Predicting morphological changes in the Alhajuela Lake

Lars Marinus Langhorst
student #4299922
l.m.langhorst@student.tudelft.nl

1st supervisor: Ken Arroyo Ohori
2nd supervisor: Hugo Ledoux

Date P2: January 27 , 2022

1 Introduction

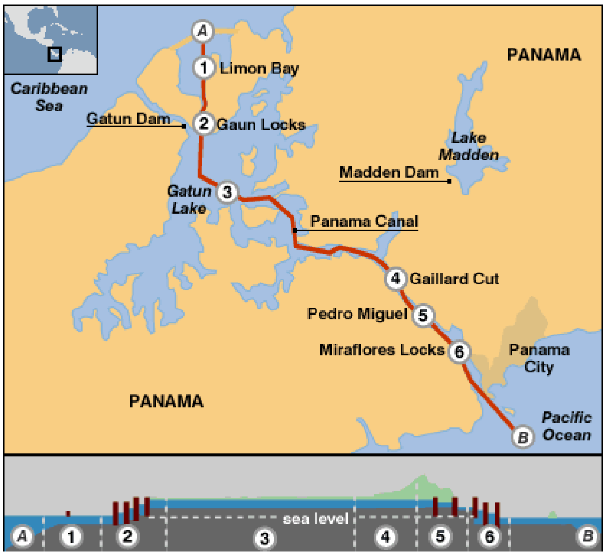


Figure 1: The Panama Canal with the Alhajuela Lake - formerly known as Lake Madden (BBC, 2006)

One of the largest threats to the continuation of operations in the Panama Canal is the shortage of water during the dry season. The canal consists of a large man-made reservoir named the Gatun Lake, and a secondary reservoir; the Alhajuela Lake reservoir which is separated by the Alhajuela Dam (formerly Madden Dam). The canal is dependent on the Alhajuela Lake to maintain the minimum required water level to operate its locks during the dry season. Sedimentation in the reservoir is the result of sediment flux coming from several rivers in the watershed. The location of sediment deposition is dependent on the morphology of the bed which determines the local flow. This results in an increasing loss of storage capacity since the reservoir was built. With an accurate prediction of sedimentation, a time frame for finding a solution to this problem can be determined on long and short term. On short term, dredging operations can slow the loss of volume and these can be better engineered with knowledge of the local sedimentation process. A long term solution may be an alternative reservoir or water source when the Alhajuela Lake no longer suffices as a buffer. Due to the multi dimensional nature of the sedimentation induced morphological response and the large number of parameters involved, current solutions are often 1D or 2D models that use analytical methods based on physical processed, and empirical methods created with results from scale experiments. These methods use flow and sediment flux parameters in combination with river cross sections to compute morphodynamic behavior. These methods require parameter values that are not always available as well as expertise and large amounts of work to apply to each individual situation.

For the Alhajuela reservoir of the Panama Canal, several predictions have been made for the number of years until the reservoir cannot function anymore due to sedimentation Loewenberg (1999). After creation of the Madden Dam in 1935 and with it the Alhajuela reservoir (formerly Madden Lake), it was estimated the reservoir would have a 26% decrease of volume capacity in a period of 3900 years. A study was later done by Alvarado (1985) with the estimated time being 100 years until a 26% decrease of capacity, ending in 2035. This estimate was done at a time in which deforestation was not controlled, and since this research many measures have been taken greatly decreasing deforestation and consequently decreasing the amount of sedimentation. Since the deforestation and sedimentation have decreased, no follow-up research has been done to predict the current rate of sedimentation in the Alhajuela reservoir. By training a Machine Learning (ML) model with geomorphological and hydrological parameters, local sediment levels could be predicted without requiring excessive amounts of time and expertise required. The hydrological parameters include all parameters commonly used in 1D analytical methods such as water flow, sediment flux, and flow speed. Geomorphological parameters are considered all parameters that can be extracted from the digital elevation model (DEM). The use of large data DEMs such as point clouds, height maps, or triangulations, in combination with ML models for the purpose of sedimentation prediction has not been widely incorporated yet since the computational power needed for such operations has only been available in recent years.

1.1 Scientific relevance

In recent years numerous studies have examined the possible benefits of ML models for predicting sedimentation. These models are often trained using geological and hydrological information. ML 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). Features from DEMs (slope, aspect, elevation, etc.) have been used as features in ML models, however no large steps have been made in the training of models with features extracted from these DEMs specifically for analysing sedimentation levels, nor the automation thereof. As information available per reservoir often varies, current researched methodologies are applicable to a single or specific type of reservoir. I foresee the possibility of an automated feature extraction and sediment prediction pipeline only requiring historical DEMs and hydrological input data. By combining ML and DEMs, predictions could then be made without any (or very little) necessary calibration and expertise.
The immense number of reservoirs around the world are all affected by sedimentation, and the Panama Canal with the connected Alhajuela Lake is no exception. This is where sedimentation could ultimately have very drastic impact on the entire world shipping industry, making insight into this process is crucial. This thesis will research a pipeline to predict sedimentation in the Alhajuela Lake using a ML model, trained with geomorphological parameters and hydrological data.

2 Related work

2.1 Sedimentation

In all watersheds, the rivers and creeks are supplied with eroded soil which gets transported through the flowing water. In places of reduced water flow, this sediment then sinks to the river bed, causing it 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 in a reservoir such as flow, relative pool height, sediment supply from upstream, and sediment size and distribution along with other geological factors, the geomorphic delivery processes, 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.

