Spring 2024

Regional Sea Level Rise Prediction in Monterey Bay with LSTMs and Vertical Land Motion

Branden Lopez

Regional Sea Level Rise Prediction in Monterey Bay with LSTMs and Vertical Land Motion

A Project

Presented to

The Faculty of the Department of Computer Science

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Branden Lopez

May 2024

The Designated Project Committee Approves the Project Titled

Regional Sea Level Rise Prediction in Monterey Bay with LSTMs and Vertical Land Motion

by

Branden Lopez

APPROVED FOR THE DEPARTMENT OF COMPUTER SCIENCE

SAN JOSÉ STATE UNIVERSITY

May 2024

Dr. Genya Ishigaki	Department of Computer Science
Dr. Behin Elahi	Department of Industrial and Systems Engineering
Dr. Faranak Abri	Department of Computer Science

ABSTRACT

Regional Sea Level Rise Prediction in Monterey Bay with LSTMs and Vertical Land Motion

by Branden Lopez

Earth system data is vast in volume and variety, and is used to forecast weather, hurricanes, floods, and sea level. Sea Level Rise (SLR) impacts various sectors, especially ecosystems, food production, industry, population, health, and the availability of clean water. Because of its broad impact, describing the behavior and forecasting SLR is an important topic. Traditional Machine Learning (ML) models vary in use, but many are not capable of capturing all the non-linear spatial and temporal properties of SLR factors. Deep learning models efficaciously handle complex time series data, noise, and high dimensional spaces, making them a focus of recent SLR research. Long Short-Term Memory (LSTM) historically performs well for SLR predictions but has underperformed when forecasting regional SLR using altimetry data such as Mean Temperature Anomaly (MTA) and Oceanic Heat Content (OHC) time-scaled to quarters. This project proposes the inclusion of Vertical Land Motion (VLM) data, which are often disregarded by existing literature due to the lack of cohesive datasets, in the dataset along with oceanic and atmospheric variables. Our experiments focusing on Monterey Bay, California demonstrates that VLM data can improve the performance of LSTMs for regional SLR prediction. We also identify key LSTM features by feature importance computation. Furthermore, we assess the viability of using VLM in the presence of missing data points and its effects on the prediction.

Keywords: *Sea Level Rise, Vertical Land Motion, Deep Learning.*

ACKNOWLEDGMENTS

There are many people I would like to thank for helping me complete my master's degree and inspiring me in both academia and life. First, I want to thank Dr. Ishigaki for his guidance and feedback on this project, as well as Dr. Elahi and Dr. Abri for taking the time to work with me. When I started at Crafton Hills College and was placed in pre-algebra, I had no idea what I wanted in life. In math class, I met Professor Hanley, who inspired me with motivational videos and encouraged me to develop mantras of discipline and affirmation. Thank you, Professor Hanley; without your encouragement, I would not have focused on math during my undergraduate studies or invested the time to understand material I once found esoteric. This journey has been enriched by the friends I have made along the way, who have acted as support groups, sources of joy, and inspiration. While I wish I could thank each of you individually, doing so would require its own book. Lastly, I would like to thank my mother, Juana. Although our relationship has had its challenges, she supported me to the best of her abilities and taught me to work hard and carry myself with integrity.

TABLE OF CONTENTS

CHAPTER

CHAPTER 1

Introduction

The global coastlines exceed 1.6 million kilometers [1], and 84% of countries of the world have a coastline [2] with approximately 10% of the world's population [3]. These coastal zones hold immense economic value; for instance, maritime transport holds 90% of the world trade.

Sea Level Rise (SLR) is a vertical change of the oceanic sea level [4]. Because of the importance of the coastal areas, the consequences of SLR are a significant threat to economic activity and threats to global population displacement. SLR exacerbates flooding in low-lying coastal areas, increases coastal erosion, affects coastal ecosystems such as salt marshes, and introduces saltwater intrusion into estuaries and aquifers [5].

Many research works have explored the various factors that cause SLR, such as oceanic, atmospheric, and coastal variables. Many prominent factors that describe Global SLR pinpoint global sea-level rise at 1.7 mm a year [6]. However, Regional Sea Level Rise (RSLR) differs from the global phenomena due to local differences in density structure, ocean currents, and frequency of low-pressure systems [5]. For instance, the Grand Isle experiences a yearly RSLR of 9.24 mm [7], which is significantly different from the global average. In response, each existing research work on RSLR needs to explore the sea level analysis in a set of targeted regions, considering the unique characteristics of the regions. The accurate prediction of RSLR enables the assessment of flooding risks and sea level conditions for maritime transport planning.

This project proposes a model that can incorporate Vertical Land Motion (VLM) data for more accurate RSLR prediction. VLM is the vertical change of land [8] such as rise (uplift) or sink (subsidence). Glacial Isostatic Adjustment (GIA), which is the rebound of the earth's crust that once held sheets of ice [8], is known to contribute to subsidence or uplift in various regions. Also, there are other non-GIA factors near the coast such as ground resource extraction, land use, natural land loads, and sediment movements that can cause changes to SLR and capture processes that typical oceanic and atmospheric variables cannot capture. Despite VLM's possible contribution to SLR changes, existing works do not incorporate VLM because of data quality and availability issues. Natural events such as El Niño induce variability and force the temporal time frame to include many decades to capture these events. Also, VLM measuring devices are relatively new and are not found in abundance yet.

The main objective of this project is to demonstrate the effectiveness of including VLM data in the RSLR prediction process and identify the necessary data processing and suitable learning methods to deal with the complex nature of the VLM data. This novel proposal is driven by the fact that other papers do not incorporate VLM due to data quality issues such as length of recording and equipment failures that cause gaps in data. To provide a proof of concept, we take advantage of Monterey Bay's rich VLM records and combine VLM in Monterey Bay's coastal region with traditional oceanic and atmospheric variables. Modeling with time series models such as Autoregressive Integrated Moving Average (ARIMA) and Long-Short Term Memory (LSTM) then using interpretability with SHapley Additive exPlanations (SHAP) to assess which factors drive RSLR in Monterey Bay. While traditional oceanic and atmospheric variables describe a large portion of RSLR in Monterey Bay, VLM is found to have importance likely from seasonal non-GIA contributions.

This paper is divided into several chapters and covers the following topics. Chapter 2 covers the three typical categories of models used to forecast SLR and their related works, while Chapter 3 is an overview of factors that historically influence SLR on a global and regional level. Chapter 4 introduces the various pre-processing, correlation analysis, models, metrics, and feature importance methods for this project. Chapter 5 is a pragmatic analysis of the various methods and results, which assess the impact of our variables. Chapter 6 goes into greater detail about the VLM history, data collection, and interpolation in the presence of missing data. Lastly, our conclusion is in Chapter 7.

CHAPTER 2

Related Works

Many works explore SLR with various models, features, spatial, and temporal frames. However modeling can generally categorized as Machine Learning (ML), Deep Learning (DL), and alternative models that use intrinsic properties of RSLR to extrapolate or model using statistical methods.

2.1 Machine Learning for Sea Level Rise

Machine Learning models for SLR includes Support Vector Machines (SVMs) [9, 10], Autoregressive Integrated Moving Average (ARIMA) [10], and Decision Trees (DT) [11]. These models have performed well when predicting for larger global regions and individual stations. SVM has consistently produced the highest metric scores in contrast to other ML methods [9, 10, 11]. Sithara [9] uses SVM for the Willingdon Island tidal gauge station and produces satisfactory testing results and improves results by extending to Wavelet SVM. In Wavelet SVM, wavelet transforms are used to denoise the input signals before training. This additional step improves testing results by 30% and showcases a difficult of SLR, noisy signals and non-linearities. While DTs deal well with non-linear complexities, among all cited ML methods it typically performs the worst. Altunkaynak [11] provides insights by forecasting on daily SLR and showcased that DTs have difficulty establishing connections between past and future SLR, this emphasizes the temporal complexity of SLR. While ARIMA models have typically thrived in forecasting temporal data, Balogun [10] shows that ARIMA models are not as capable as SVMs, because ARIMA is a linear model and is not able to capture the non-linear correlation of each feature. Lastly, recent work has gone further and compared DL models to traditional ML methods and found that it outperforms all covered methods.

