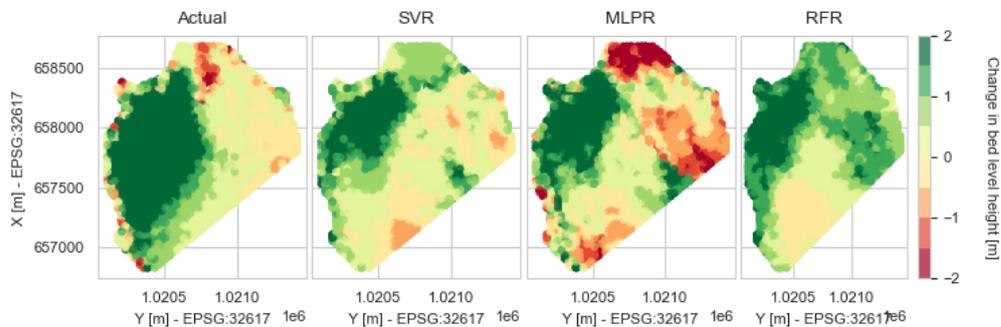
params number: 11



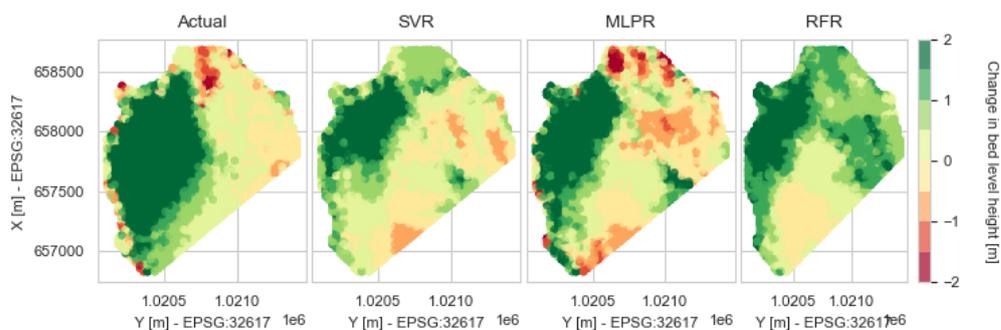
params number: 12

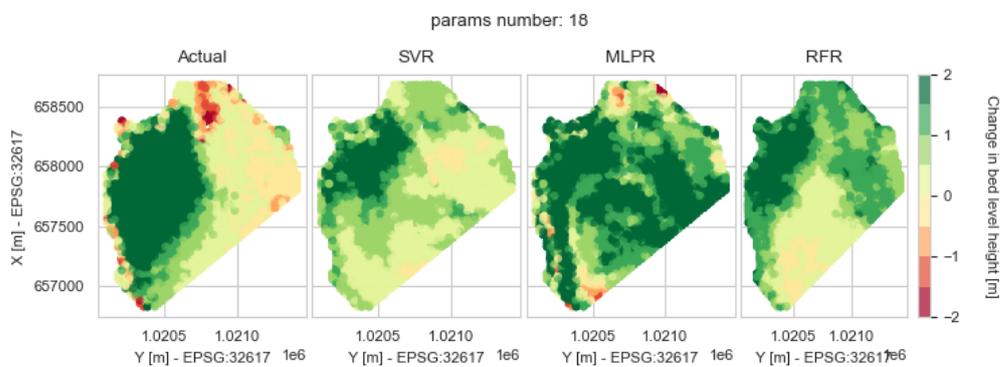
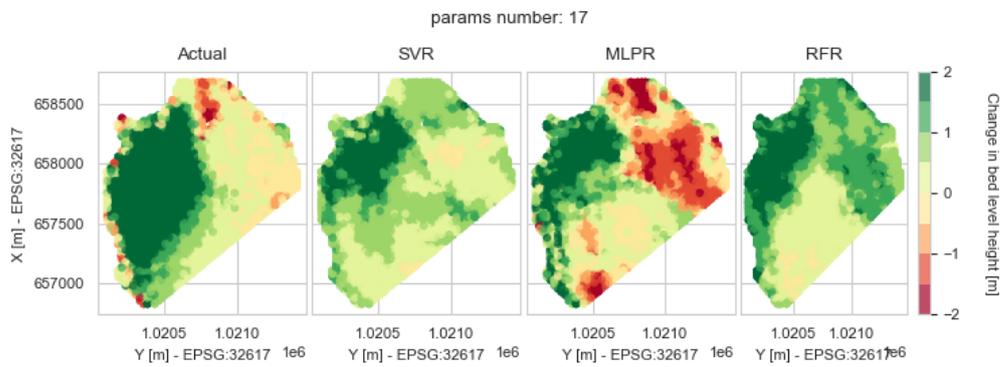
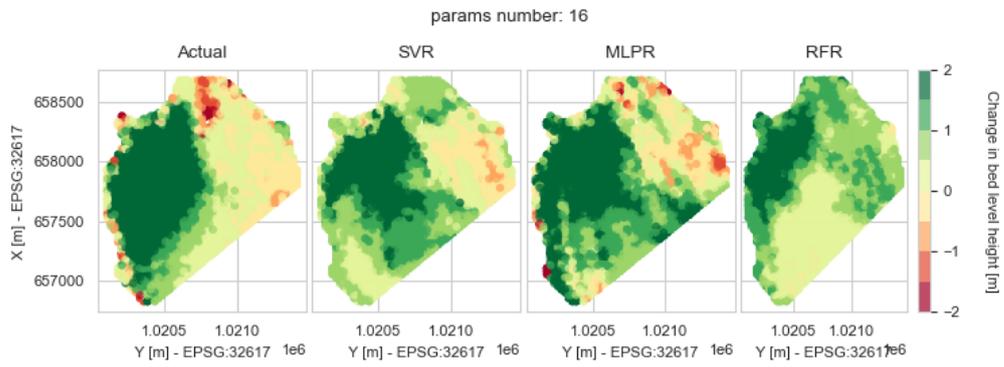
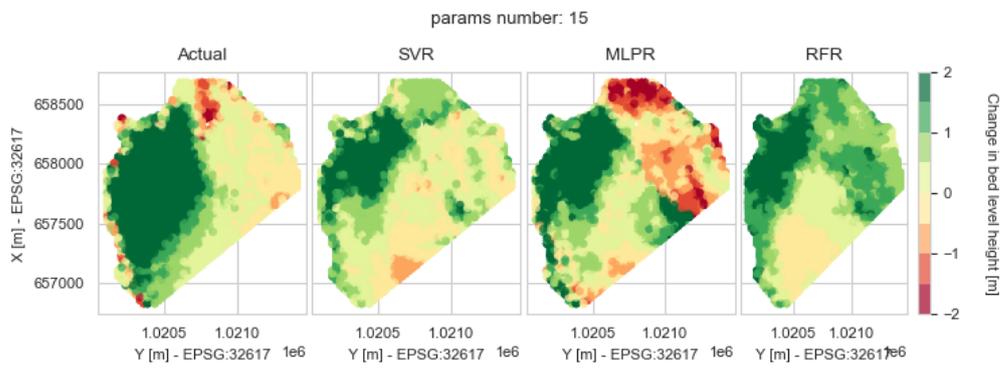


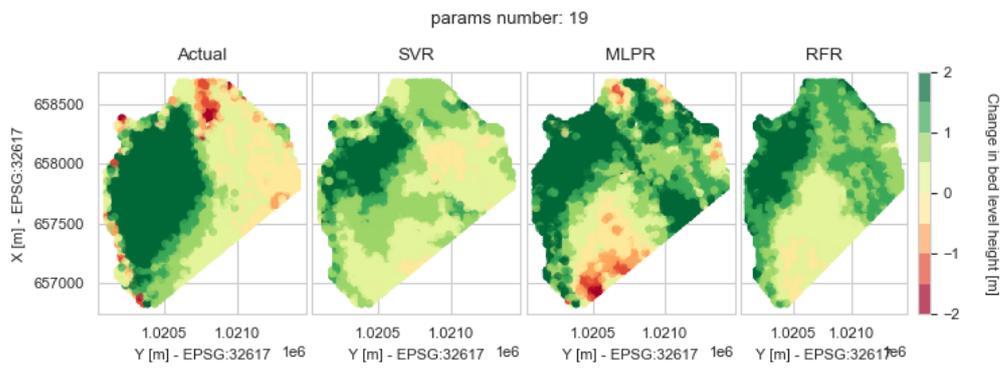
params number: 13



params number: 14

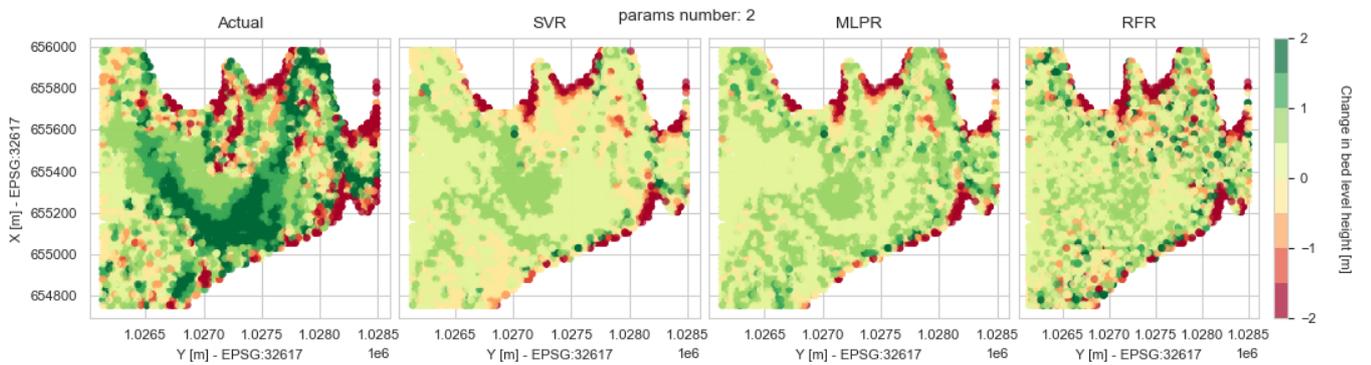
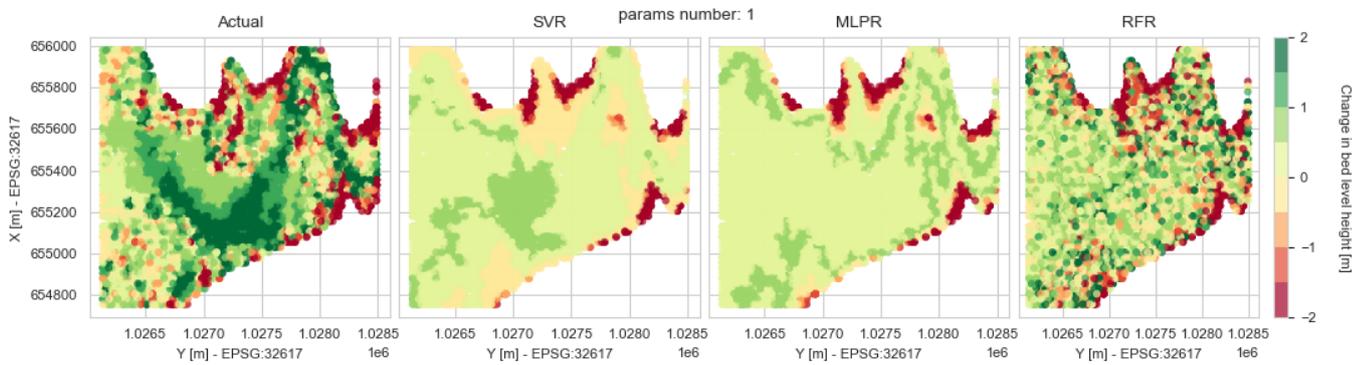