In present day, with powerful 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 inflow of sediment into the reservoir, 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 of the inflow as well as not taking reservoir operations into account (Idrees et al., 2021). These methods make use of parameters such as:

- Channel width [m]
- Flow velocity [m/s]
- Chezy coefficiënt [$m^{1/2}/s$]
- Porosity of reservoir bed
- Median particle diameter of suspended sediment [m]
- Sediment flux [m^3/s]

These parameters are estimations and 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 these methods of acquirement largely affects the accuracy of tasks depending on these results (Stefanyshyn et al., 2021). Although these analytical models are interpretable as they are based on physical processes, the models cannot always be applied due to lack of data or knowledge.

2.2 ML and ANN

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 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 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 RF with sedimentation data, however the predictions made are on a spatial scale. In this thesis the aim is to predict sedimentation on a temporal scale, making the prediction dependent on different parameters and processes.

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 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.

2.3 Predictions with geomorphological features

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.
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 processes (GPs) and evolutionary support vector machines (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 Figure 2.

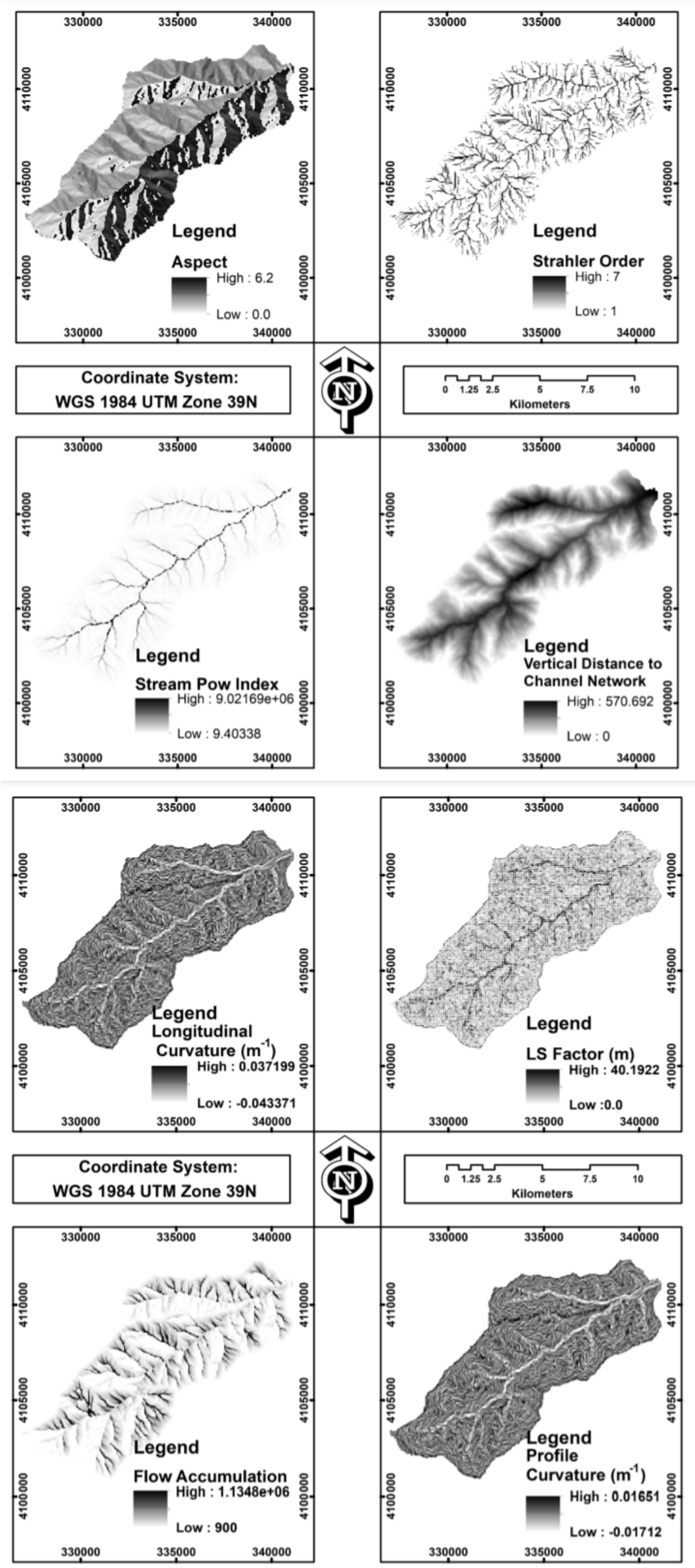


Figure 2: Maps of parameters extracted from DEM (Asadi et al., 2021)

As seen in Figure 2, 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 the actual relation of these features to the morphological process studied by Asadi et al. (2021) is entirely different from the process that is affecting the underwater bed morphology.
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 Figure 3. Random noise grids are added to establish a maximum median error for the predictor grids.