2.2 Deep Learning for Sea Level Rise

Balogun utilizes Long Short-Term Memory (LSTM), a DL model, in his study the LSTM outperforms traditional ML models. This is expected as DL models can learn from complex non-linear features, noisy signals, and interactions. More so, the LSTM a type of Recurrent Neural Network (RNN), a DL model that exceeds in establishing past connections from temporal data. However, the use of DL for SLR is not new and a 2003 paper by Huang [12] used a simple 3 layer Neural Network to predict for SLR at the South Shore of Long Island, New York. This simple DL model produces excellent results, results better than Balogun's LSTM but this is expected as Huang uses the tide gauge SLR measurements of sites within 100m of his target. Using SLR from other tide gauge stations in a small spatial region should intuitively lead to high accuracy models because SLR variability generally doesn't change significantly at small distances. Instead, Balogun utilizes a wide array of Oceanic and Atmospheric Variables for a much larger spatial region; therefore, his study is concerned with predicting from factors that drive SLR for an entire region. Similar to this, Nieves [13] used a paltry quantity of oceanic variables that describe heat to predict and describe SLR behaviors on a quarterly frequency. Nieves compares the performance of an LSTM to a Gaussian Process (GP) model to an LSTM with the same inputs. With the GP it's found that oceanic heat variables describe at least 62% of coastal SLR variability in various coastal regions, but the inputs with an LSTM were incapable of predicting with much success; Nieves notes that LSTM forecasting can be improved by increasing the number features that affect SLR.

2.3 Alternative Models for Sea Level Rise

While we have only covered AI models to predict SLR there are alternative ways predict SLR. For instance, Hamlington [14] extrapolates SLR using only the rate and acceleration of historical observations, while applying corrections for Glacial Isostatic Adjustment (GIA), and the Inverse Barometric Effect. Results for various regions in the USA largely produce results consistent with The Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6), a reputable source that attempts to describe SLR behavior. Often these other sources attempt to quantify the impact of land coastal changes on SLR, which is often overlooked with AI models. Harrison [15] utilizes a Monte-Carlo model with a number of features that drive global SLR, 3 local variable at a 21 tide gauge stations, and corrects for aquifer depletion. Poppe [16] combine two models, Marsh Equilibrium Model (MEM) and the Relative Elevation Model (REM) in response to SLR to predict for changes in sea grass marshes. The results show that RSLR can impact elevation changes.

Because of DL's of versatility this work intends to use LSTMs to predict RSLR in the Monterey Bay Region. This region is greatly impacted by climate modes such as El Nino, has much of it's variability described by OHC, and VLM measuring equipment are of high quality. However, only adding coastal land feature limits the dimension and the ability to capture the important drivers of RSLR. Therefore this study will be one of the few to study RSLR in Monterey Bay, driving features of RSLR, and quantify the importance of coastal land variables to RSLR at various temporal frequencies.

CHAPTER 3

Sea Level Rise Factors

In this chapter we will explore the importance of SLR and it's driving oceanic, atmospheric, and coastal variables.

3.1 RSLR Data

Estimations of current and future RSLR can vary dramatically due to time-series length, evaluation methods, and data collection methods. For instance tide gauge data at Monterey Bay produced significant differences in average yearly SLR in contrast to altimetry data over similar regions [7]. A large reason for this it that altimetry data records the mean SLR over an area defined by measuring equipment and processing. While altimetry SLR might not allow us to specifically assess southern Monterey Bay as our tide gauge station does, the increase of volume and quality of spatial altimetry data allow us to gather factors of SLR in the entire Monterey Bay region. Capturing the broader Monterey region is important as the open ocean influences coastal region. Therefore we capture oceanic descriptors such as water heat content at the surface, at various depths, melting glaciers, tides, and natural inter-annual variations, such as the Pacific Decadal Oscillation (PDO) by analyzing the greater Monterey Bay.

3.1.1 Oceanic Heat Content

The total amount of heat stored by the oceans is called Ocean Heat Content (OHC) and when studied over long time frames OHC derived products can describe 70% of the variations in North Eastern Pacific Ocean (NE-PO), which encompasses the American West Coast [13]. Various reasons for this modeling capacity exists, but much of it can be attributed to the ocean's physical interaction with the atmosphere.

The heat capacity of the ocean is about 1000 times larger than that of the atmosphere [5] and due to this the release of Greenhouse Gases (GHGs) and increasing atmospheric temperatures, leads to oceans absorbing much of this heat through the surface layer; where water temperature and density are nearly uniform due to strong mixing of surface waters by the wind. The heat stored at the surface layer is diffused to the deep layers. It is estimated that the heat storage in the ocean accounts for about 90% of the heat which the earth absorbed for the past 40 years. With this increase in heat, there is a thermal expansion of sea water and SLR occurs.

3.1.2 Melting Ice

The melting of land-based ice consists of two major sources, the melting of mountain glaciers and caps, and ice sheets melting at Greenland and Antarctica. The water content stored in glaciers is considered to be equivalent to 0.5m of Global Mean Sea level rise [5], while the Greenland Ice Sheet (GIS) and West Antarctica Ice Sheet (WAIS) contain ice equivalent to about 7m and 3–5m sea level rise, respectively [4]. Therefore melting due to global warming can result in significant increases to SLR and RSLR due to runoff from mountains.

Despite significant importance on global SLR, studies based on ocean general circulation models [17, 18], confirm that regional sea level trend patterns reported by satellite altimetry are mainly due to regional variability in thermal expansion; thus, ice glaciers are insignificant for modeling RSLR but can be reflected thru other variables.

3.1.3 Tides and Currents

Tides are the alternate rising and falling of the sea, usually twice in each lunar day at a particular place, due to the attraction of the moon and sun. Because of this the sea levels at a coast varies at each time of the day. The National Oceanic and Atmospheric Administration (NOAA) dedicates itself to measuring products such as SLR at Tide Gauge Stations across the US coast; while this provides a coastal SLR data source, NOAA's default datum is Mean Higher High Water (MHHW).

A datum is simply the reference elevation used to record water heights and depths. MHHW is the average of the higher high water height of each tidal day observed over the National Tidal Datum Epoch. This datum allows us to capture extreme events such as storm surges, capture relevant water levels for coastal communities, and better asses risks of flooding.

Despite this most studies use the Mean Sea Level Rise (MSLR) which is simply SLR [10, 13, 5], using this datum makes it easier to integrate data sources from different locations. MSLR which is the arithmetic mean of hourly heights observed over the National Tidal Datum Epoch, and provides a stable reference point for SLR over time. However it's important to note that extreme events can increase coastal erosion, alter estuary dynamics, morphodynamics, and flooding. In addition Price [19] found that sea level fluctuations on the west coast of the United States propagate across the ocean as long Rossby waves and are coherent with sea level fluctuations at Hawaii; again in 1986 they showed similar results in the Atlantic and Europe. Therefore, it might more important choose other datums when modeling short term behavior and tide gauge stations on the Eastern Pacific could describe Monterey Bay tides.

3.1.4 Climate Modes

Climate modes or climate oscillations are natural climate variabilities that occur over various spatial and temporal regions that can influence weather patterns, temperatures, precipitation, and other climate related variables. Two examples of climate modes in the NE-PO are the Niño-Southern Oscillation (ENSO) and Pacific Decadal Oscillation (PDO).

ENSO events occur irregularly every few years and are characterized by warmer-than-average sea surface temperatures in the central and eastern Pacific Ocean, along with changes in atmospheric pressure, and wind patterns. Because of this large interannual variability, substantial trends to the global water cycle have been found but the sign and magnitude of this trend strongly depend on the time window over which the trend is estimated [20].

PDO is a decadal climate mode characterized by variations in sea surface temperatures and atmospheric pressure in the North Pacific Ocean. Positive PDO phases are associated with warmer-than-average sea surface temperatures in the eastern North Pacific and cooler conditions in the Central North Pacific, while negative PDO phases exhibit the opposite pattern. It has been found that PDO explains the significant multidecadal variability that was manifest as a sea-level rise along the Pacific coast of North America from 1993 to 2009 that was considerably smaller than the global average [21], while sea level in the Western Pacific rose at a level three times higher than the global mean [22, 20].

Climate modes introduce a significant issue when modeling because of their inter-annual and decadal nature, and the temporal longevity of datasets. Because of this predictions are typically done with datasets spanning as far as the 1955s when modeling SLR in hopes of capturing the natural oscillations [13]. Others have modeled RSLR by using coastal factors and climate modes extracted from SLR using mathematical modeling, and found that climate modes account for up to 9% of SLR in the Mediterranean sea [23].

Chelton [24] showed that ENSO-related coastal Kelvin waves propagate rapidly along the eastern boundary of the Pacific Ocean and appear prominently in tide gauge series there, with delays on the order of a few months. Rather than remove this variability, we intend to learn from it as prior methods that remove internal variability utilize regression analysis despite ENSO events not being perfectly linear and so the regression model can underestimate the full effect [14].

3.2 Atmospheric Variables

Atmospheric variables are numerous and have been extensively studied for their ability to describe variability in SLR. Like oceanic variables, atmospheric variables can effect SLR on a global and local scale because of this it is important consider various atmospheric variables when modeling for RSLR. Some variables are air temperature, air pressure, and wind.

3.2.1 Air Temperature

Air temperature is a major factor influencing seawater temperature because it is a medium of heat transportation and heating of the water causes thermal expansion. When sea surface temperature data quality is lacking, air temperature has been used as an indicator due to seasonal constituents between sea level and air temperature being maintained [25].