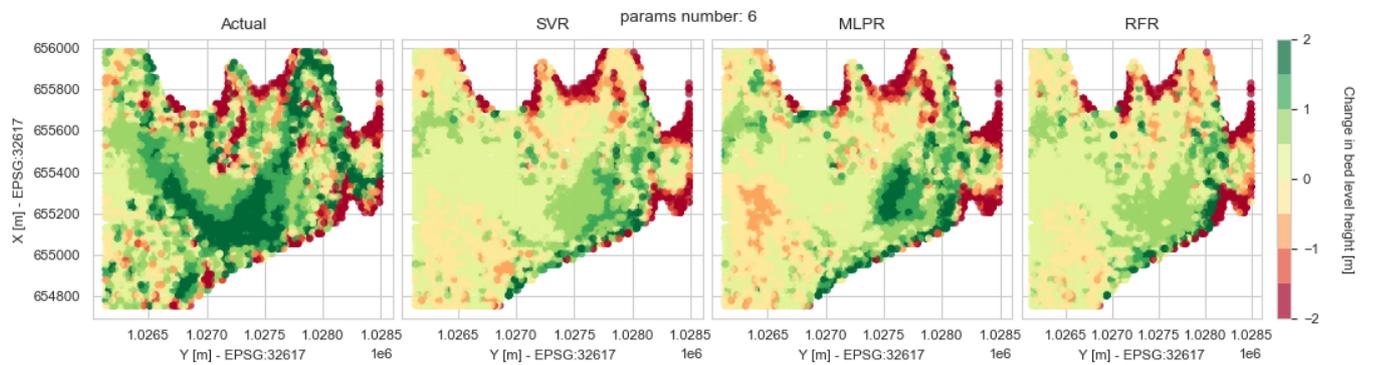
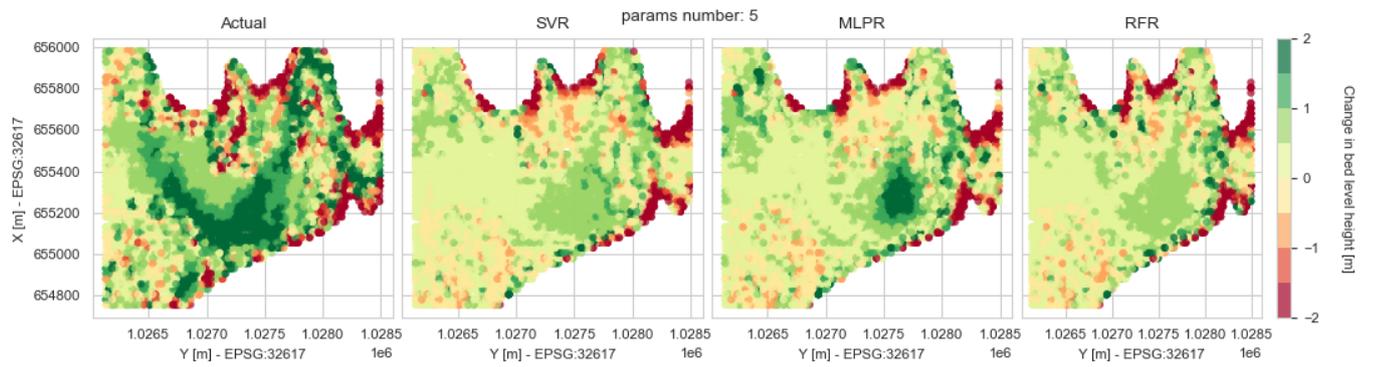
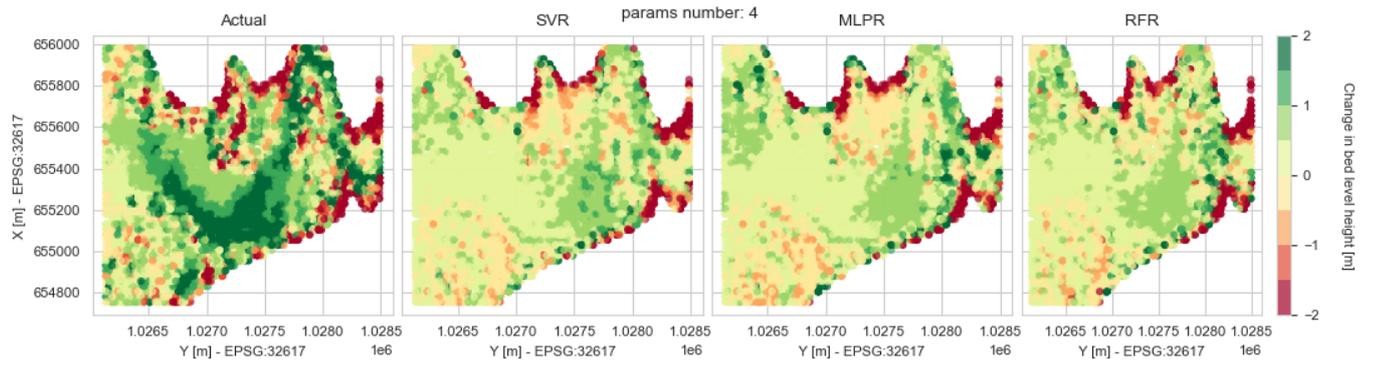
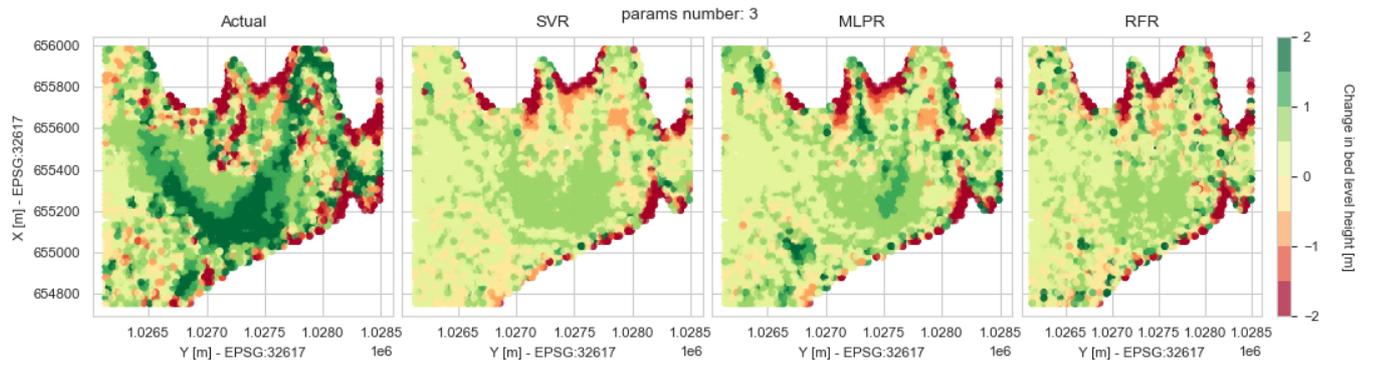


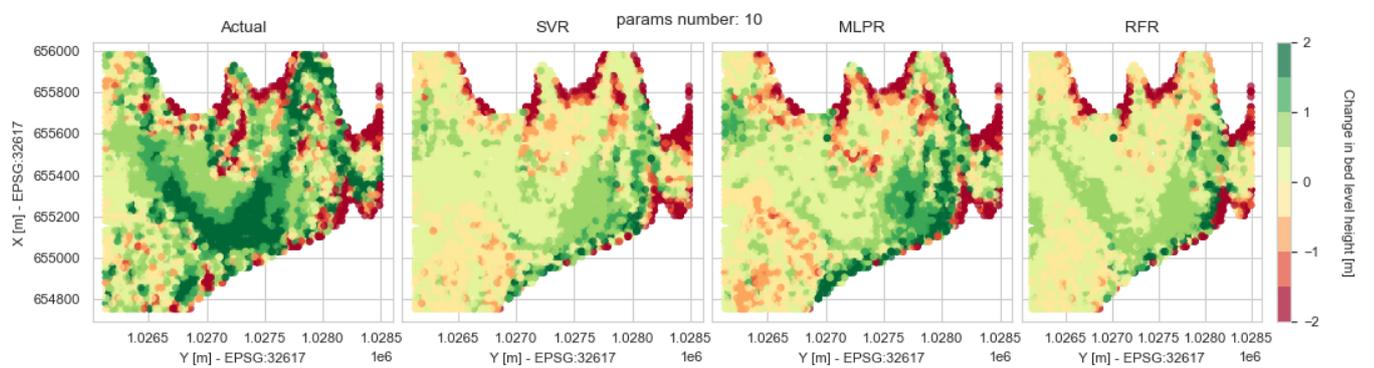
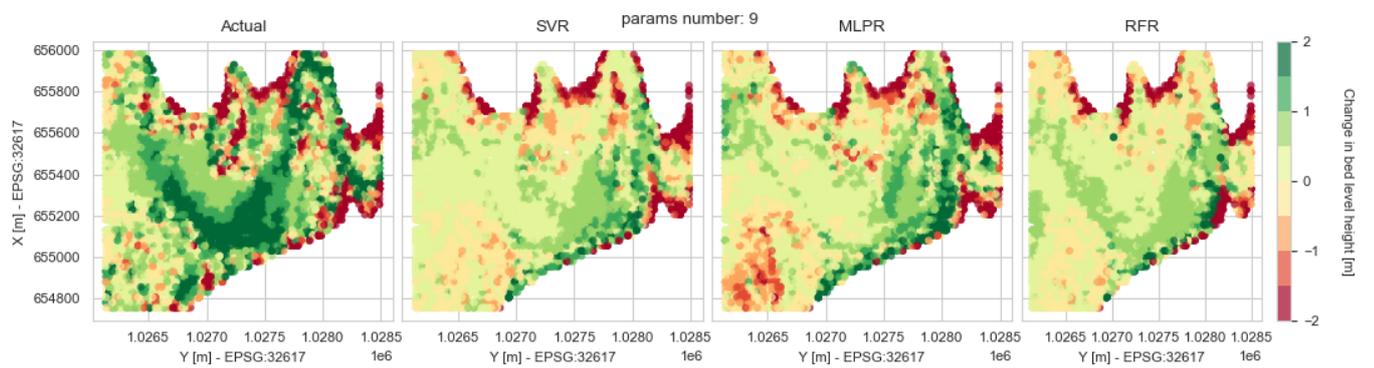
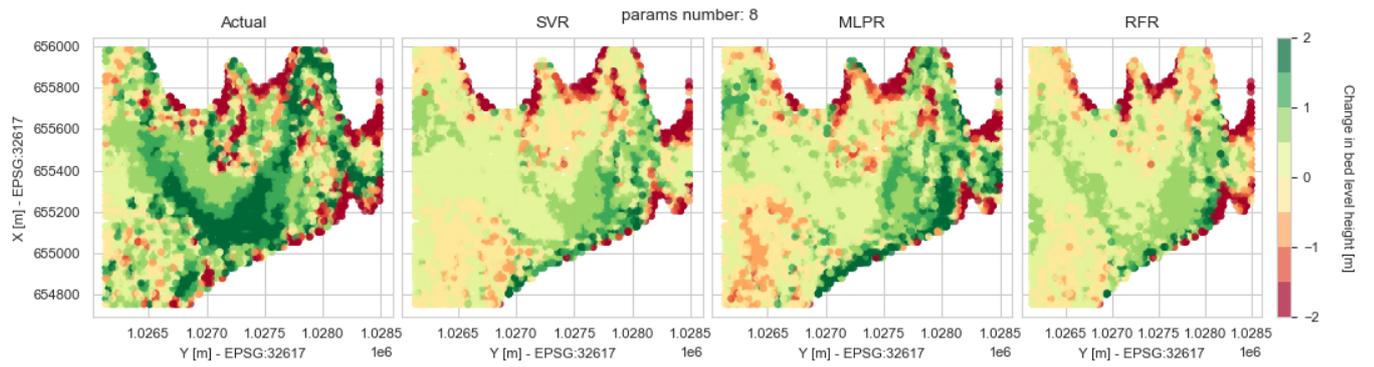
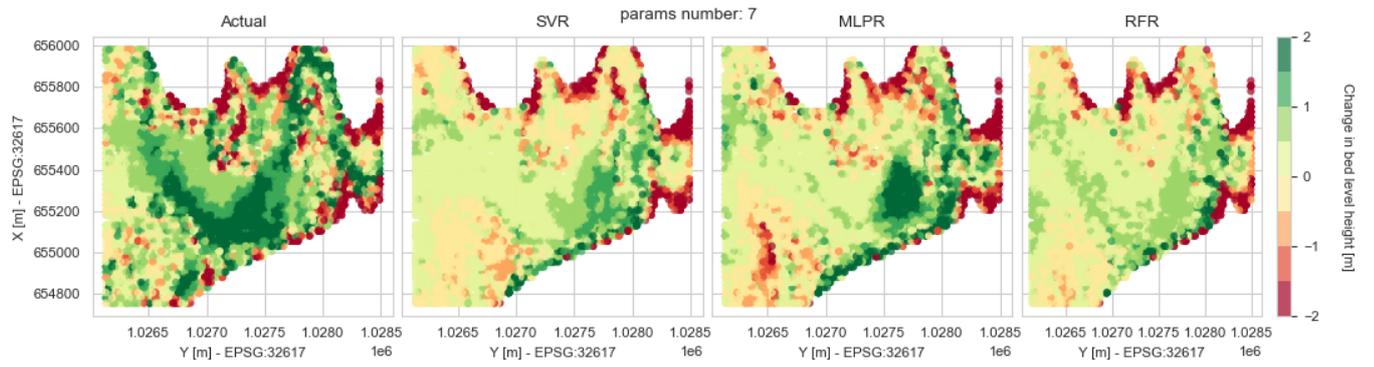