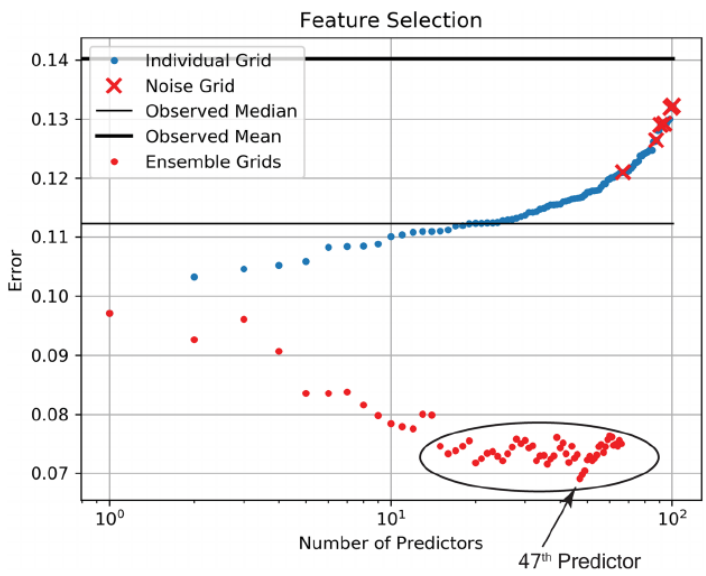


Figure 3: Predictor grid median prediction error (Graw et al., 2021)

The smallest error in Figure 3 was obtained using 47 predictor grids, while there was a total of 66 grids with less error than the grids containing random noise. 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).
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, both analytical and empirical, 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 uses hydrological

and geological parameters such as soil types, suspended sediment values, and flow velocities. 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. To fill this gap, this thesis will focus on developing a methodology to predict local morphological changes using a ML algorithm trained with hydrological and geomorphological parameters.
As previous estimations of sedimentation levels in the Alhajuela reservoir are very global and done without the available DEMs and computational possibilities presently available, this thesis will append to this by providing new estimations of the decrease in reservoir volume, as well as predictions of local morphological changes.
Features will be extracted from the available DEM and tested with a number of different ML models. These new features will be selected according to their feasible correlation with the morphological changes in the terrain. The methodology that will be constructed with this research allows for engineers to obtain a prediction of sedimentation in reservoirs using DEMs and hydrological data, without requiring specific knowledge on the complex hydrological and morphological processes.

3 Research questions

In this thesis, the main research question is:

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

The main objective of this research is the creation of a pipeline, providing predicted sedimentation levels of the Alhajuela Lake up to 10 years after last bathymetric measurements, which is also the planned period of the next measurements. This requires results from a ML algorithm, using features extracted from DEMs and hydrological data about 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, and give an indication of the operational years left for the reservoir in case no dredging takes place. 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 Alhajuela reservoir?*

3.1 Scope

This research will focus on the application of the ML algorithms for prediction of sedimentation levels in the Alhajuela Lake. The data that will be used for the prediction are point clouds of the reservoir's bathymetry from within the last 25 years. This data consists of the original single beam sonar measurements as well as interpolated points of the lake bed. Additionally the precipitation in the watershed and hydrological data of the dam from the past 20 years will be used.
The ML algorithm will be used to replace the methodologies ordinarily taken using solely hydrological data, and use the ML model in combination with the DEMs of the lake to predict the global and local sedimentation. The final pipeline will be designed for the Alhajuela Reservoir specifically, and will not be an out-of-the-box solution for all reservoirs. The method will be

adaptable for similar reservoirs, if sufficient historical data is available. This thesis will not go in to the details of, or analyse, the hydrological aspects of the result.
The final predictions will consist of the predicted future DEM of the reservoir with the local sedimentation hot-spots, a global prediction of the total volume loss, as well as a prediction for the number of years left until the reservoir has lost 26% of its original capacity in order to compare the result with the prediction made by Alvarado (1985).

4 Methodology

In this section, the steps necessary to obtain a prediction of sedimentation using a ML model is described. A pipeline to give an overview of these steps is shown in Figure 4. The main components of the pipeline are the pre-processing of the input data (DEMs and hydrological data), extracting features from the data, selection of the provided features, followed by the training and validating of the ML model. The accuracy of the prediction made with this model is then analysed, and a new set of features can be chosen to improve the prediction, or the future prediction can be made when the current model accuracy is sufficient.