3.2.2 Air Pressure

Sea level is known to be affected by local atmospheric pressure in a phenomenon known as the "inverted barometer effect" and an increase in atmospheric pressure by 1 millibar should depress the de-tided sea level by 1 cm and vice versa. Air pressure has been found to be a major contributor to SLR in the Red Sea [25].

However there are some caveats. In studies of coastal-trapped waves, it is often found that the local inverted barometer effect is not unity, where wind effects are large and there is a long coastline equatorward of the observations. In regions where atmospheric pressure fluctuations are significant and can introduce noise or bias into sea level observations, sea level measurements have been corrected to remove the contribution of atmospheric pressure variations. It is commonly found that this adjustment is effective at high frequencies but not at lower [26].

3.2.3 Wind

Wind stress is the shear stress exerted by the wind on the surface of large bodies of water. This stress can lead to SLR through its role in moving warm water from one part of the sea to another [27].

There are numerous studies on wind stress, for instance, sea level variability off California is correlated with equatorial wind stress variability at periods as long as 20 years and is associated with ENSO events [28]. Likewise, as with the western coast of Europe during times of higher sea level, the coastal winds are either stronger to the north or weaker to the south. In fact much of variability in sea level at San Francisco is consistent with variability in longshore wind forcing, which are winds blowing parallel to the coastline [26]. Further studies have found that wind speed along the coastline are critical to predicting SLR on the Malaysian coast [10].

3.3 Coastal Land Variables

Sea level varies spatially according to the shape and geometry of the coast. Despite this prior studies will only use one or a combination or oceanic and atmospheric variables. To better capture SLR variability in specific coastal regions it is beneficial to capture coastal land variables.

3.3.1 Vertical Land Motion

Vertical Land Motion (VLM) is the vertical change of land over time, in VLM land can experience uplift or subsidence. This change can be attributed to number of natural and anthropogenic factors.

There are many natural factors that affect VLM, most notable is Glacial Isostatic Adjustment (GIA). GIA is the long-term rebound of the earth's crust that once held ice sheets during the last glacial period and represents the single largest factor driving VLM [8]. Gauges that measure SLR need to be adjusted for VLM to find the absolute

SLR but data on VLM at these Tide Gauges have not always been present. For this reason many prior studies [29, 30] assumed that the vertical land motion at many stations were negligible or discarded stations with considerable VLM.

Non-GIA sources have about 70% of the impact of GIA's contribution to VLM [6]. However this includes tectonic uplift, isostatic adjustment following erosion, sediment loading, and changes in mantle flow. Shorter term and unpredictable non-GIA changes can arise from elastic response to contemporary ice loss owing to climate change, storage changes within aquifers and hydrocarbon reservoirs, surface loading owing to changes in the terrestrial hydrosphere, and sediment compaction due to land use changes [31].

Many of these non-GIA processes are difficult to track due to their spatial dimension and interaction with one another. For instance, hydrocarbon extraction along the cost can lead can lead to significant sediment loading into nearby oceans, this increases regional seawater salt content and can cause haline contraction. That is, changes in seawater salinity can also influence sea level. Because when water becomes more saline, it becomes denser and contracts slightly, leading to a decrease in sea level. Conversely, when seawater becomes less saline, it becomes less dense and expands slightly, contributing to an increase in sea level. Likewise, extraction of oil or aquifers leads to land subsidence.

Because these natural and anthropogenic factors can lead to measurable changes to VLM, it's important to correct tide gauge data for VLM at the measuring station [32]. Likewise, tracking VLM around the greater coast region may be useful capturing SLR variability.

3.3.2 Bathemetry

While individual sensors and satellite imaging can estimate topographic maps for VLM, bathymetric or seafloor maps exist but not to a large extent [2]. However recent advances have used a mix of various satellite imaging to approximate bathemetry at various depths.

GIA and non-Gia processes affect bathymetric changes much as it does with topographic VLM changes, for this reason SLR prediction with underwater VLM changes could describe variability in SLR. However we leave this variable from our input variables as the focus of this paper is VLM.

3.4 Region

Usually RSLR predictions focus around big city areas and land that is highly susceptible to flooding. This often leaves information about other coastal areas unknown and therefore the driving factors of RSLR in many regions are unknown. A coastal land is affected by large spatial regions which have reduced variability due to RSLR being an average of the spatial area. Instead, this paper focus on a single point in Monterey Bay, a tide gauge station in the south of the bay. Focusing on a single point will introduce natural variation caused by waves at the shore, but should tell us how a individual locations could be greatly impacted by greater SLR factors. Attributing to the difficulty of this region is the fact that it's a bay. Holleman [33] shows that the channeling of water in the San Francisco Bay Area alters flux in the region and changes tidal energy, which typically amplify the tides in a region.

CHAPTER 4

Sea Level Rise Estimation Methods

As discussed there are various methods to model SLR. Many of these methods start before modeling, data must be pre-processed, correlation analyized, metrics must be selected. Afterwards its important to compare traditional models to LSTMs and determine feature importance.

4.1 Pre-Processing

Pre-processing is an important step when modeling, this is because many forecasting models assume that data is equally spaced, with no null values, and is smooth.

4.1.1 Linear Imputation

Linear interpolation or imputation is a method of curve fitting using linear polynomials. The interpolation is between two known points with unknown values in between our indices x_0 x_1 and known values y_0 and y_1.

$$y = y_0 + (x - x_0)\frac{y_1 - y_0}{x_1 - x_0}$$

4.1.2 Cubic Imputation

Cubic interpolation is curve fitting with polynomials of three degrees. Polynomial interpolation is better for approximating non-linear data and generally produces a smoother curve than linear interpolation. There are many methods of cubic interpolation, our module Pandas uses Akima interpolation. Akima interpolation uses piecewise cubic polynomials to produce continuously differentiable sub-splines which are more smooth and natural.

4.1.3 Differencing

A stationary time series is one whose properties do not depend on the time at which the series are observed [34]. Therefore, SLR is not stationary because we know it has been increasing year over year and for this reason the data must be differenced.

Therefore we subtract the current value of the series from the previous one, this helps bring the mean closer to zero and eliminate the trend. We can do this as many times as needed, therefore we need a test to determine how many times we should difference.

The Autocorrelation Function (ACF) measures the linear correlation of an observation at time t against prior times $t-k$ [35], while the value k is the lag. Therefore differencing 1 time is indicative of a high value for lag 1. Luckily, many modules exist for assessing ACF and automatically plot the values. If a lag value is outside of the 95% confidence interval, then it should be differenced. If multiple values lie outside of the confidence interval, then we difference, and iteratively use the ACF until stationarity is achieved.

4.1.4 Smoothing

Oftentimes, there might be too much noise in a signal and it must be smoothed. The smoothing generally reduces the total variation and therefore limits interpretability but allows models to capture the general trend that influence variability.

4.1.4.1 Exponential Smoothing

Simple Exponential Smoothing (SES) is fitting for data with no clear trend or pattern. In SES [35] forecasts are calculated using weighted averages, where the weights decrease exponentially as observations come from further in the past. It takes one parameter, alpha (α) with higher values prioritizing current observations rather than past values. This allows us to smooth much of the original data without over-smoothing. Allowing our models to capture the general trend without overfitting to outliers.

4.2 Correlation Analysis

Correlation Analysis can unveil important variable interactions, point to collinear variables, and point to feature importance. Often times feature correlation is quantified linearly but other methods such as SHAP can implicitly determine feature important after modeling.

4.2.1 Linear Correlation Coefficient

While data patterns can be explored through graph, this is largely subjective and cannot numerically quantify correlation. For this reason the Linear Correlation Coefficient exists, to describe the magnitude and direction of the linear relationship.

$$r = \frac{\sum_i^n \frac{x_i - \bar{x}}{s_x} \frac{y_i - \bar{x}}{s_y}}{n-1}$$

Where $\bar{x}$ and s_x are our sample mean and sample deviation of the x's, and $\bar{y}$ and s_y are our sample mean and sample deviation of the y's. This method not only will tell us the correlation between response and predictors but between the predictors themselves.

The correlation coefficient is therefore a great tool to use for pre-processing and some papers [10] use it to measure how input features can linearly prove results. However, r for quantifying input-output improvement is not beneficial with a non-linear model.

4.3 Models

Having high quality data is important for modeling and choosing a model that can capture the characteristics of our data is just as important. In this section we cover a few models that should perform well.

4.3.1 ARIMA

To understand ARIMA we much first cover its individual components, Autoregression (AR), Integrated (I), Moving Average (MA).