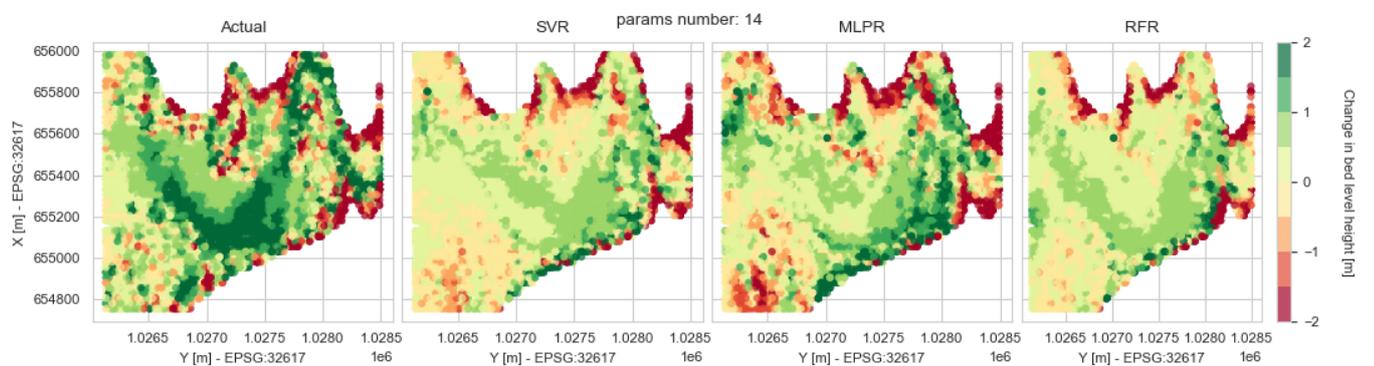
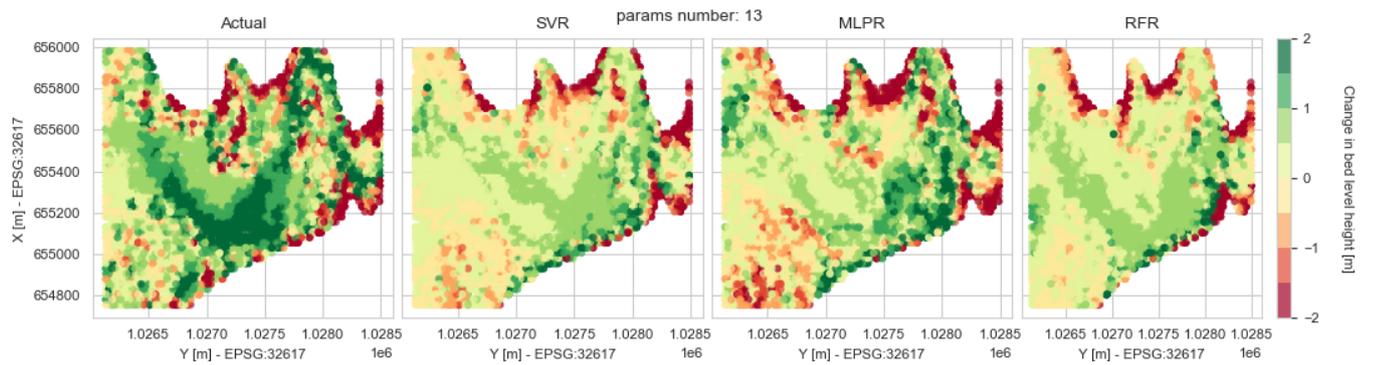
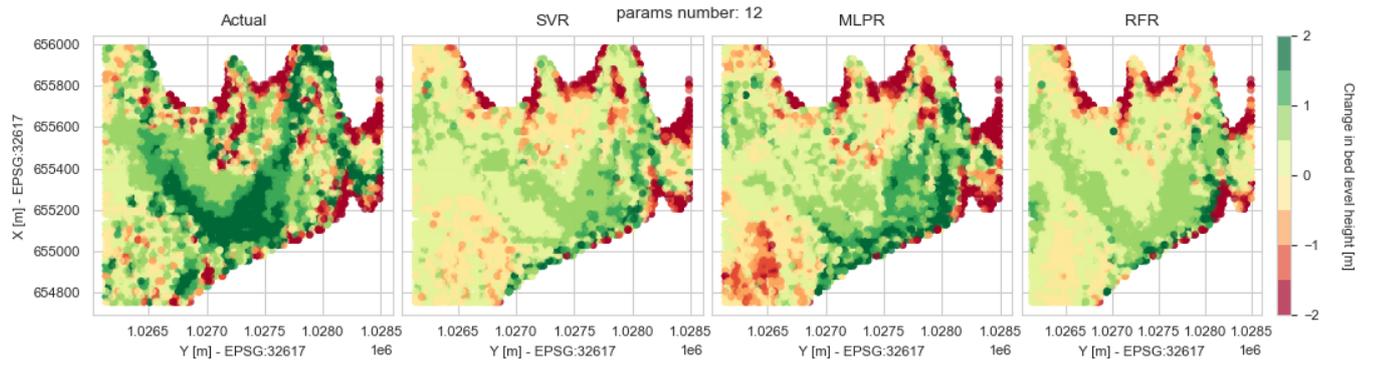
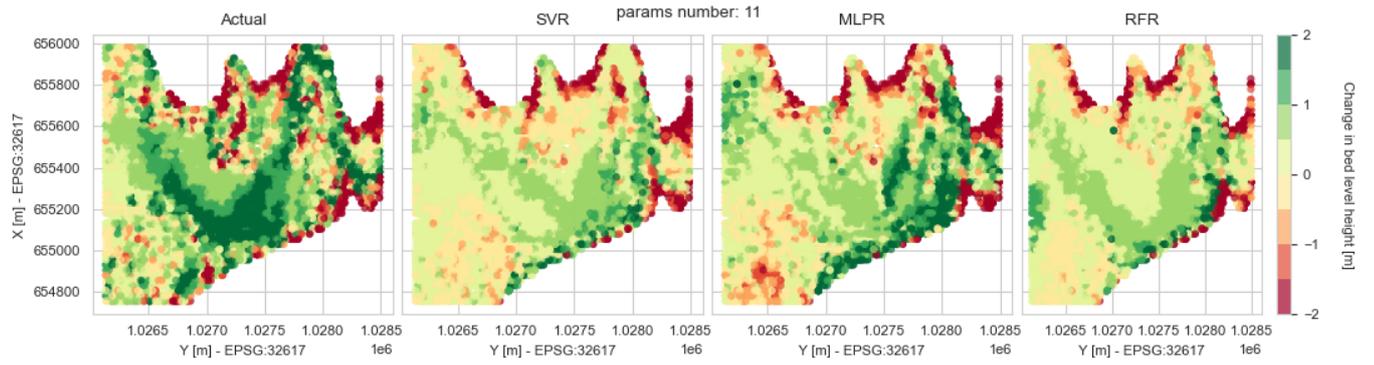
C.2 Río Pequení

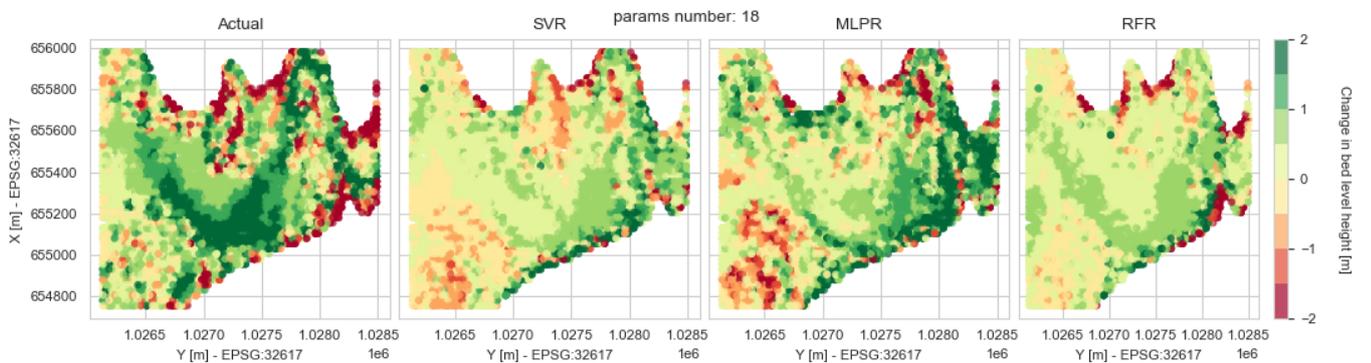
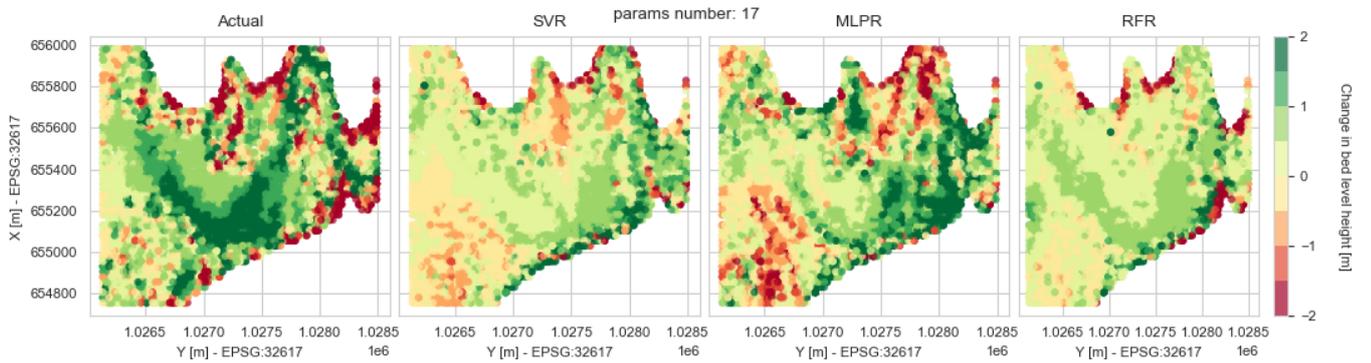
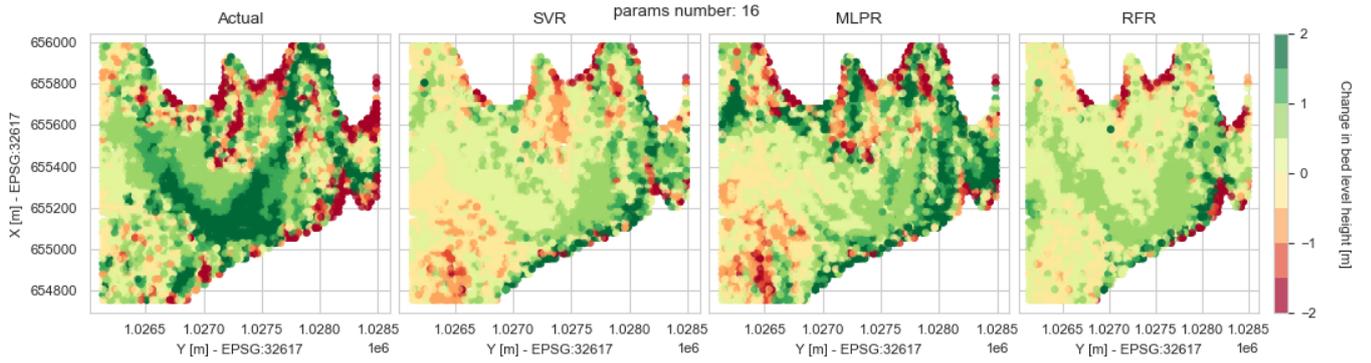
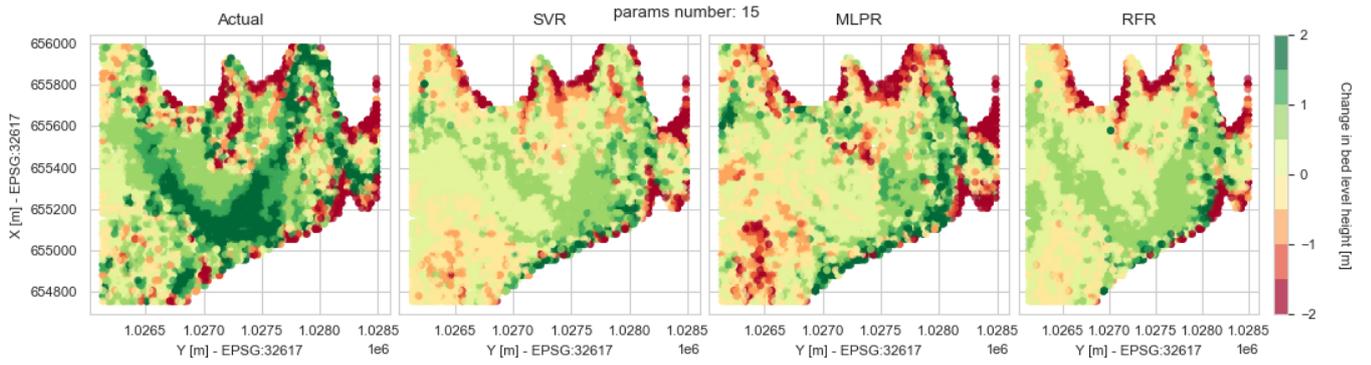
The next figures are models tests for a range of parameters in the Río Pequení basin.