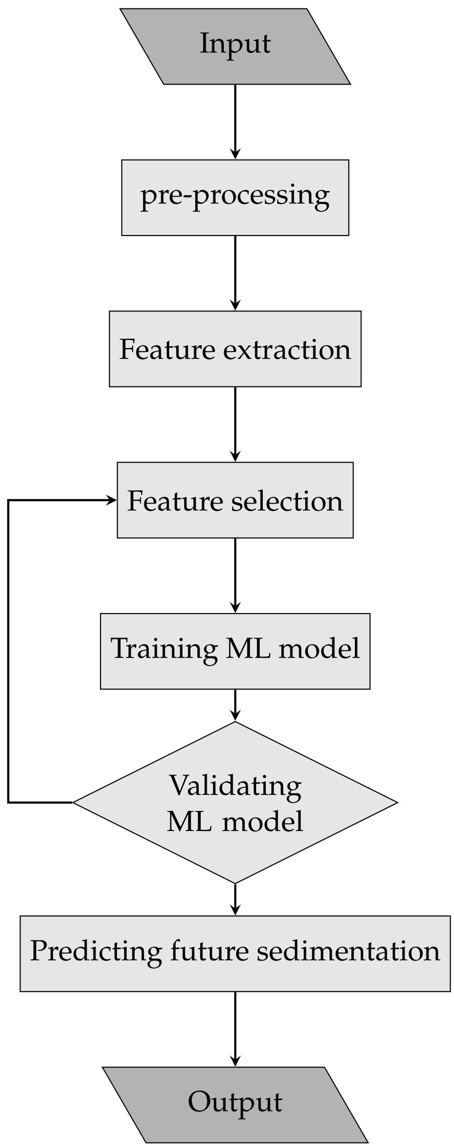


Figure 4: Research Pipeline

4.1 Feature extraction

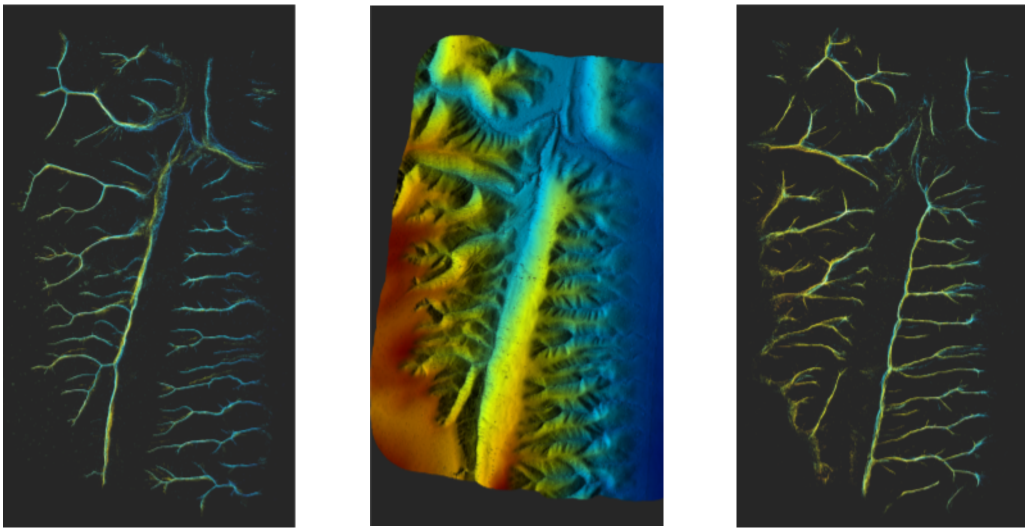
The feature extraction step in the pipeline is an important part of the process in the earlier phases of the research, however throughout the research new features can always be added, tested and removed when necessary. Using FME, features can be extracted from point clouds and triangulations thereof. By overlaying a grid of points on several computed triangulations in FME, layers of features are collected which can then be appended to a .csv file. This way, the height values of the points are interpolated using the TIN through FME which can be adapted to compute for a specified number of vertices. In the preliminary prediction, simple features were computed and extracted such as:

- X coordinate
- Y coordinate
- Slope [Deg]
- Aspect [Deg]
- Height relative to lowest point [m]
- Depth compared to average water level [m]
- Euclidean distance to a specified point (e.g a river mouth) [m]
- Year
- Time elapsed since last measurement [years]

To add to these features and improve the prediction, more features will be computed using grids, triangulation, but also the Medial Axis Transform (MAT). A number of these additional features can be extracted using FME in the same manner as was already done, but different scales. Slope and aspect are computated for triangulations with different levels of detail, meaning different total numbers of vertices.

A pipeline to extract additional features will be created in C#. Here, all vertices can be stored with appended features and extracted into bins that collectively make up layers of grids as mentioned in Section 2. In C# the slope, aspect, height, and other features of a point can be compared to that of the surrounding points providing the relative value of this feature to its surroundings on a specific scale. This is important since a small change in slope in a terrain has very different implications than change of slope on much larger scale. The same scaling principle applies for several other features such as relative height and aspect.

The MAT will be extracted for the terrain per data set. The Medial Axis Transform (MAT) was first introduced by the biologist Blum et al. (1967), and Peters (2018) explains how to apply this transform in a robust way to geographic point clouds Transforming the model to a MAT, we move the axis to all valleys which are naturally close to center of the lake bed and rivers. An example of this transform on a terrain is shown in figure 5.



(a) Exterior MAT (b) Terrain point cloud (c) Interior MAT

Figure 5: MAT on terrain (Peters, 2018)

This allows for the coordinate system to be centered along the most important locations of the terrain(valleys and deepest points in the lake), whereas in a longitude-latitude grid only Euclidean distances can be computed. The longitudinal and transverse axis of the MAT relative to the flow of a river is likely to have correlation with sedimentation values as was found by Xuejun et al. (2011). To connect the points in the grid to the MAT, the MAT will be created in 3D using the Shrinking Ball Algorithm introduced by Ma et al. (2012). This is then projected to 2D, and every grid point will be corresponding to its closest point on the medial branches. Some features that will be extracted from the using are:

- Distance closest branch
- ID of closest branch
- length and magnitude of closest branch
- Width to Depth ratio along cross section
- Curvature of medial branch
- Spoke vector length compared to up- and downstream directions of medial branch

More possible features to be extracted from the MAT will be researched by studying the MAT properties along with the effecting parameters in the sedimentation and landscape morphology processes.

4.2 Feature selection

The goal of the feature selection step is to filter the provided features, and create a set of features that will train the model with the highest possible prediction accuracy. Some features have stronger correlation than others, and too many features may impact the prediction in a negative way, specifically when using the RF (Mitchell et al., 2021; Graw et al., 2021) Different combinations and numbers of features will be used to train and validate the models, searching for the set that will result in the highest prediction accuracy. One of the methods to search for the best set of features is the method used by Graw et al. (2021) as is described in Section 2.

Features are additionally analysed for selecttion based on their importance values. From a RFR model, the Gini importance, or Mean Decrease in Impurity (MDI) can be computed per feature, where the times a feature is used to split a node is counted and then weighted by the number of samples it splits (Perrier, 2015). The total sum of importance values of the features equates to zero, and features with higher values have higher importance.
Every iteration, a prediction of sedimentation in 2018 is made using the model trained with data from 1997 to 2012. The result is then analysed, compared with the actual sedimentation of the predicted period, and global and local accuracy are assessed as done in Section 5.