4.3.1.1 Autoregression

The term autoregression indicates the regression of the variable against itself; therefore an autoregressive model predicts using a linear combination of past values or lags of the variable. The term p is used to indicate the number of autoregressive variables used. An AR model with constant c, autoregressive weight ϕ, and error ϵ can be written as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t$$

4.3.1.2 Moving Average

While a moving average is often thought as a smoothing method similar to exponential smoothing, MA models are a linear model where our variables are past errors. The term q is used to indicate the number of errors used. Therefore a moving average model with error weight θ

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q}$$

4.3.1.3 Integrated

To simply put it, the integrated (I) simply means differencing has been integrated. The differencing d occurs before hand in models that automatically select our AR and MA degrees. The term d is equal to the differencing used. Therefore an ARIMA model with response differenced as y' can be written as:

$$y'_t = c + \phi_1 y'_{t-1} + ... + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}$$

4.3.2 Seasonal ARIMA

Similar to ARIMA is Seasonal ARIMA (SARIMA). The SARIMA model is more suitable in modeling time series with seasonality, or lags at occur at specific periods. SARIMA introduces three new but similar parameters, P, D, and Q. Where P is the number of seasonal AR terms, D the seasonal differencing, and Q indicates the number of seasonal errors used. For quarterly data, P will use lags of 4 (the frequency) times the lag number to model; its similar for M but this time lags are 12 periods apart. The same can be said for the seasonal moving average Q, where the frequency indicates the lag k spaces apart.

4.3.3 Deep Neural Networks

Deep Neural Networks (DNNs) are feed-forward models with stacked layers of artificial neurons. Arguably the most used neuron is the Perceptron, which takes a weighted sum of the inputs and transforms this with a non-linear activation function. This value is then passed to every neuron in the next layer until our prediction is reached. This output is compared to the actual observation with some metric and then backpropagation finds how each connection weight should be altered in order to reduce the error [36]. This DNN is capable of learning complex nonlinearities and interactions.

4.3.4 Recurrent Neural Networks

Feed-forward models will only pass activations forward but RNNs have connections backward. This backward connection allows it to take multiple observations in sequence and predict with the activation of the prior step. Therefore the output at time step t is a function of all the inputs from previous time steps, which is what gives an RNN its "memory". The memory of an RNN unit is referred to as the hidden state. Backpropagation is similar to an DNN but this time all cells associated with output are updated and because every cell is represented by a prior hidden state, backpropagation works over all time steps.

4.3.5 LSTMs

LSTMs are RNNs but utilize an improved cell for faster convergence, and better memory. There are many intricacies to the cell that give to better memory.

For instance, LSTMs have two memories. A short term memory $h_{(t)}$ and long term memory $c_{(t)}$. In the figure below the long term memory $c_{(t-1)}$ is passed into the LSTM cell, where our Forget Gate will drop unimportant memories. From here newer memories are added from the Input gate. Without further transformation we have arrived at our new long-term memory $c_{(t)}$, but the short term memory requires a few more steps. To attain the short term memory, $c_{(t)}$ is transformed with tanh and the important parts are kept to create our new $h_{(t)}$.

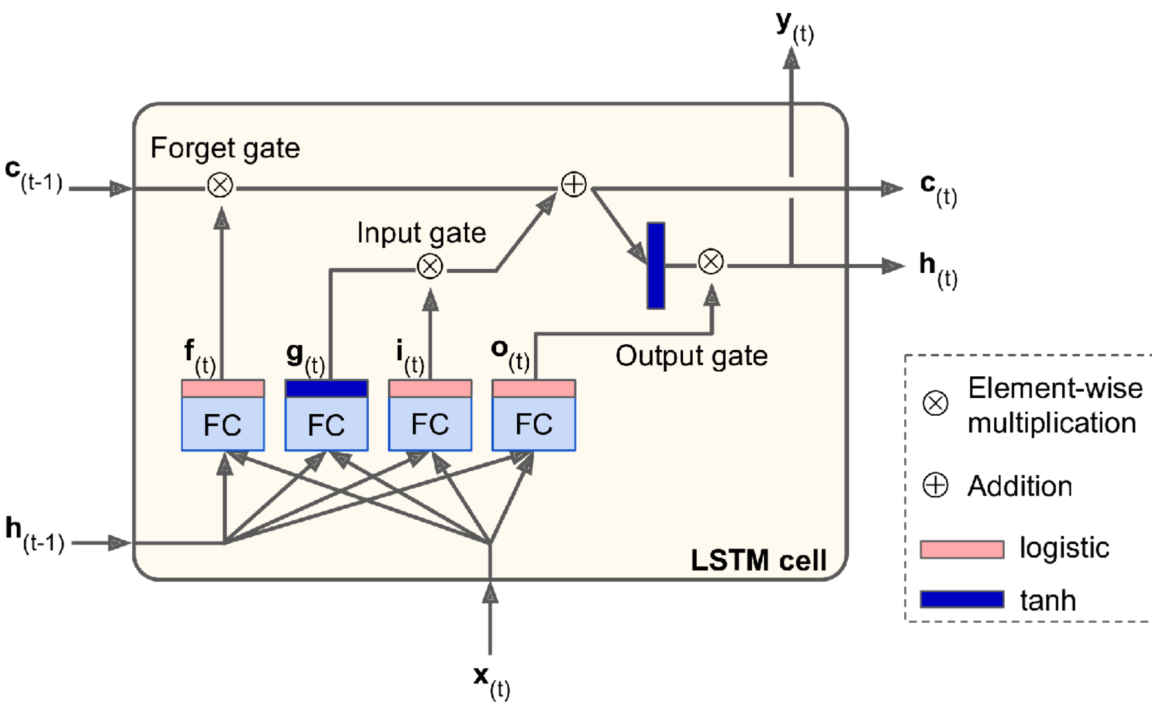


Figure 1: LSTM Cell [36]

While coding this might appear difficult open source languages such as PyTorch give access to optimized LSTM layers.

4.4 Metrics

There are several metrics that are commonly used to assess SLR and interpolation results. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are just a few.

4.4.1 Root Mean Squared Error

Root mean squared error is the square root of the average squared difference between predicted value and actual value. It is represented below

$$RMSE = \sqrt{\frac{1}{n}\sum^{n}(Y_i - \hat{Y}_i)^2}$$

Where n is the total number of observations, Y_i is the observation i, and $\hat{Y}_i$ is the predicted value for observation i. RMSE is sensitive to large errors but is also insensitive to outliers.

4.4.2 Mean Absolute Error

Mean Absolute Error (MAE) is the average absolute difference between predicted value and actual value. This is different from RMSE as errors now scale linearly and the error is less impacted by outliers. MAE is represented below.

$$MAE = \frac{1}{n}\sum^{n}|Y_i - \hat{Y}_i|$$

While, both capture residual error their use is based on the scenario. For example, Monterey Bay is not as prone to flooding so if we simply want to describe the overall trend in forecasting we could use MAE. However, if we modeled a region with significant risk of flooding then it might be beneficial to use RMSE when analyzing forecasting accuracy in order to better capture seasonal SLR prediction to ensure any safety measures have near zero error. Likewise, it's best to train our LSTM with RMSE as it will help penalize the error caused by seasonal fluctuations.

4.5 Feature Importance

ARIMA models are linear and linear coefficients give description about an input's ability to change the response, but LSTMs do not have this intrinsic property. Instead, we must assess feature importance of an LSTM with methods such as SHAP.

4.5.1 Shapley Additive Explanations

Shapley Additive explanations (SHAP) determines feature importance by calculating the contribution of each feature and is formed from coalitional game theory [37]. In which a player (feature) working in conjunction with others, receive a value associated to how much they contribute to a payout (prediction).

To accomplish this there are 5 steps we must satisfy to find these values.

1. Sample coalitions $z_k^{'} \in \{0, 1\}$ (1 = feature present in coalition, 0 = feature absent).
2. Get prediction for each $z_k^{'}$ by first converting $z_k^{'}$ to the original feature space and then applying model $\hat{f} : \hat{f}(h_x(z_k^{'}))$
3. Compute the weight for each $z_k^{'}$ with the SHAP kernel.
4. Fit weighted linear model.
5. Return Shapley values ϕ_k, the coefficients from the linear model.

Where the function h_x maps 1's to the corresponding value from the instance x that we want to explain. 0's are replaced by random feature values from data. In figure 2 we see the general methods flow chart used in this project.