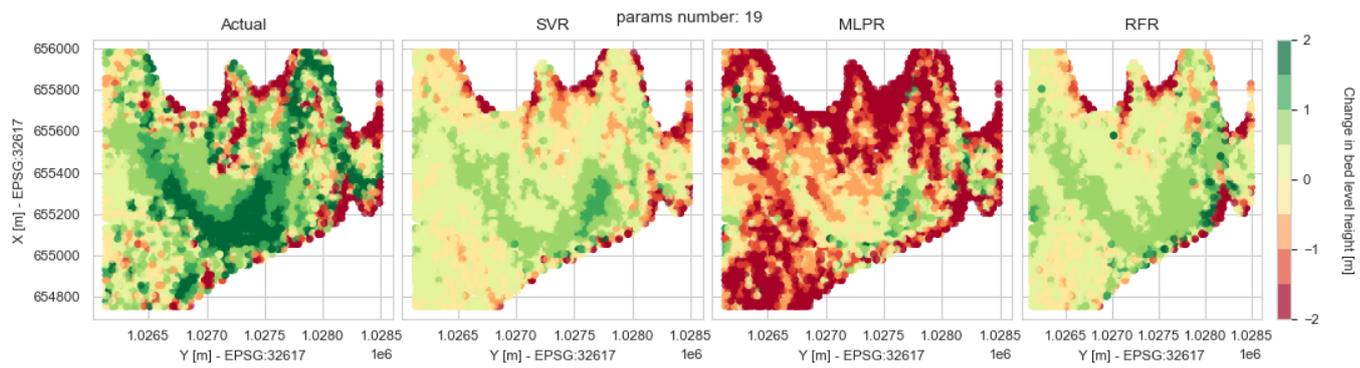












Appendix D

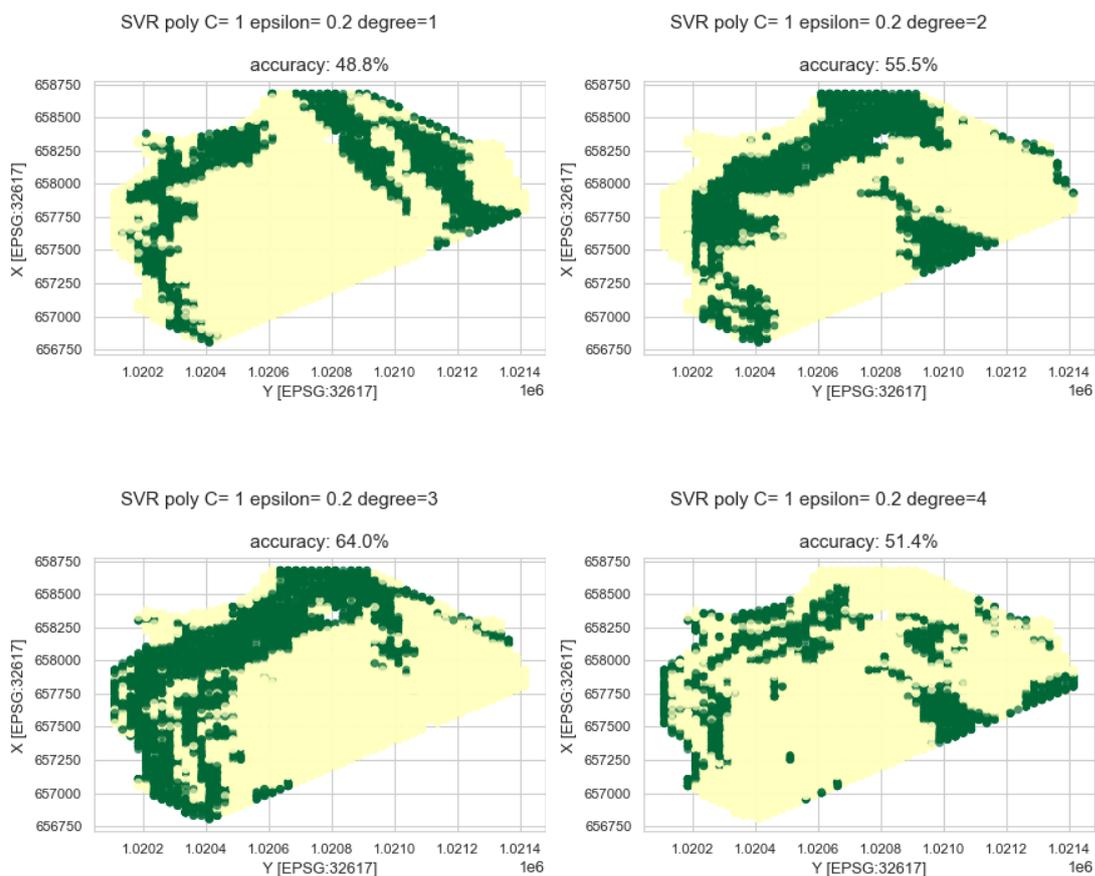
Hyperparameter tuning results

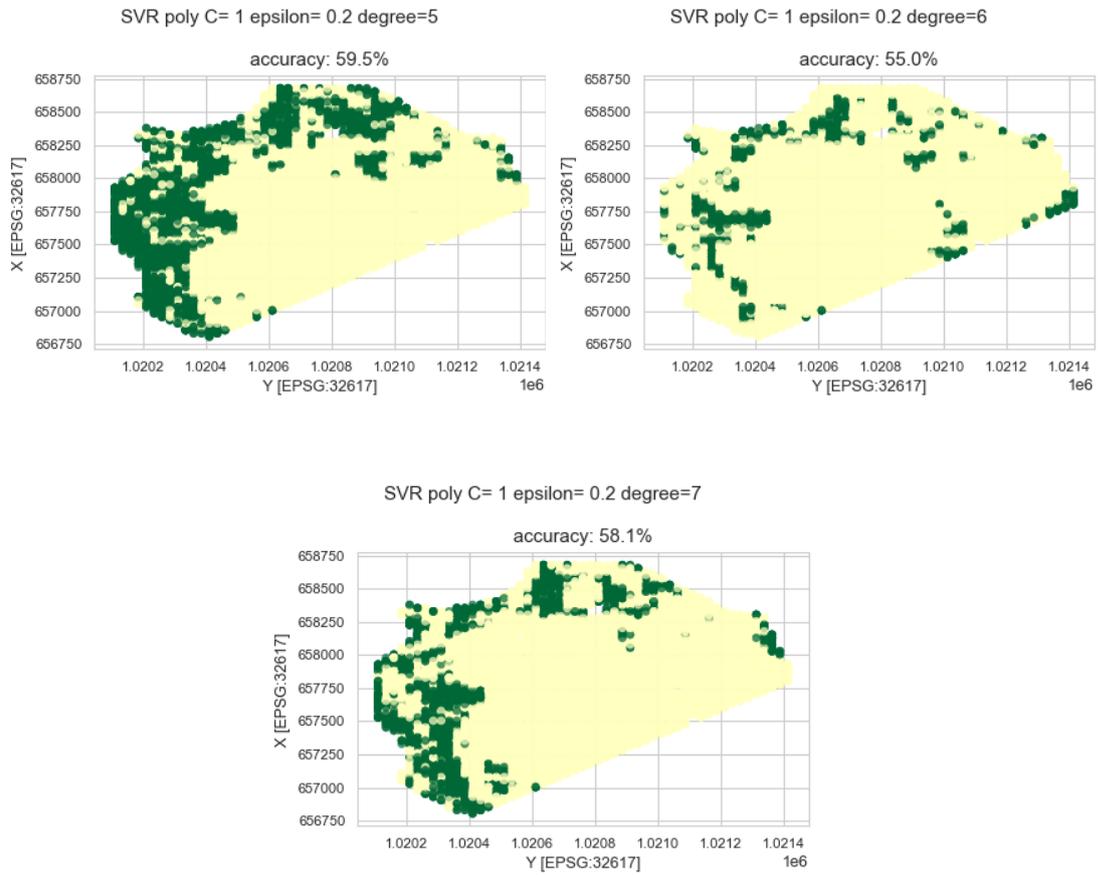
Performing the hyperparameter tests, a large number of tests with different combinations of parameters were performed. The following sections show extended ranges of the tests performed.