4.2.1 Training and validating the ML model

Several ML methods will be trained and tested for accuracy in this research. The main candidate for the prediction is the the Random Forest Regression (RFR) model due to the fact that it can handle many potentially correlated predictor variables, it can contain a large amount of trees without overfitting (Hastie et al., 2009), and it does not require extensive tuning (Yates et al., 2018). The out-of-bag cross-validation prediction error and relative estimates of variable importance which can be extracted from the RF model give the advantage of direct insight into the parameters used for each prediction (Mitchell et al., 2021). The results from the first RFR predictions with the data set are shown in Section 5.
Additionally, binning the data on grids to improve predictions as was done by Graw et al. (2021) will be trialed in order to attempt to improve accuracy. Since the area of study is of a smaller scale, the binning process must be adapted to work on this scale and can additionally be adapted with MAT coordinates.
The accuracy of the prediction is then tested and visualised as done in Section 5, to check possible areas for improvement. The MASE accuracy metric as defined by Hyndman and Koehler (2006) shown in equation 1 will be applied to assess prediction accuracy as it was found to be most effective for different types of predictions (Hyndman et al., 2006).

$$MASE = mean\left(\frac{|e_j|}{\frac{1}{T-1}\sum_{t=2}^{T}|Y_t - Y_{t-1}|}\right) = \frac{\frac{1}{J}\sum_{t=2}|e_j|}{\frac{1}{T-1}\sum_{t=2}^{T}|Y_t - Y_{t-1}|} \tag{1}$$

This is repeated iteratively as explained in section 4.2.

5 Preliminary results

To compute the preliminary results, bathymetric models in the form of point clouds of the Alhajuela Lake are available of the years 1997, 2008, 2012 and 2018 were used which were originally made using single beam sonar equipment. The datasets were alligned and clipped to create a grid where, only the cells are kept which contain data from every year. To account for the hydrological difference in rainfall and natural events between the different periods, an arbitrary value 'Xfactor' is added. To obtain the first results, the data was filtered and aligned and a number of features were extracted from the DEMs. These features were then added to grid cells creating a number of arrays with features for the ML pipeline. The sklearn library was used to split the data, train and test the model, and make predictions with the RFR algorithm.

5.1 Feature importances

In Figure 6 the features used to get the preliminary predictions are shown in order of their computed importance. For every grid cell the slope, aspect, last measured height (horiginal), depth, x and y coordinates, the distance to a number of river mouths (dRM1 - dRM7) and the combined value of a points distance to all river mouths (dRMtot) is computed. Additionally, the time since the last measured height (T) is known, and the year this was measured. As mentioned the 'Xfactor' is added to take into account differences in rainfall and natural events that may trigger high amounts of erosion or sedimentation. In 2010, such an event took place during the rainy season where extreme rainfall and mudslides resulted in large amounts of morphological changes in the reservoir bed. The 2008 measurements were thus deemed inaccurate, and an extra survey of the reservoir to be done 4 years later in 2012 instead of the standard 10 years.

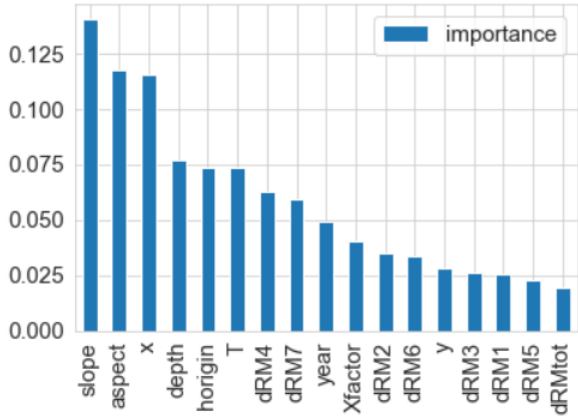


Figure 6: Feature importance

5.2 Prediction residual

In Figure 7, a boxplot is shown for the average annual bed level changes over the period of 2012 to 2018. The prediction of the annual sedimentation in the period of 2012 to 2018 is done with a RFR model that is trained with data from 1997 to 2012. As can be seen in Figure 7, the model is generally underestimating the amount of sedimentation resulting in the residual values (predicted value minus the actual value) to be slightly negative.

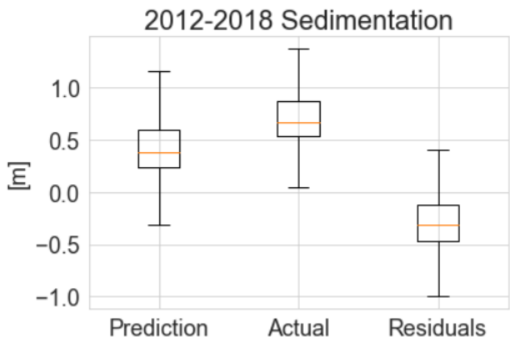


Figure 7: Sedimentation prediction results 2012-2018

In Figure 8 the prediction alongside the actual sedimentation value and the residuals for this same period is visualized along longitudal and latitudal axes. There are areas that are correctly predicted by the model, e.g. in the northern part of the lake where the Rio San Miguel and Rio Pequeni feed into the reservoir, with the Salamanca catchment area experiencing a lot of erosion(red branch in the northern part of the lake).