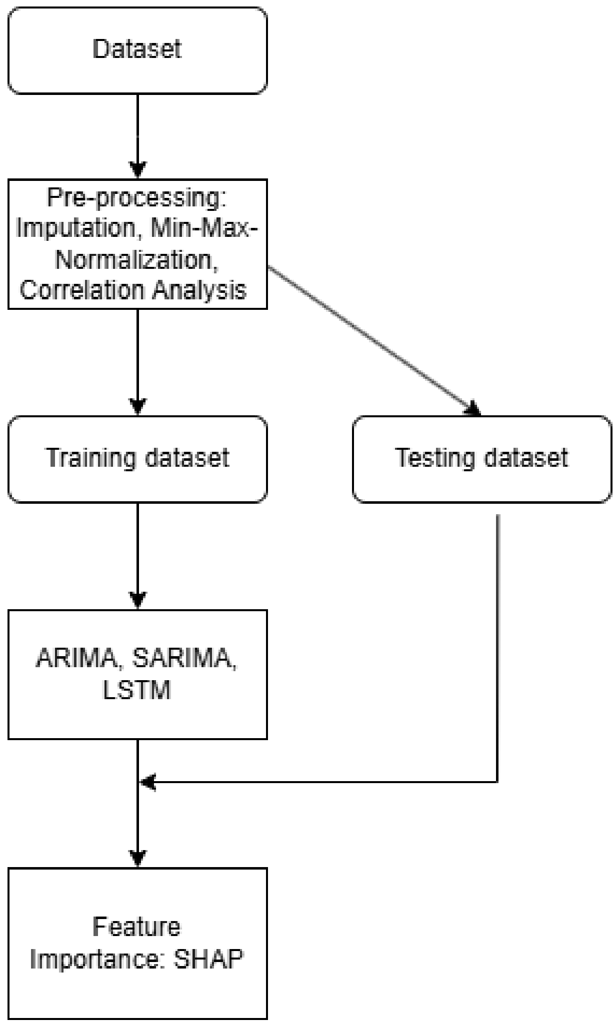


Figure 2: Training Flow Chart

CHAPTER 5

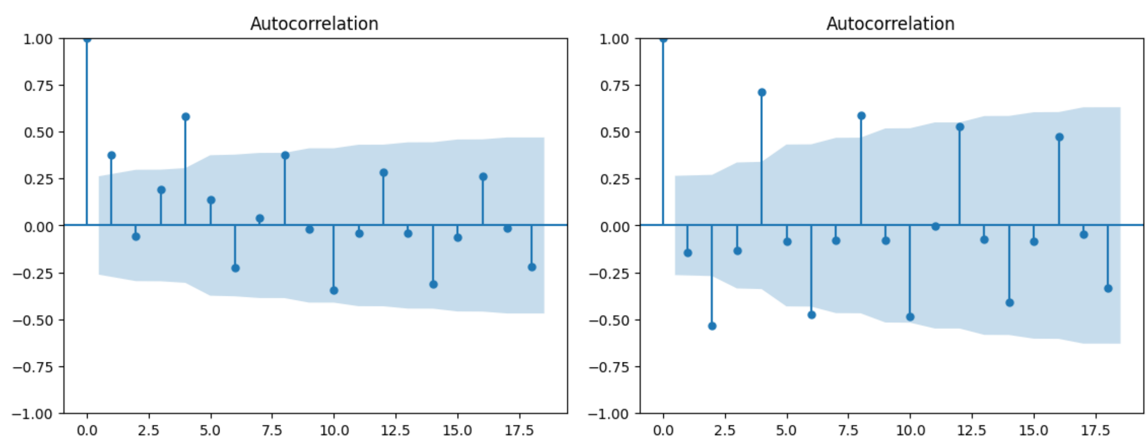
Experiments and Results

As discussed SLR variables have various impacts depending on the temporal frequency. For this reason two experiments are set up to predict SLR on a monthly and quarterly frequency. This chapter will cover experiment setup, correlation analysis, results, and analyze what our results mean on RSLR at the Monterey Bay area.

5.1 Quarterly Frequency

While a quarterly frequency might not describe daily variations, it allows us to encapsulate the trends and seasonality of our data while reducing the amount of noise. Furthermore, the frequency allows us to describe data with less interpolation, use variables which are not measured at higher frequency, and it allows us to generalize further because it emphasizes long-term periods. Due to the noise and trend differences among frequencies, quarterly data must be analyzed in it's own frame.

5.1.1 Differencing



(a) ACF without differencing

(b) ACF 1 difference

Figure 3: Quarterly autocorrelation function results

Our data is analyzed for autocorrelation using ACF. If our lags are found to be outside the confidence interval we difference our data according to the number of lags outside the interval. In Figure 3 we see that our ACF indicates a correlation with the first lag and the 4th. For this, we apply first-order differencing to achieve stationarity. After doing this we see that lags in multiples of 4 stand out, this indicates there is seasonality in our data and we could even implement seasonal differencing. Despite this it's beneficial to model and determine if further differencing is needed.

5.1.2 Linear Correlation

With data differenced for stationarity we now analyze our variables for correlation. While prior papers have influenced our data collection process to include a vast collection of variables, each region is affected by the variables differently. Despite this, many variables measure the same feature at different depths. For this reason, correlation analysis is utilized to remove variables with weak response correlation. In figure 4 we showcase our variables after correlation analysis. Surprisingly, our oceanic heat content variable, T_dc_mt_700, shows little linear correlation with our response mean_height, while a prior study used this to explain at least 62% of the variability in RSLR. Intuitively, this likely indicates a strong non-linear correlation between the two.

While the eastern and northern current velocities have a high linear correlation together, it's expected as the two often work together and for this reason, they remain in the dataset. Lastly, salinity shows little correlation with other variables despite thermohaline circulation indicating that seawater density and circulation are affected by this.

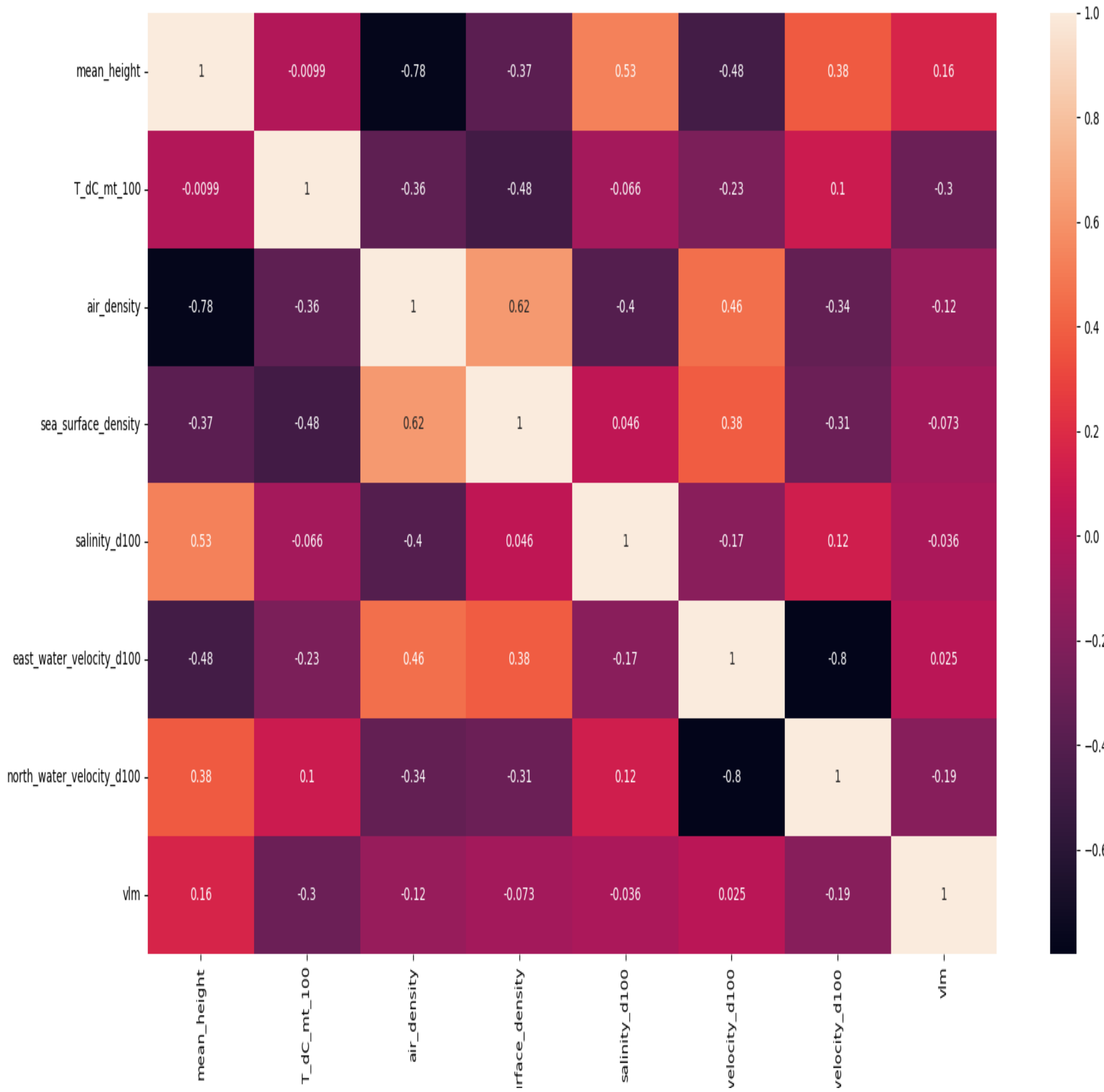


Figure 4: Quarterly correlation analysis results

5.1.3 Modeling

ARIMAX enables us to model with the features in figure 4, while it's typically best to find a subset of features that accurately describe our response, it can indicate the best possible fit with traditional time series forecasting models. Because of this, we overfit an ARIMA model with all possible features and also let auto

SARIMAX select a best model. Table 1 showcases the results of SARIMAX and the LSTM. In comparison to LSTM results, traditional ML methods are not as capable.