D.1 Poly kernel degree tests

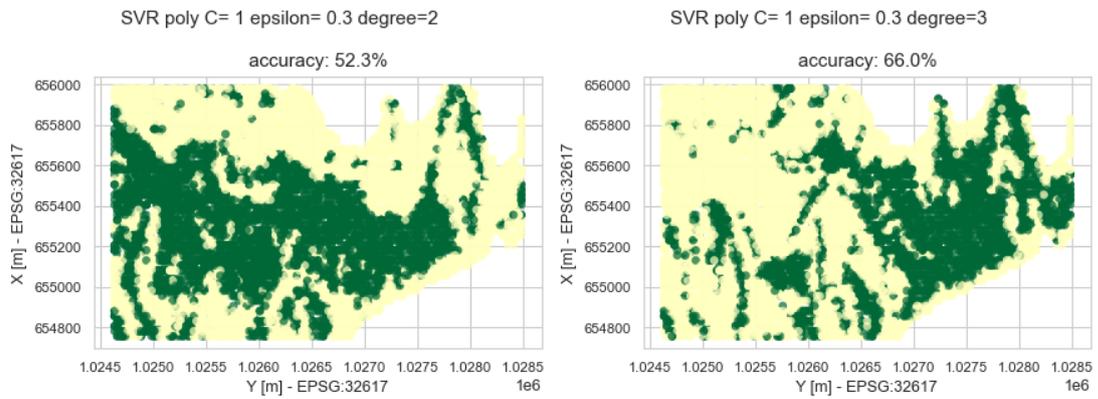
First the poly kernel was tested on a range of degrees. The results for the Río Chagres basin and the Río Pequeni Basin are shown in the figures below.

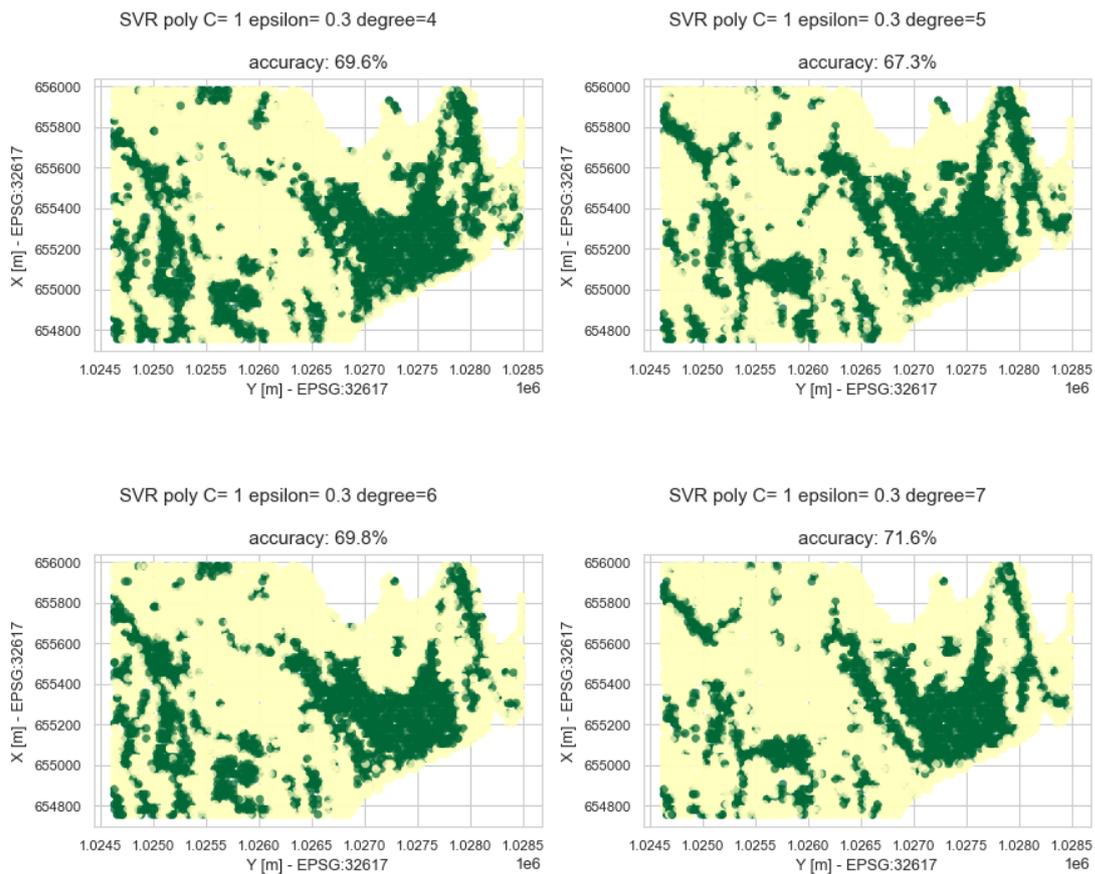
D.1.1 Río Chagres





D.1.2 Río Pequení

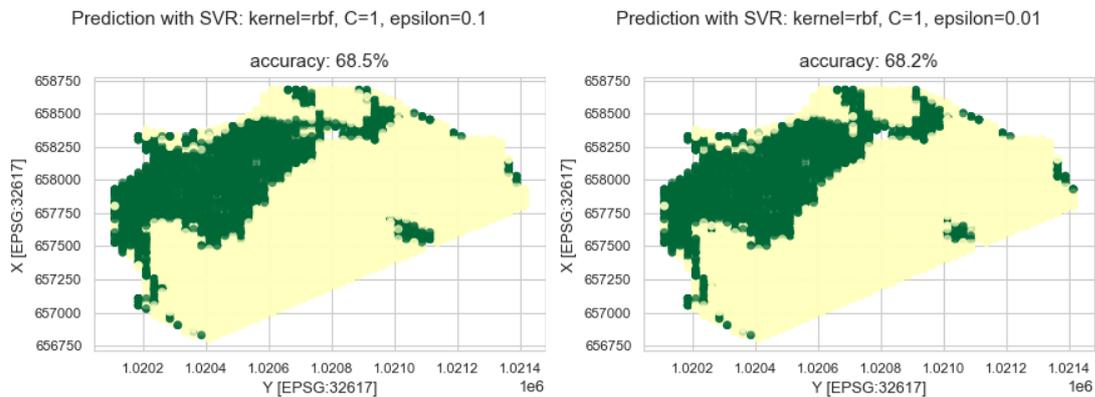




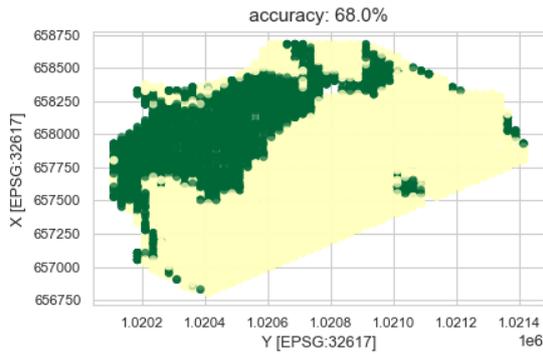
D.2 Rbf kernel C and epsilon tests

To provide coverage for the possible parameter optimization for the rbf kernel, a grid of possible parameters was tested, a range of which is showed in the figures in the following sections. This was done for both the Río Chagres basin and the Río Pequení basin.

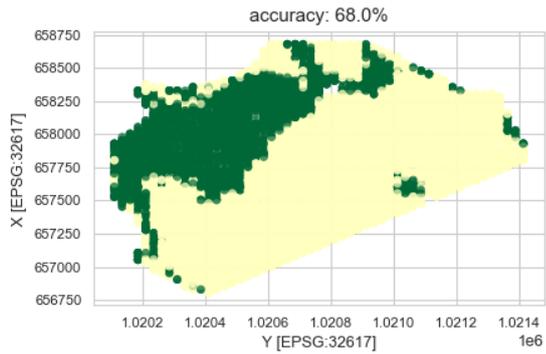
D.2.1 Río Chagres



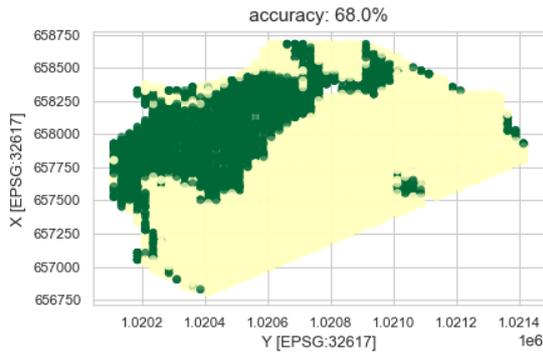
Prediction with SVR: kernel=rbf, C=1, epsilon=0.001



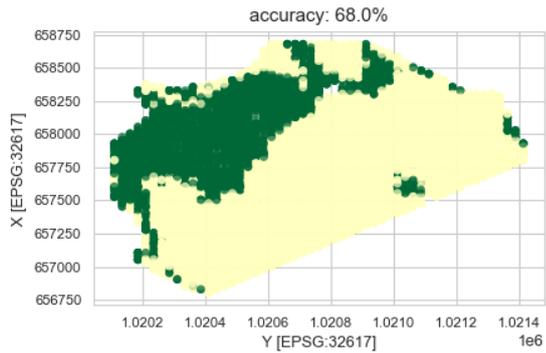
Prediction with SVR: kernel=rbf, C=1, epsilon=0.0001



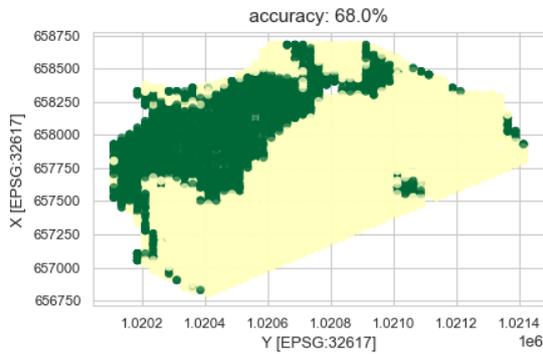
Prediction with SVR: kernel=rbf, C=1, epsilon=1e-05



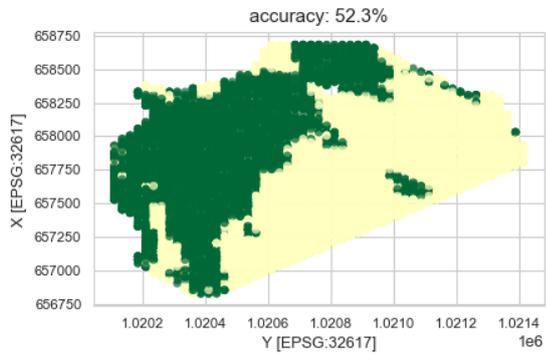
Prediction with SVR: kernel=rbf, C=1, epsilon=1e-06



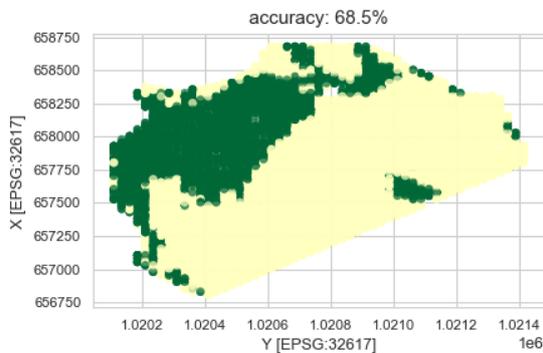
Prediction with SVR: kernel=rbf, C=1, epsilon=1e-07



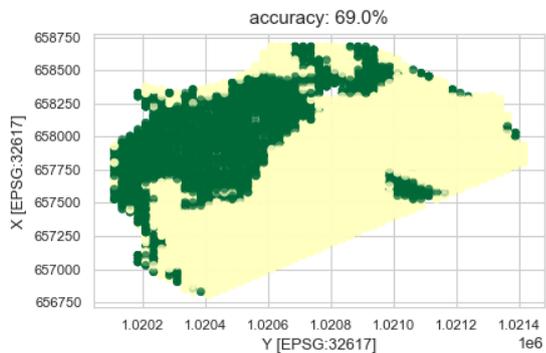
Prediction with SVR: kernel=rbf, C=1, epsilon=1



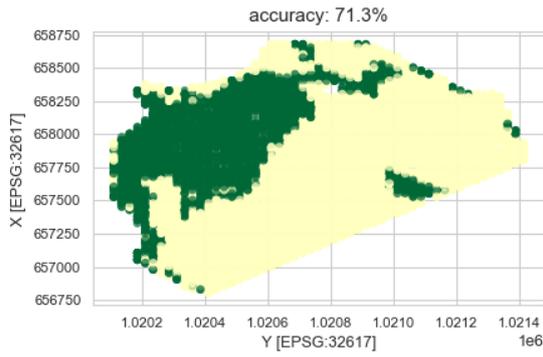
Prediction with SVR: kernel=rbf, C=1, epsilon=0.2