In the south east side of the reservoir, the Rio Chagres brings a constant supply of sediment which is clearly visibly in the first two plots of Figure 8 as darker blue patches are present at the mouth of the river. These dark patches are sediment deposition location and are dynamic in the sense that the location of deposition not remain the same. This is demonstrated as the model is trained on the data until 2012, whereas in the years leading up to 2018 the sedimentation has been taking place in a slightly more southern part of the river mouth. This is also visible in the right plot in Figure 8 showing the residual as a red patch near the mouth of the Rio Chagres.

From these preliminary results, several regions are seen to have been predicted more accurately than others. Overall, the model is predicting the annual build up of sedimentation to be slighly less than the actual values, which could be correlated to external causes such as agriculture or heavy rain seasons. These external factors are thus represented with the Xfactor having a higher value in periods of high rainfall, and this variable will be replaced or supplemented by hydrological data of the dam and rivers along with and the watershed rainfall statistics. The slope being a feature of high importance is a promising result, as this means that features to be extracted using the MAT axis may also be successful since the slope is closely correlated with the distance of a point from the center of a river or river mouth.

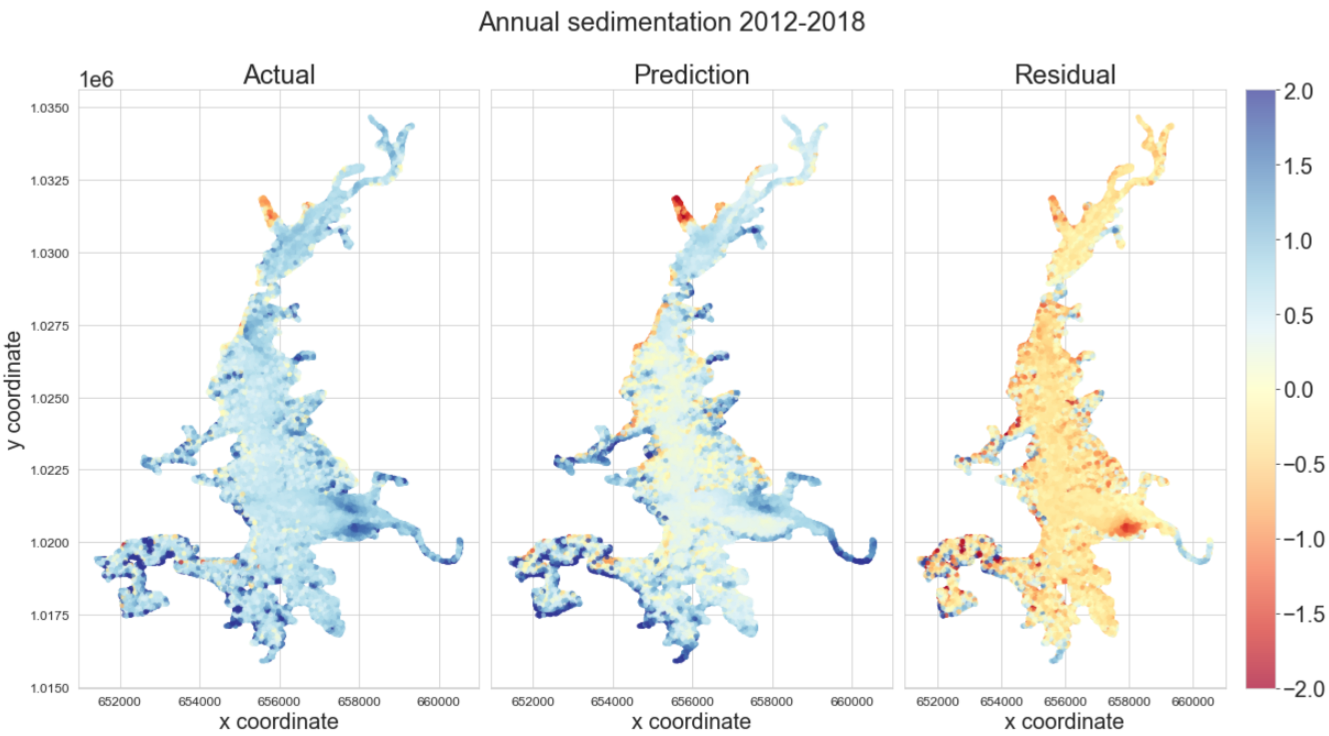


Figure 8: Annual sedimentation

5.3 Prediction of local sedimentation

To predict the annual sedimentation levels in the period of 2018 to 2022, the model was trained and tested with data from the years 1997, 2008, 2012, and 2018. The input features from the 2018 DEM were given to the RFR model, with the Xfactor set to the 'normal' value, which produced the prediction in Figure 9.