Table 1: Quarterly modeling results in millimeters

Model	VLM	RMSE train	RMSE test
ARIMAX All Vars	Yes	17178	61
SARIMAX Auto	No	6642	38.8
LSTM All Vars	Yes	32.1	16
LSTM Subset	Yes	28	24.7
LSTM Subset	No	28.69	28.94

Training the LSTM with 1 layer and 50 hidden units, we achieve lower RMSE training and testing values. While deeper models typically capture more complexities, increasing the number of layers decreased model performance and generally leads to linear predictions. Despite this, an LSTM with one layer still captures non-linear interactions but indicates that RSLR for Monterey day is not highly non-linear. Likewise, reducing our dataset to 7 input features from 18 decreases the RMSE. While it is only a 4mm difference in training RMSE, its more noticeable with our testing RMSE with a difference of 8mm. Indicating that some of the removed variables likely have the ability to describe SLR behavior in the region.

Likewise, removing VLM from the subset model decreases training RMSE by only 2mm and testing RMSE by 4mm, a large difference for a single feature. Because the smaller number of subset features, intuitively it makes sense why removing a single variable will impact RMSE so much. However, the decrease in test RMSE indicates that this variable is useful as it leads to better generalization performance. Therefore the idea that VLM could add meaningful predictive capability is proving to be true.

To help assess the what the model is learning, we utilize KernelSHAP to determine feature importance.

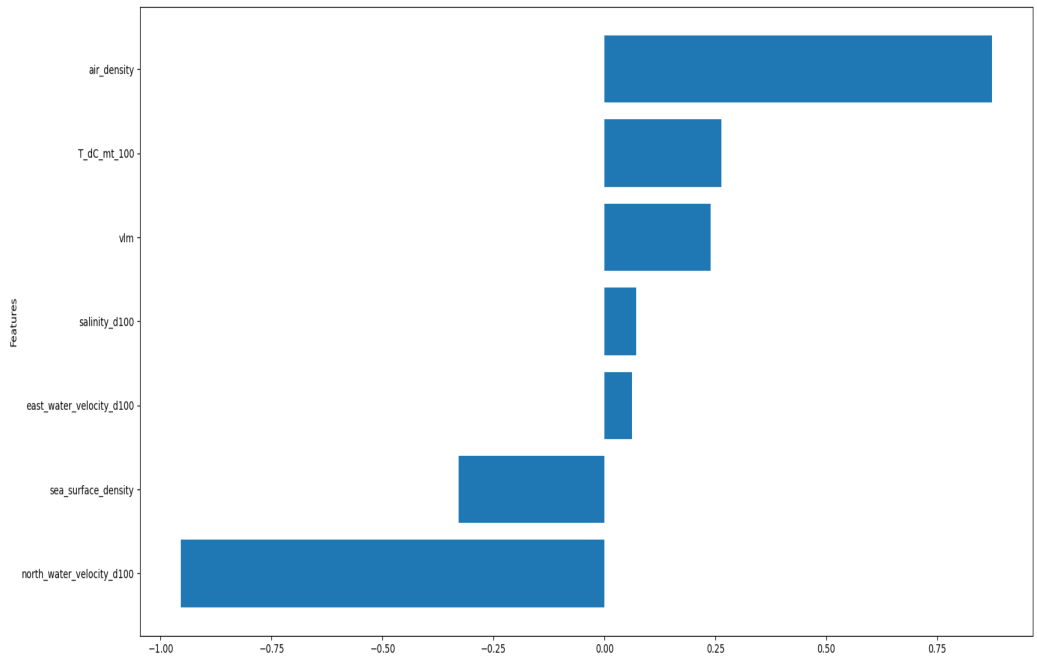


Figure 5: Quarterly SHAP average feature importance

Most of the results are rather intuitive. As north water velocity increases, the mean sea level at the Monterey bay station will decrease. Because the Monterey bay station is southern, this could indicate that water is being moved away from the station. Likewise, as water density increases our water will contract in volume, leading to a decrease in sea level. Heat a leading component to water density increases RSLR as expected.

Air density changes with variations in atmospheric pressure, temperature, and humidity. Due to the inverse-barometric effect we would expect air density to decrease RSLR and not increase it but the other components of air density are likely driving this change. In Northern California winter experiences low temperatures and low humidity which increase atmospheric pressure but this period also experiences increases of storms, and east to west ENSO events which drive currents. For this reason, it's logical to assume that air density reflects strong seasonal trends.

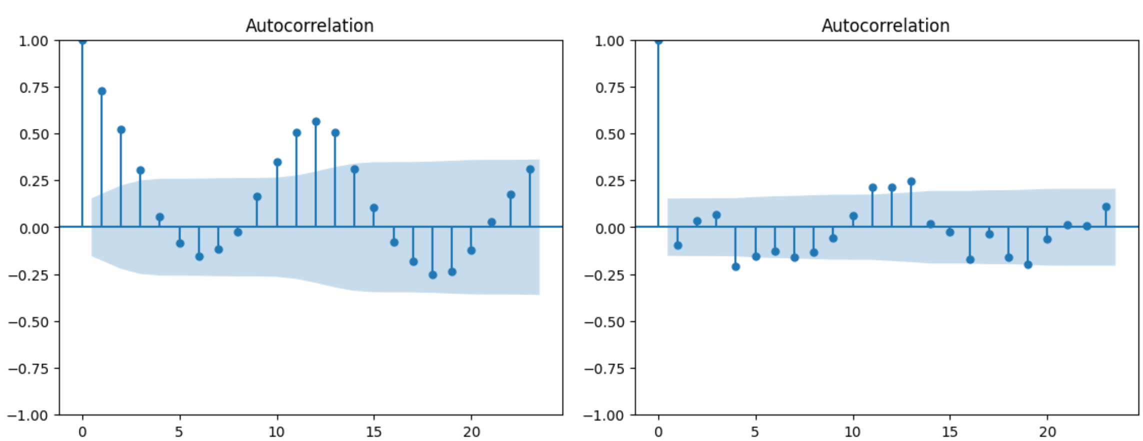
VLM has a positive impact and thus increases to VLM will increase RSLR. There are many reasons for why this could occur. For example the weight of snow could cause our stations at higher altitude to experience higher load and therefore subsidence during the winter, while storms can lead to lead to groundwater recharge uplift. Considering the positive effect of VLM it's logical to assume that VLM is affected by similar process that drive up SLR in this region.

5.2 Monthly Frequency

With quarterly data observed, it's time to see how monthly data varies with our features.

5.2.1 Differencing

Like the quarterly case, we observe lags outside of the confidence interval. Rather than difference by 3 lags at once, I aim to iteratively difference and then check for stationarity. In Figure 6b we see that with one lag our response has been made stationary, with the exemption of seasonal lags. For this reason we stop here.



(a) ACF without differencing

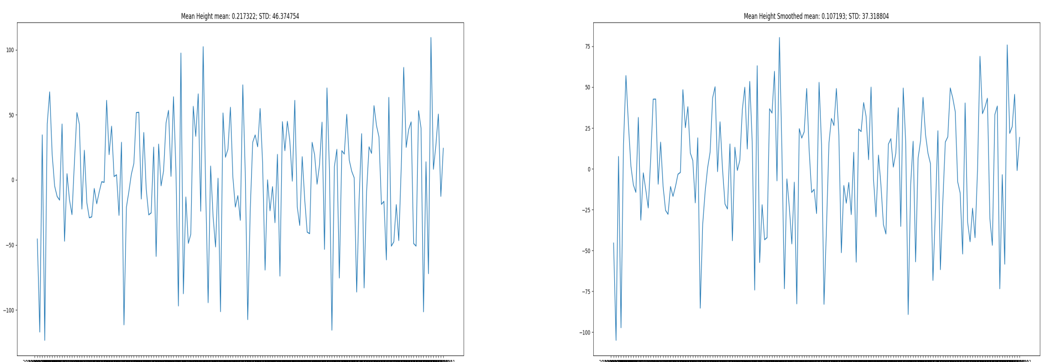
(b) ACF 1 difference

Figure 6: Monthly autocorrelation function results

5.2.2 Smoothing

When modeling for quarters there were 4 observations a year and now there are 12. For this reason there is now more noise and this noise prevented the LSTM from learning any meaningful pattern to the data. For this reason we utilize exponential smoothing with the monthly frequency. Because smoothing often widens the gap between modeling and interpretation of results we set our parameter alpha to .80. This takes pays more attention to the current sample at hand rather than historical values.

If figure 7 the overall structure of our SLR is similar in both images but 7b exhibits significantly less variance, a 9mm difference for standard deviation.