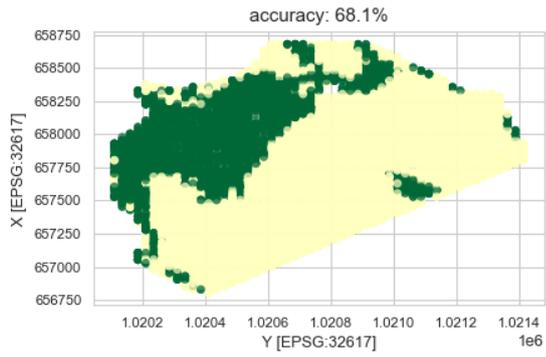
Prediction with SVR: kernel=rbf, C=1, epsilon=0.3



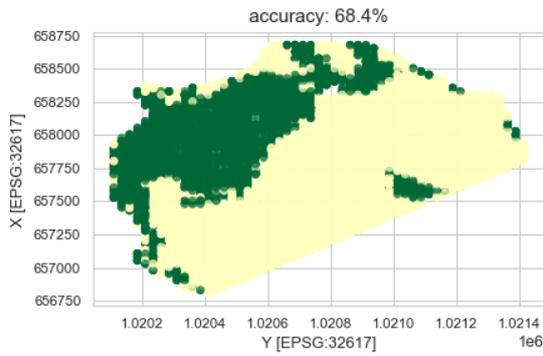
Prediction with SVR: kernel=rbf, C=1, epsilon=0.4



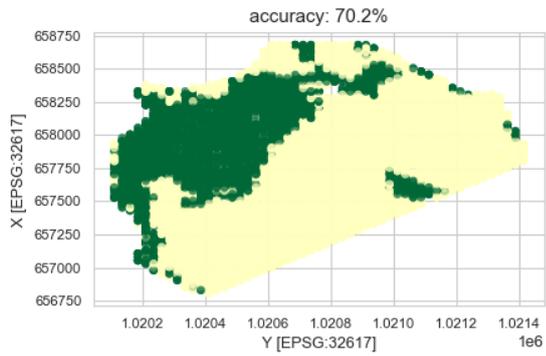
Prediction with SVR: kernel=rbf, C=1, epsilon=0.15



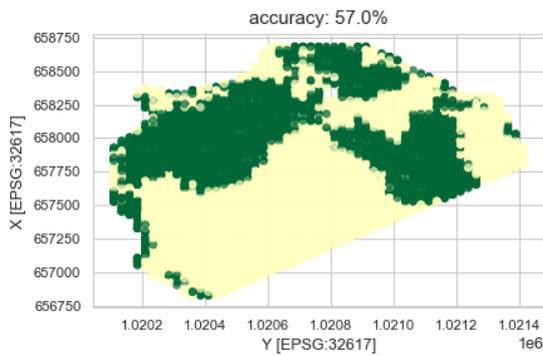
Prediction with SVR: kernel=rbf, C=1, epsilon=0.25



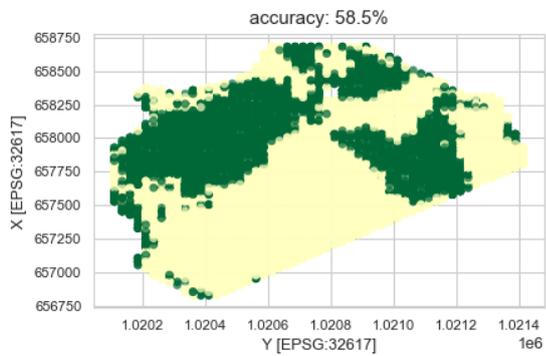
Prediction with SVR: kernel=rbf, C=1, epsilon=0.35



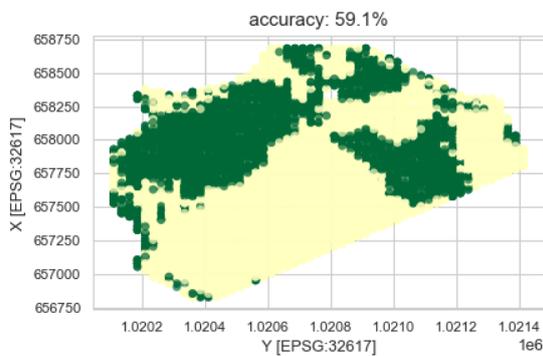
Prediction with SVR: kernel=rbf, C=5, epsilon=0.1



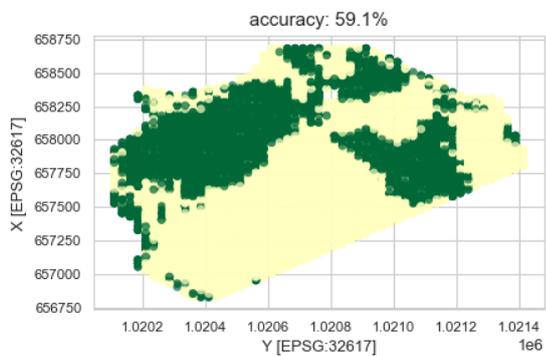
Prediction with SVR: kernel=rbf, C=5, epsilon=0.01



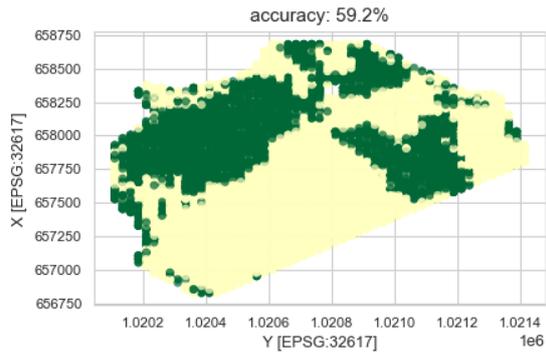
Prediction with SVR: kernel=rbf, C=5, epsilon=0.001



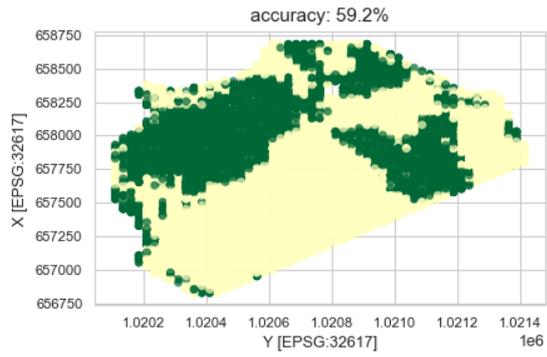
Prediction with SVR: kernel=rbf, C=5, epsilon=0.001



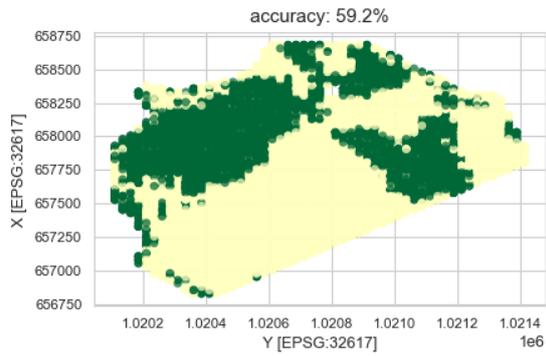
Prediction with SVR: kernel=rbf, C=5, epsilon=1e-05



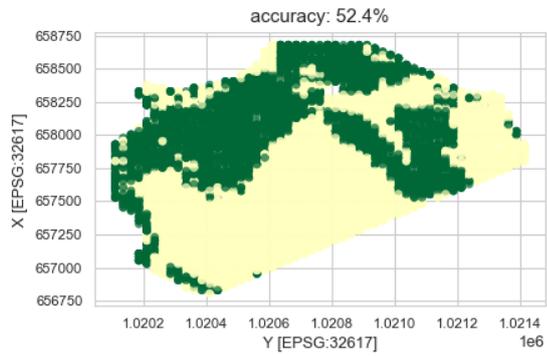
Prediction with SVR: kernel=rbf, C=5, epsilon=0.0001



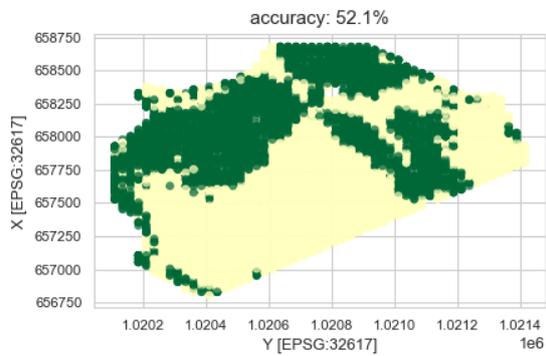
Prediction with SVR: kernel=rbf, C=5, epsilon=1e-06



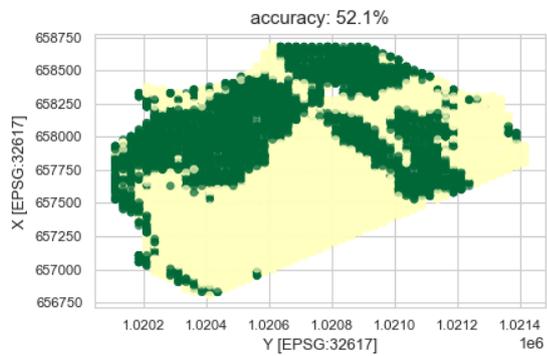
Prediction with SVR: kernel=rbf, C=10, epsilon=0.1



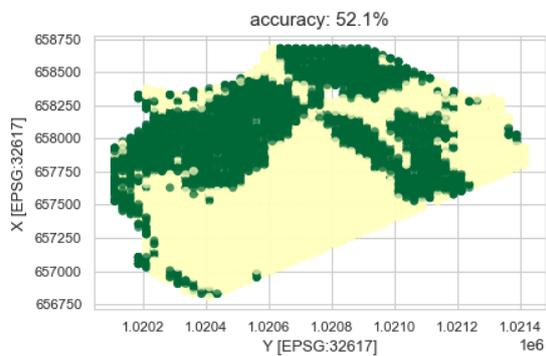
Prediction with SVR: kernel=rbf, C=10, epsilon=0.001



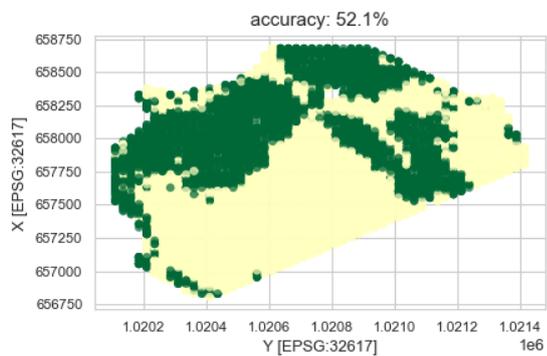
Prediction with SVR: kernel=rbf, C=10, epsilon=0.0001



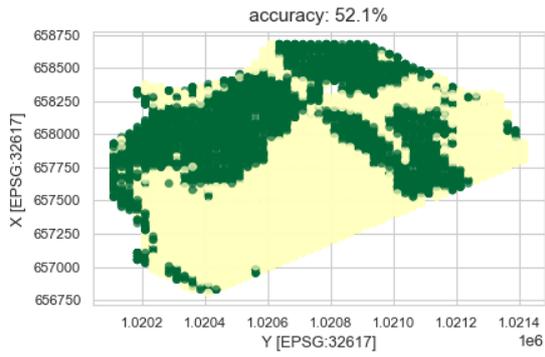
Prediction with SVR: kernel=rbf, C=10, epsilon=1e-05