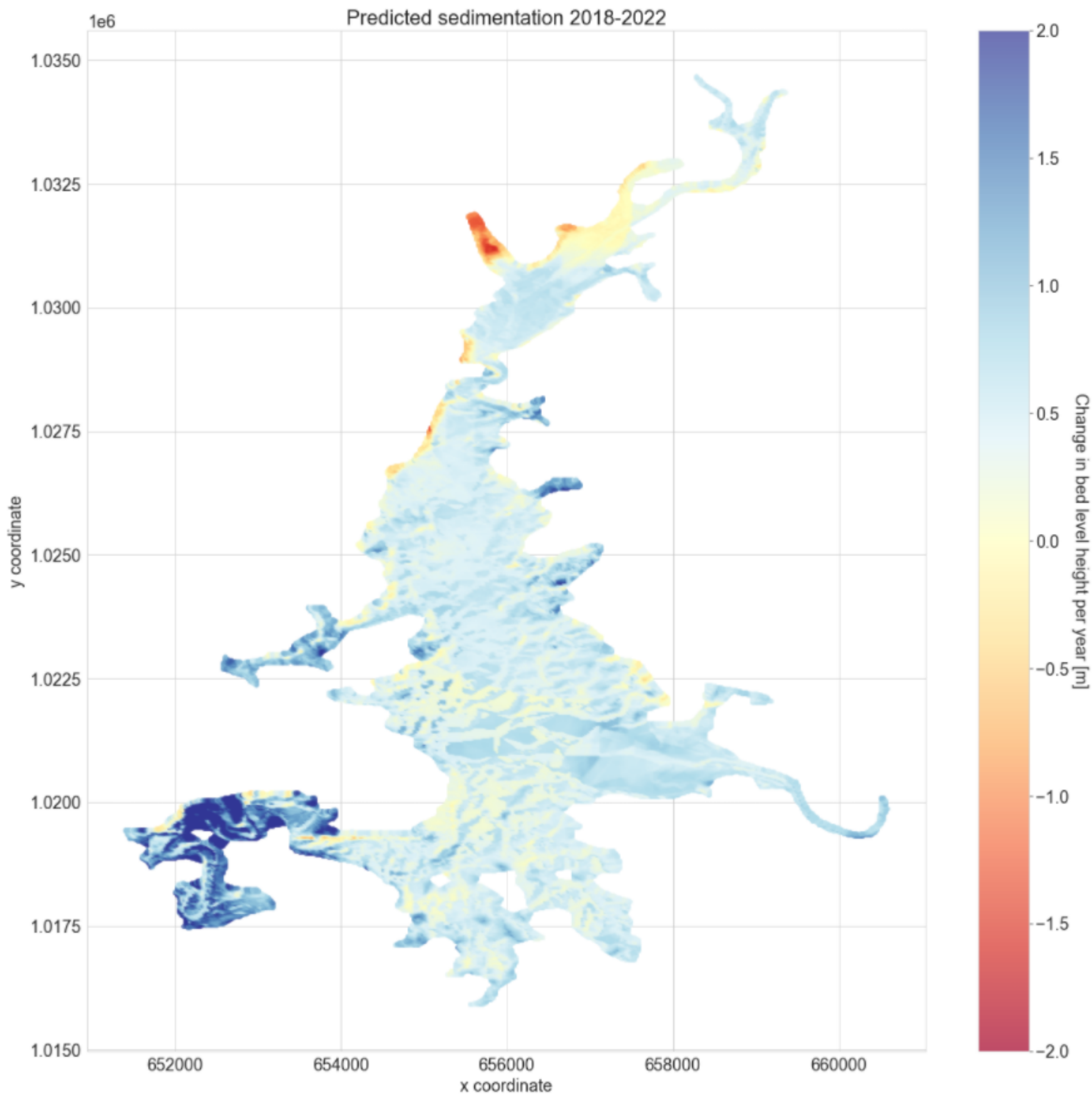


Figure 9: Annual sedimentation prediction 2018-2022

6 Time planning

To show the planned schedule for this thesis, the Gantt chart in Figure 10 shows the planned soft and hard deadlines for the individual parts in the entire duration of the research.

6.1 Meetings

A weekly half-hour meeting is held with the primary supervisor. Every other week the secondary supervisor will join this meeting to provide additional support and feedback.