(a) Unsmoothed SLR

(b) Smoothed SLR

Figure 7: Monthly mean height (SLR) smoothing differences

5.2.3 Linear Correlation

Smoothing affects correlation on the response, for this reason I remove redundant variables and arrive at a different subset. Air density has been removed from the dataset but air temperature, air pressure, water temperature have entered. The removal of air density could indicate that due to the increased noise, granular features provide a better measure of noise. Here air pressure appears to be variable with the highest linear correlation, in a direction that agrees with the inverse barometric effect. In quarterly frequency our heat content variable had little linear correlation with respect to our response, but the correlation has increased significantly indicating a higher linear correlation in the monthly frequency. This change showcases that variable importance can change in RSLR depending on the frequency.

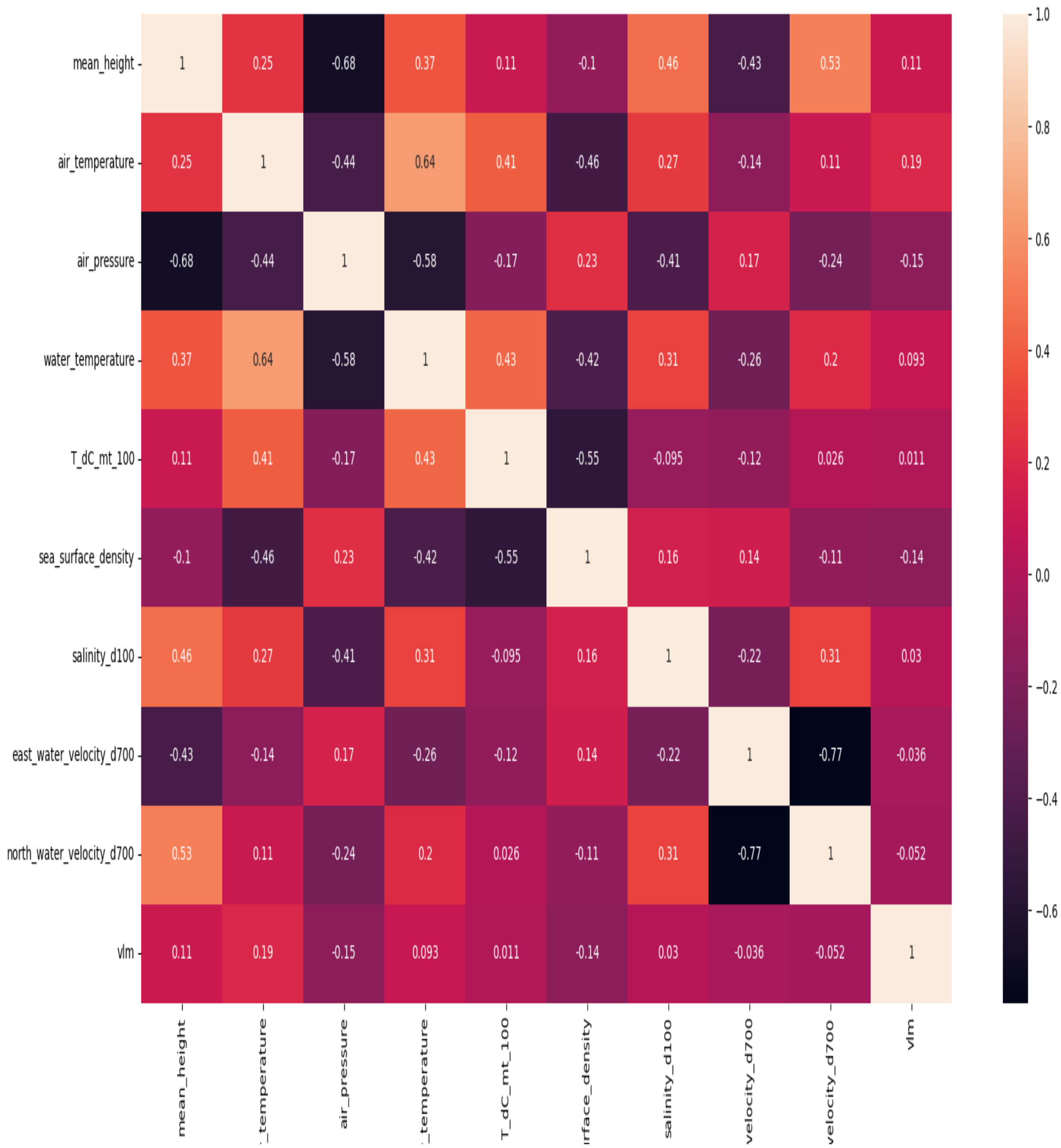


Figure 8: Monthly correlation analysis results

5.2.4 Modeling

Again overfitting an ARIMAX model with every possible variable we see in Table 2 that scores are worse than its auto SARIMAX counter part and LSTM competitors. This time the RMSE test between the LSTM models and SARIMAX are not as wide. In fact, our LSTM with all variables does worse than SARIMAX and LSTM with subset data.

Table 2: Monthly modeling results in millimeters

Model	VLM	RMSE train	RMSE test
ARIMAX All Vars	Yes	830	80.5
SARIMAX Auto	No	41.4	33.5
LSTM All Vars	Yes	28.3	34.8
LSTM Subset	Yes	25.9	28
LSTM Subset	No	25	29.8

The best performing LSTM model was a model with 50 units and 2 layers. There are several possible reasons as to why the LSTM doesn't do a significantly better job when measuring RMSE. The large variations that occur are likely the reason, RMSE is more sensitive to outliers and the LSTM appeared to capture the general trend and didn't appear to overfit for the spikes presented in monthly data. Our overfit LSTM does worse than the traditional time series model and this could be due to several reasons. For instance, the additional variables may contribute irrelevant information, the dimensionality makes it harder to find relevant patterns, and the collinearity can make it difficult to establish which variables are most important. Simply reducing the collinearities and subsequently the dimensionality showcases that these are all possible scenarios. The LSTM with the reduced subset does at least 5mm better in our testing case in comparison to any other model. Allowing it not only to have less training error but prediction error as well. Showcasing the importance of nonlinearities for monthly frequency and performing correlation analysis. Lastly removing VLM reduced the training error but increases the training error, showing that VLM has measurable impact on SLR in this region.

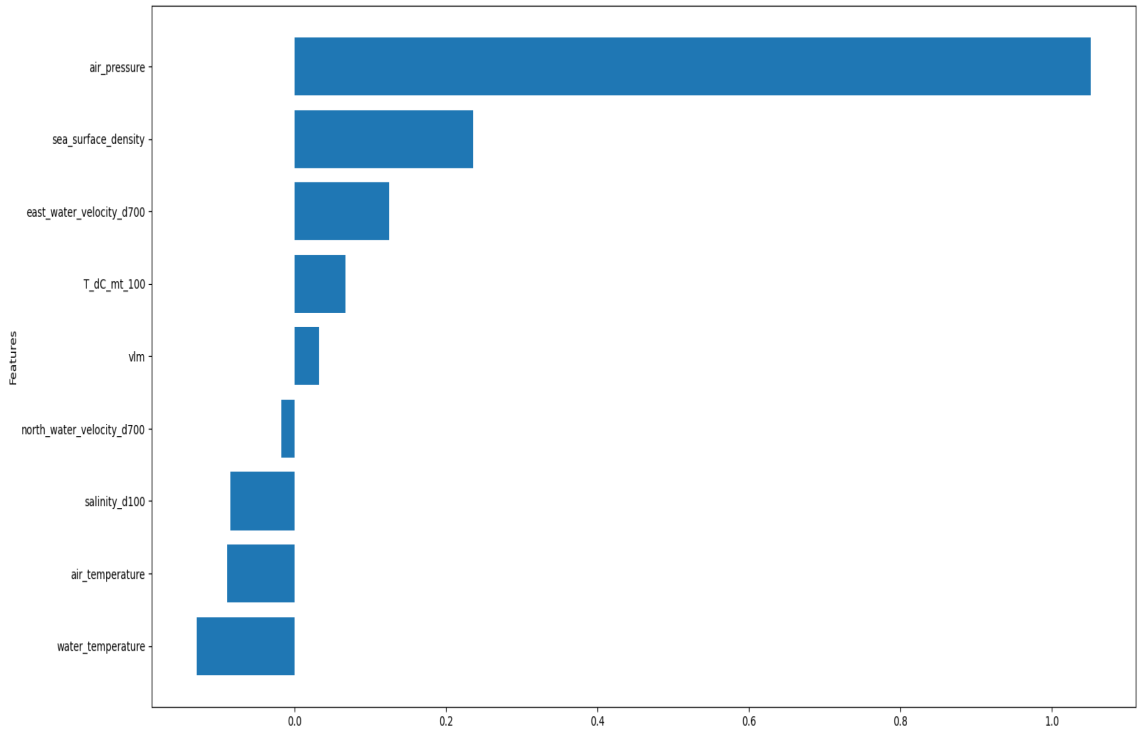


Figure 9: Monthly SHAP results

Air pressure has the largest positive contribution to SLR monthly variation while its fellow component of air density, air temperature, is the second largest negative contribution. In both frequencies air pressure has a positive impact on our response, counter-intuitive to the inverse-barometer effect. However, it's found that air pressure is generally highest in the winter [38], the time Monterey bay experiences more tide variation and increased storms which influence RSLR. East water velocity, heat content, and VLM all have positive contributions which was expanded upon in the quarterly analysis. Oddly sea surface density increases SLR, counter intuitive to thermal expansion. However, this could indicate that atmospheric variables play a more significant part in shaping RSLR in Monterey bay. VLM's small positive importance indicates that oceanic and atmospheric variables dominate RSLR in this region but VLM has some importance in a direction that agrees with quarterly importance.