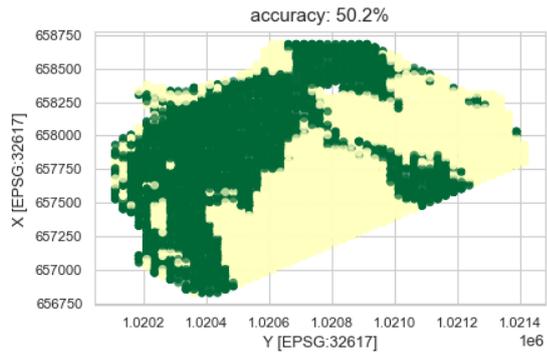
Prediction with SVR: kernel=rbf, C=10, epsilon=1e-06



Prediction with SVR: kernel=rbf, C=10, epsilon=1e-07

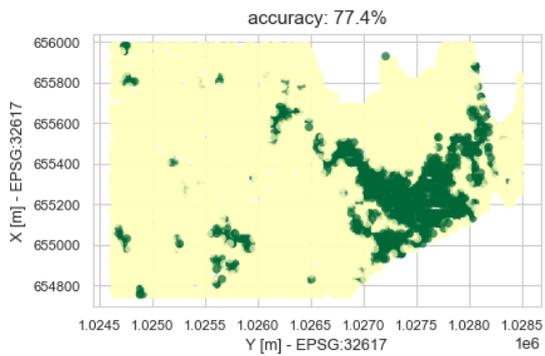


Prediction with SVR: kernel=rbf, C=10, epsilon=1

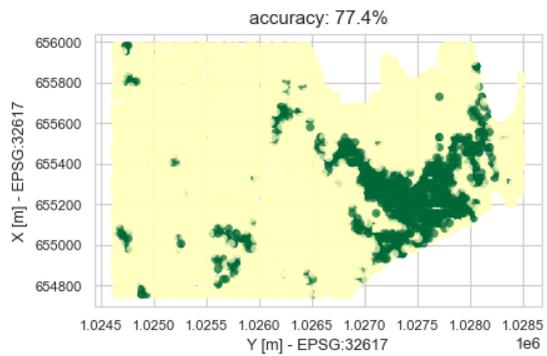


D.2.2 Río Pequení

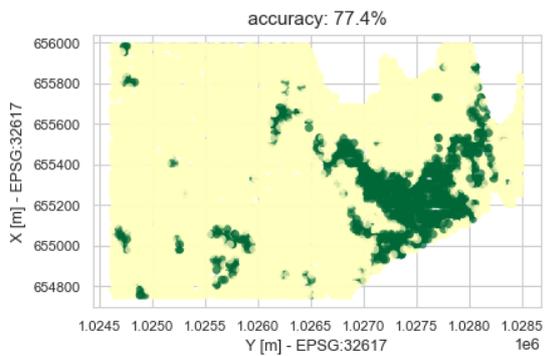
Prediction with SVR: kernel=rbf, C=1, epsilon=0.1



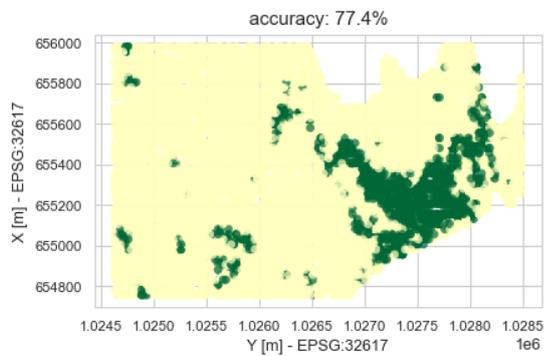
Prediction with SVR: kernel=rbf, C=1, epsilon=0.01



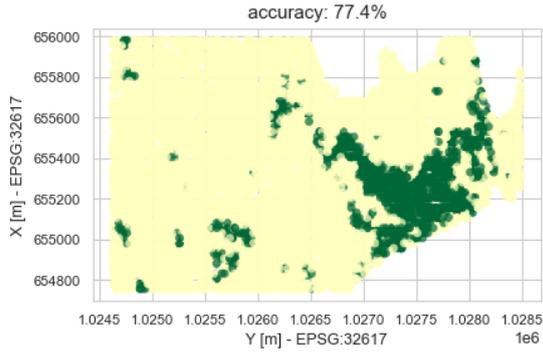
Prediction with SVR: kernel=rbf, C=1, epsilon=0.001



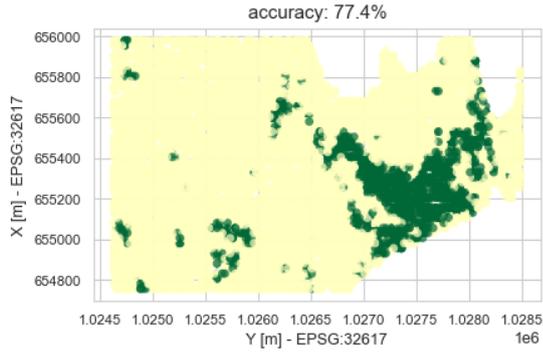
Prediction with SVR: kernel=rbf, C=1, epsilon=0.0001



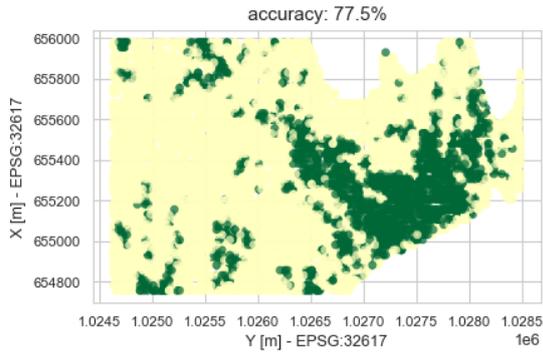
Prediction with SVR: kernel=rbf, C=1, epsilon=1e-05



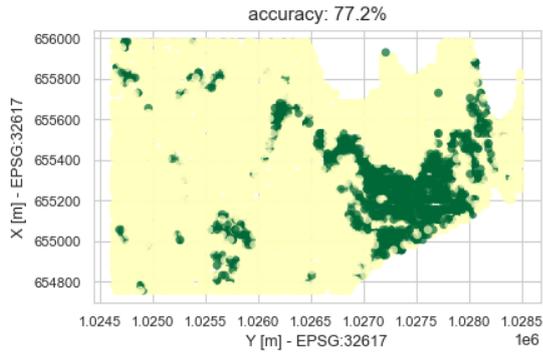
Prediction with SVR: kernel=rbf, C=1, epsilon=1e-06



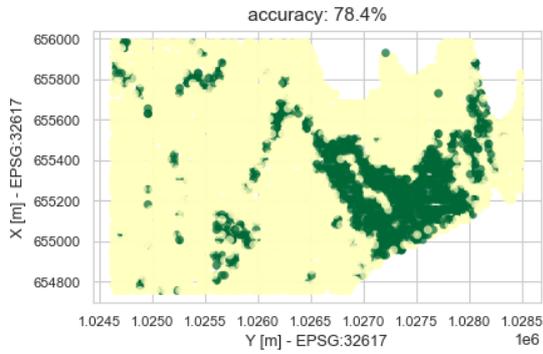
Prediction with SVR: kernel=rbf, C=1, epsilon=1



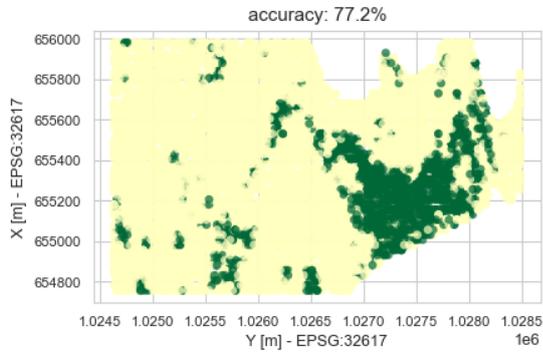
Prediction with SVR: kernel=rbf, C=1, epsilon=0.3



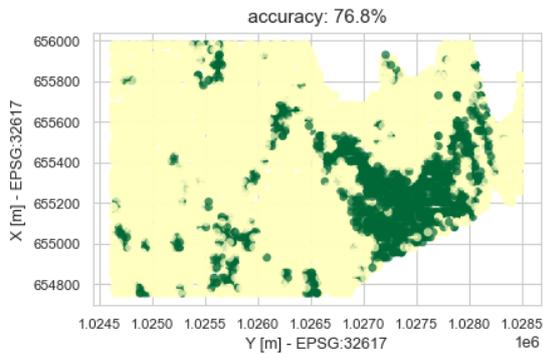
Prediction with SVR: kernel=rbf, C=1, epsilon=0.5



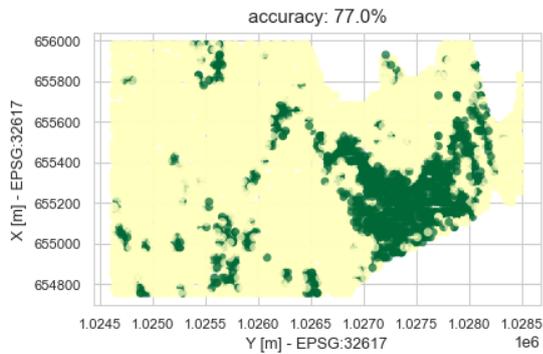
Prediction with SVR: kernel=rbf, C=5, epsilon=0.1



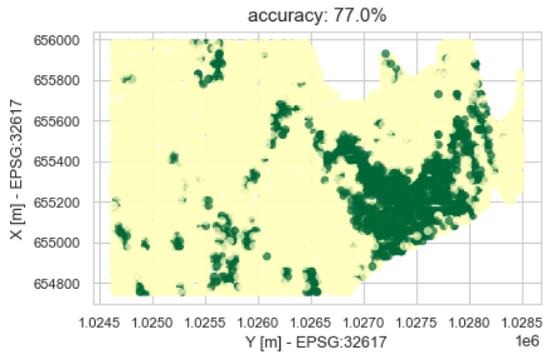
Prediction with SVR: kernel=rbf, C=5, epsilon=0.01



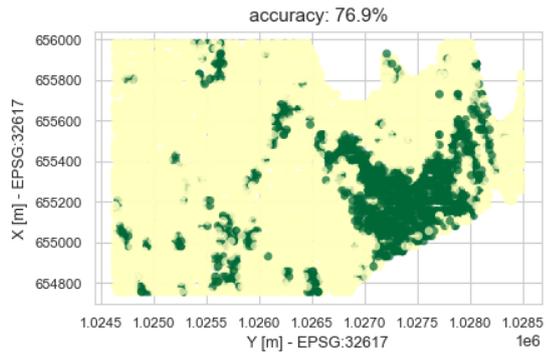
Prediction with SVR: kernel=rbf, C=5, epsilon=0.001



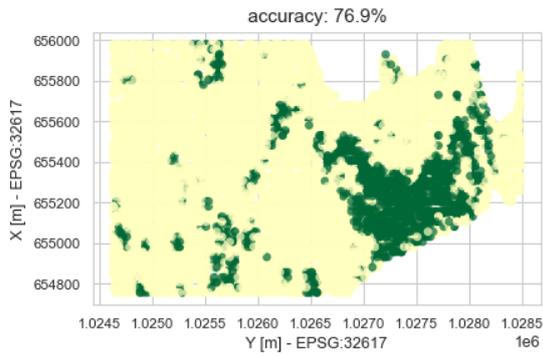
Prediction with SVR: kernel=rbf, C=5, epsilon=0.0001



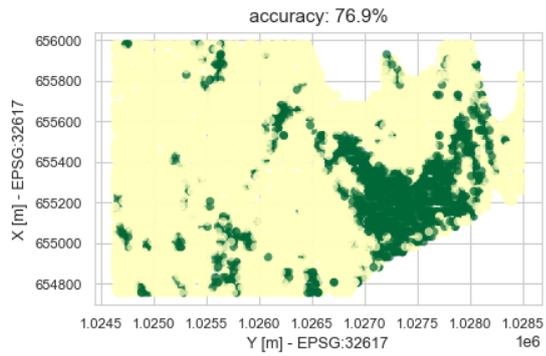
Prediction with SVR: kernel=rbf, C=5, epsilon=1e-05