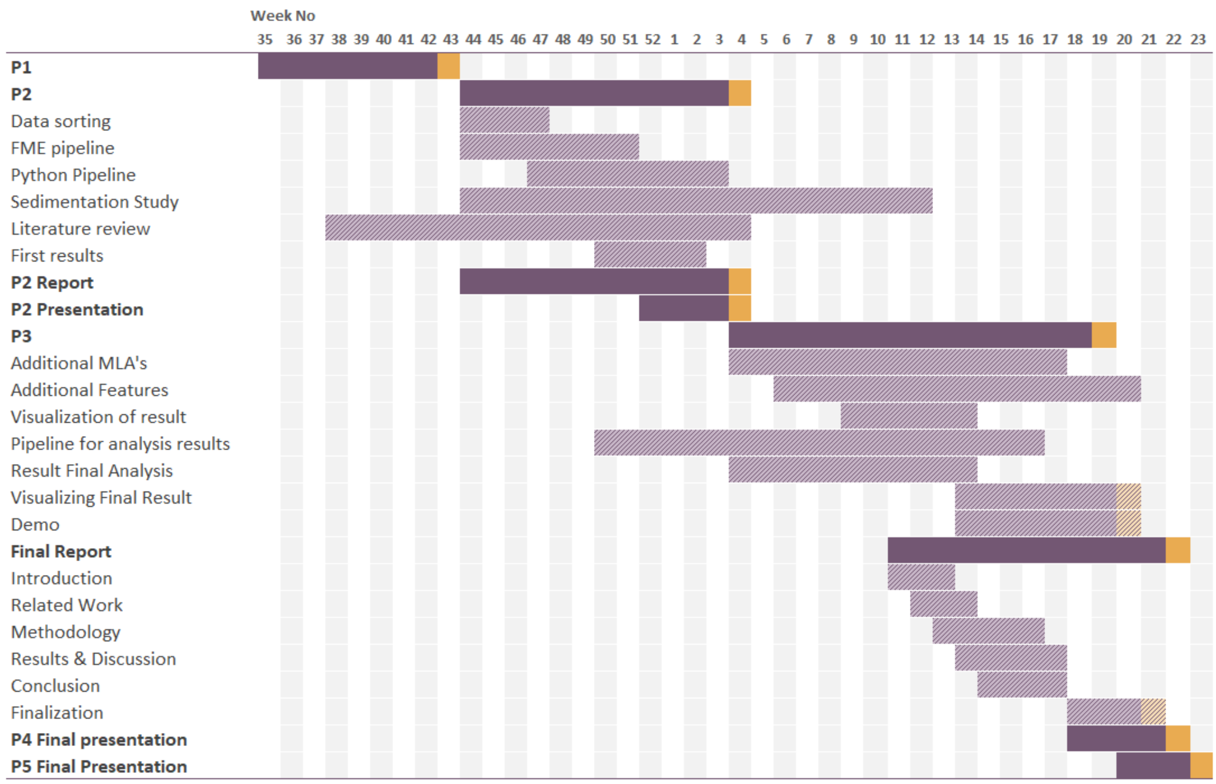


Figure 10: Gantt chart

6.2 Tools

During this research, the following tools are used:

- FME software by Safe– The raw point cloud data is not yet aligned. Thinning, filtering, and interpolating is done in FME from which a .csv file is then exported.

- C# - Unity software – C# is used for extraction of features that are not easily obtained with FME. The Medial Axis Transform is done in C# along with the extraction of features related to it. Results can be inspected immediately by means of the Unity engine.

- Microsoft Excel – The extracted data is organised in Excel before importing to Python

- Python – Used to process data, perform predictions, and visualize and analyse results.

- Scikit-learn ML library by Pedregosa et al. (2011)– ML library for Python.

6.3 Data

The data used is a collection of bathymetric measurements in the form of point clouds. Datasets of the last 25 years will be used to train the model for making predictions into the future. These datasets are pointclouds of the bathymetric models created with single beam echo sounding techniques, as well all point clouds created from a triangulations of the original pointclouds. All data is provided by the Panama Canal Department of Cartography.

Additionally, hydrological data of available of the Madden Dam along with rainfall statistics from the past 20 year is provided by the Hydrological Department of the Panama Canal.

References

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.

BBC. U.s. energy information administration, 2006. URL `http://news.bbc.co.uk/1/hi/world/americas/6074400.stm`. [Online; accessed January 10, 2022].

H. Blum et al. *A transformation for extracting new descriptors of shape*, volume 43. MIT press Cambridge, MA, 1967.

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.

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.

T. Hastie, R. Tibshirani, and J. Friedman. Random forests. In *The elements of statistical learning*, pages 587–604. Springer, 2009.

R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2006.03.001. URL `https://www.sciencedirect.com/science/article/pii/S0169207006000239`.

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.

M. Loewenberg. Sedimentation in the panama canal watershed. *Protecting watershed Areas: Case of the Panama Canal*, 1999.

J. Ma, S. Bae, and S. Choi. 3d medial axis point approximation using nearest neighbors and the normal field. *The Visual Computer volume 28, pages7–19, 2012*, 2012. doi: https://doi.org/10.1007/s00371-011-0594-7.

M. A. Melton. A derivation of strahler's channel-ordering system. *The Journal of Geology*, 67 (3):345–346, 1959.

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 spatialand 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.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

A. Perrier. Feature importance in random forests, Aug 2015. URL `https://alexisperrier.com/datascience/2015/08/27/feature-importance-random-forests-gini-accuracy.html`.

R. Peters. Geographical point cloud modelling with the 3d medial axis transform. 2018.

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.

Safe. Fme - data integration platform, safe software inc., 2021. URL `https://www.safe.com/fme/`.

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.

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. stimating the chézy roughness coefficient as a characteristic of hydraulic resistance to flow in river channels: a general overview, existing challenges, and ways of their overcoming. , 39(3):16–43, 2021.

C. Xuejun, Z. Baoping, Y. Jue, M. Rigen, and L. Jingjuan. Analysis on soil displacement and distribution of ground displacement field of shielding tunnel under the binhai thorough fare of shenzhen. In *2011 International Conference on Electric Technology and Civil Engineering (ICETCE)*, pages 5745–5749. IEEE, 2011.

K. L. Yates, P. J. Bouchet, M. J. Caley, K. Mengersen, C. F. Randin, S. Parnell, A. H. Fielding, A. J. Bamford, S. Ban, A. M. Barbosa, C. F. Dormann, J. Elith, C. B. Embling, G. N. Ervin, R. Fisher, S. Gould, R. F. Graf, E. J. Gregr, P. N. Halpin, R. K. Heikkinen, S. Heinänen, A. R. Jones, P. K. Krishnakumar, V. Lauria, H. Lozano-Montes, L. Mannocci, C. Mellin, M. B. Mesgaran, E. Moreno-Amat, S. Mormede, E. Novaczek, S. Oppel, G. Ortuño Crespo, A. T. Peterson, G. Rapacciuolo, J. J. Roberts, R. E. Ross, K. L. Scales, D. Schoeman, P. Snelgrove, G. Sundblad, W. Thuiller, L. G. Torres, H. Verbruggen, L. Wang, S. Wenger, M. J. Whittingham, Y. Zharikov, D. Zurell, and A. M. Sequeira. Outstanding challenges in the transferability of ecological models. *Trends in Ecology Evolution*, 33(10):790–802, 2018. ISSN 0169-5347. doi: https://doi.org/10.1016/j.tree.2018.08.001. URL https://www.sciencedirect.com/science/article/pii/S0169534718301812.