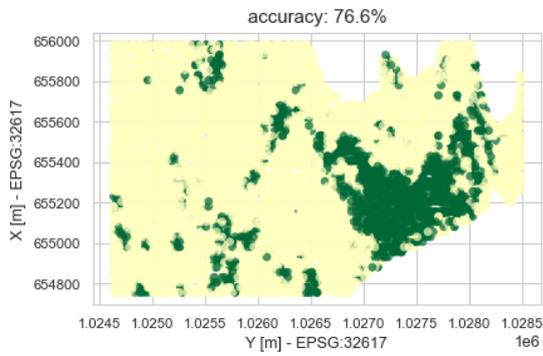
Prediction with SVR: kernel=rbf, C=5, epsilon=1e-06



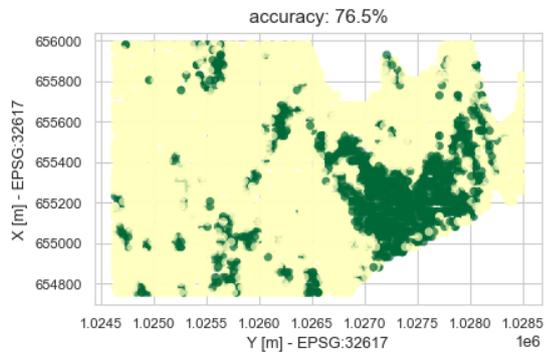
Prediction with SVR: kernel=rbf, C=10, epsilon=0.1



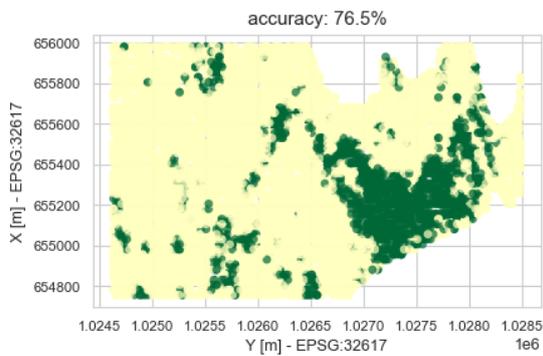
Prediction with SVR: kernel=rbf, C=10, epsilon=0.01



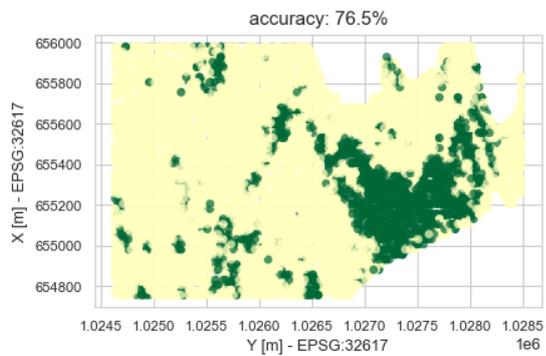
Prediction with SVR: kernel=rbf, C=10, epsilon=0.001



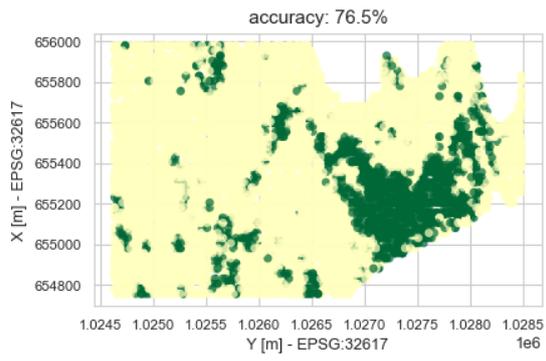
Prediction with SVR: kernel=rbf, C=10, epsilon=0.0001



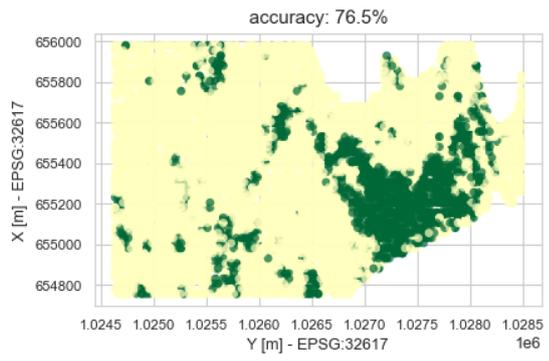
Prediction with SVR: kernel=rbf, C=10, epsilon=1e-05



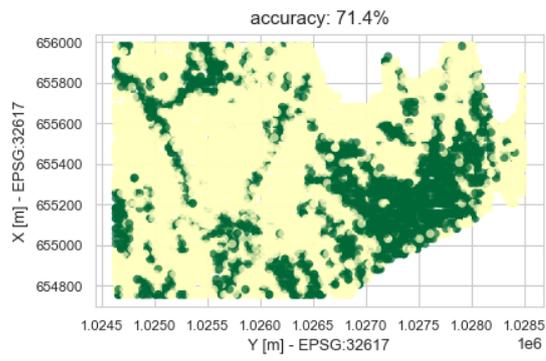
Prediction with SVR: kernel=rbf, C=10, epsilon=1e-06



Prediction with SVR: kernel=rbf, C=10, epsilon=1e-07



Prediction with SVR: kernel=rbf, C=10, epsilon=1



Bibliography

- L. Alvarado. Sedimentation in madden reservoir: Balboa heights, panama: Meteorological and hydrological branch, engineering division, panama canal commission. 1985.
- E. Alzaghoul, M. B. Al-Zoubi, R. Obiedat, and F. Alzaghoul. Applying machine learning to dem raster images. *Technologies*, 9(4):87, 2021.
- M. Asadi, A. Fathzadeh, R. Kerry, Z. Ebrahimi-Khusfi, and R. Taghizadeh-Mehrjardi. Prediction of river suspended sediment load using machine learning models and geomorphometric parameters. *Arab J Geosci*(2021 14: 1926, 2021. doi: <https://doi.org/10.1007/s12517-021-07922-6>.
- M. Avand, A. Kuriqi, M. Khazaei, and O. Ghorbanzadeh. Dem resolution effects on machine learning performance for flood probability mapping. *Journal of Hydro-environment Research*, 40:1–16, 2022.
- M. Awad and R. Khanna. Support vector regression. In *Efficient learning machines*, pages 67–80. Springer, 2015.
- A. Blom. River dynamics 1. <https://brightspace.tudelft.nl/d21/1e/content/399264/viewContent/2476235/View>, 2021. Accessed: 2022–15-05.
- M. Buchhorn, M. Lesiv, N.-E. Tsendbazar, M. Herold, L. Bertels, and B. Smets. Copernicus global land cover layers—collection 2. *Remote Sensing*, 12(6):1044, 2020.
- R. Costache and D. T. Bui. Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the putna river catchment of romania. *Science of The Total Environment*, 691:1098–1118, 2019.
- A. EL Bilali, A. Taleb, B. EL Idrissi, Y. Brouziyne, and N. Mazigh. Comparison of a data-based model and a soil erosion model coupled with multiple linear regression for the prediction of reservoir sedimentation in a semi-arid environment. *Euro-Mediterranean Journal for Environmental Integration* (2020) 5:64, 2020. doi: <https://doi.org/10.1007/s41207-020-00205-8>.
- F. Engelund and E. Hansen. A monograph on sediment transport in alluvial streams. 1967.
- Esri. How curvature works. URL <https://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/00q900000000t000000>.
- F. M. Exner. *Zur physik der dünen*. Hölder, 1920.
- J. H. Graw, W. T. Wood, and B. J. Phrampus. Predicting global marine sediment density using the random forest regressor machine learning algorithm. *Journal of Geophysical Research: Solid Earth*, 126, e2020JB020135 (2021), 2021. doi: <https://doi.org/10.1029/2020JB020135>.

- J. Huang, Y. Sun, and J. Zhang. Reduction of computational error by optimizing svr kernel coefficients to simulate concrete compressive strength through the use of a human learning optimization algorithm. *Engineering with Computers*, pages 1–18, 2021.
- R. J. Hyndman et al. Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, 4(4):43–46, 2006. URL <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.218.7816&rep=rep1&type=pdf>.
- M. B. Idrees, M. Jehanzaib, D. Kim, and T.-W. Kim. Comprehensive evaluation of machine learning models for suspended sediment load inflow prediction in a reservoir. *Stoch Environ Res Risk Assess* 35, 1805–1823 (2021), 2021. doi: <https://doi.org/10.1007/s00477-021-01982-6>.
- B. Jagers. A comparison of predicted methods for medium-term planform changes in braided rivers. pages 713–722, 2001.
- M. Loewenberg. Sedimentation in the panama canal watershed. *Protecting watershed Areas: Case of the Panama Canal*, 1999.
- E. N. Lorenz. The predictability of hydrodynamic flow. Technical report, MASSACHUSETTS INST OF TECH CAMBRIDGE, 1963.
- M. A. Melton. A derivation of strahler’s channel-ordering system. *The Journal of Geology*, 67(3):345–346, 1959.
- A. Mendoza, J. D. Abad, E. J. Langendoen, D. Wang, P. Tassi, and K. E. K. Abderrezzak. Effect of sediment transport boundary conditions on the numerical modeling of bed morphodynamics. *Journal of Hydraulic Engineering*, 143(4):04016099, 2017.
- E. Meyer-Peter and R. Müller. Formulas for bed-load transport. 1948.
- J. T. Minear and G. M. Kondolf. Estimating reservoir sedimentation rates at large spatial and temporal scales: A case study of california. *WATER RESOURCES RESEARCH*, VOL. 45, 2009. doi: <https://doi.org/10.1029/2007WR006703>.
- P. Mitchell, M. Spence, J. Aldridge, A. Kotilainen, and M. Diesing. Sedimentation rates in the baltic sea: A machine learning approach. *Continental Shelf Research*, 214:104325, 2021. ISSN 0278-4343. doi: <https://doi.org/10.1016/j.csr.2020.104325>. URL <https://www.sciencedirect.com/science/article/pii/S0278434320302788>.
- P. Panagos, P. Borrelli, and K. Meusburger. A new european slope length and steepness factor (ls-factor) for modeling soil erosion by water. *Geosciences*, 5(2):117–126, 2015.
- O. Rahmati, N. Tahmasebipour, A. Haghizadeh, H. R. Pourghasemi, and B. Feizizadeh. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. *Geomorphology* 298 (2017) 118–137, 2017. doi: <https://doi.org/10.1016/j.geomorph.2017.09.006>.
- V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71:804–818, 2015. doi: <https://doi.org/10.1016/j.oregeorev.2015.01.001>.
- A. Sahoo, A. Barik, S. Samantaray, and D. K. Ghose. Prediction of sedimentation in a watershed using rnn and svm. In *Communication Software and Networks*, pages 701–708. Springer, 2021.

- P. Samuels. Backwater lengths in rivers. *Proceedings of the Institution of Civil Engineers*, 87(4): 571–582, 1989.
- M. R. Segal. Machine learning benchmarks and random forest regression. 2004. URL <https://escholarship.org/uc/item/35x3v9t4>.
- D. V. Stefanyshyn, Y. V. Khodnevich, and V. M. Korbutiak. Estimating the Chezy roughness coefficient as a characteristic of hydraulic resistance to flow in river channels: a general overview, existing challenges, and ways of their overcoming. *Ecological safety and nature management*, 39(3):16–43, 2021.
- R. D. Williams, J. Brasington, and D. M. Hicks. Numerical modelling of braided river morphodynamics: Review and future challenges. *Geography Compass*, 10(3):102–127, 2016.
- L. Wright. Sediment transport and deposition at river mouths: a synthesis. *Geological Society of America Bulletin*, 88(6):857–868, 1977.

Colophon

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