5.3 Analysis

The subset variables indicated driving SLR factors in the Monterey Bay region. Wind was a prominent predictor in prior research, but doesn't appear to be an important predictor with correlation analysis in either frequency; this is likely due to the channeling caused by the bay. Depth at 100 appears to be the strongest predictor in comparison to other heat and oceanic depth variables, intuitively this occurs as the shallow depth reflects more energy and does not account for the slowly changing heat content at the bottom of the ocean. Overall, the main drivers at the Monterey Bay station are atmospheric and oceanic. Atmospheric variables such as air pressure, temperature, and air density play an important role in both frequencies. Behind those in importance are oceanic variables which mainly influence Monterey Bay with oceanic currents, and salinity of the ocean. Lastly our coastal variable, VLM has shown importance in driving more accurate predictions.

While the importance given to VLM by SHAP is small, this likely reflects the slow change in VLM. VLM is affected by GIA and non-GIA processes but considering that non-GIA processes drive a large portion of seasonal VLM changes it's important to measure VLM to capture processes that traditional SLR variables cannot. Such changes to the water cycle in a region due aquifer pumping, reloading, land usage, land loading due to weather, and even plate drift. Therefore the small importance is likely attributed to VLM's natural behavior but due to its extrinsic factors it drive us to more accurate predictions. Regardless of the reason, this project has shown that

VLM has ability to explain changes in SLR in both quarterly and monthly frequencies but shows the most promise with quarters. For this reason, VLM's attribution should be explored in regions that typically more vulnerable to flooding such as Malaysia.

CHAPTER 6

Plausibility of VLM

VLM showed ability to increase generalization accuracy in the quarterly frequency and gave a minor performance boost in the Monthly Frequency. With this knowledge future models could incorporate this depending on their accuracy needs. However, GNSS VLM data has typically been difficult to attain due to data quality issues. This study used the mean of four GNSS stations with at least 95% complete data from 2008-2023 which was interpolated linearly. Oftentimes data sources are going to be more sparse with frequent gaps and gaps that span years, often deriving from equipment malfunction. In this chapter we cover why individual stations were used and cover several interpolation methods at various frequencies to determine which methods is best for VLM.

6.1 Historical VLM

While we have covered driving sources of VLM in Chapter 2, we have not covered VLM collection, why we use individual stations and cons associated with it.

6.1.1 VLM Collection

As with every data source, VLM needs to be collected and historically this has been accomplished by using ground-based methods such as levelling and GNSS. Levelling is typically a time consuming process that involves measuring the height of the ground at multiple points, often using specialised equipment. Our GNSS data uses satellites to transmits a signal that can be picked up by a receiver, the ground station, by measuring the time delay between transistor and receiver we can calculate the distance from the satellite and relative elevation. Of course this comes with measuring error, for example [39] satellite clocks alone can cause 2m of error to our VLM measurements! Despite this, many methods exist to correct errors which can bring estimated VLM errors to as low as 5mm, which is what our data source Nevada

Geodetic Laboratory [40] accomplishes. While there are newer forms of VLM tracking such as analyzing stereoscopic pairs of satellite optical images and processing of SAR images but this study couldn't find these datasets with the right temporal frame or were not available for the Monterey Bay region.

6.1.2 Station Selection

While error might be low for tracking VLM with GNSS data, there is an spatial issue with using GNSS. GNSS receivers describe VLM at that specific station using them to describe a larger spatial region usually provides an over-simplified representation of the actual pattern of the land displacements due to the limited number of monitoring locations. Therefore, finding better sources of VLM data could provide better and more accurate results.

In addition, this study used the mean of four stations along the coast of Monterey Bay and excludes VLM data in the mountains which shows VLM trends opposite to those on the coast. Being able to accurately interpolate for stations which didn't have at least 95% of its data intact could be a step in the right direction for fixing the speciality. Likewise, tracking the change in the mountains could unveil interactions with the land below and SLR.

6.2 Interpolation

With the goal of increasing the number of usable GNSS stations we look at several methods of interpolation and assess which is best. To determine the best we assessing RMSE, MAE, and observe the graphs of interpolated data for our two frequencies.

The methods are simple, we use linear interpolation, SARIMA, and AKIMA. The metrics in table 3 below are produced from randomly subsetting our tide gauge station, as it is the most complete, for 50 iterations and averaging the interpolation results. The results are for two different temporal frames, 4 quarters and 10 quarters. The results are clear, linear does best according to metrics and a model such as ARIMA with parameters (1,0,1)(1,0,1) don't do as well. However, one interesting note is the graphs that accompany the results.

Table 3: Quarterly Interpolation Results

	4 Quarters		10 Quarters	
Method	RMSE	MAE	RMSE	MAE
Linear	2.76	9.12	3.34	12.45
SARIMA	3.24	12.86	3.65	16.16
AKIMA	2.97	13.12	3.6	14.59

The graphs in discussion can be found in Appendix A.10. In good scenarios SARIMA appears to capture seasonality well, and this is important because VLM's attribution to accurate predictions are likely due to non-Seasonal GIA changes. Thus, maintaining seasonality in long term interpolations are likely beneficial and nudge the use of use SARIMA for individual stations. However, short term predictions of a few time steps are better served by linear interpolation in the quarterly case.

Table 4: Monthly Interpolation Results

	6 Months		24 Months	
Method	RMSE	MAE	RMSE	MAE
Linear	2.67	8.6	3.65	15.12
SARIMA	2.93	10.64	4.21	19.72
AKIMA	2.75	9.12	5.45	37.53

Table 4 shows similar performance results as those in the Quarterly case. That is linear outperform our other interpolation metrics but it is burdened with seasonality preservation. The graphs are in Appendix A.11 and showcase a similar intuition. For short term periods of missing data, linear does just fine as seasonality is not as prominent. However, for longer periods of time, linear interpolation doesn't preserve seasonality. Therefore SARIMA should be used in long term cases, while linear should be used for short term cases.

CHAPTER 7

Conclusion

RSLR is important to predict because maritime commerce, industry, population, and agriculture depend on safe water levels. Accurately predicting and understanding RSLR can help us determine water levels for incoming maritime vessels, better assess flooding, and help understand what drives local tides. Vertical Land Motion (VLM) is typically ignored when predicting RSLR for oceanic and atmospheric variables which have been recorded for longer temporal periods and at reliable frequency. Monterey Bay, an area where reliable VLM measuring devices are available, was used to observe VLM's affect on SLR in this project. Using a large set of atmospheric and oceanic variable that historically impact SLR, VLM's importance is studied by using AI models and SHAP values. While atmospheric and oceanic variables largely describe RSLR in Monterey Bay, VLM which displayed little linear correlation and has more importance when used by an LSTM and analyzed using SHAP values. Modeling metrics show that VLM can narrow residual error in training and when forecasting unknown values in the long term. Seasonal non-GIA processes such aquifer recharge and seasonal land loading are intuitive reasons to VLM's ability to capture long seasonal trends in VLM. More so, we have showed that while using GNSS VLM data we can interpolate accurately in the short gaps in data using linear interpolation and for gaps as far as 2 years with SARIMA in the monthly frequency. This interpolation allows us increase the usable number of GNSS stations for VLM while preserving seasonality.

LIST OF REFERENCES

[1] L. Burke, Y. Kura, *et al.*, *Coastal ecosystems.* World Resources Institute Washington, DC, 2001.

[2] A. Melet, P. Teatini, *et al.*, ‘‘Earth observations for monitoring marine coastal hazards and their drivers,’’ *Surveys in Geophysics*, 2020.

[3] G. McGranahan, D. Balk, and B. Anderson, ‘‘The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones,’’ *Environment and Urbanization*, vol. 19, no. 1, pp. 17--37, 2007. [Online]. Available: https://doi.org/10.1177/0956247807076960

[4] A. Cazenave and W. Llovel, ‘‘Contemporary sea level rise,’’ *Annual review of marine science*, vol. 2, no. 1, pp. 145--173, 2010.

[5] N. Mimura, ‘‘Sea-level rise caused by climate change and its implications for society,’’ *Proceedings of the Japan Academy. Series B, Physical and biological sciences*, 2014.

[6] G. Wöppelmann and M. Marcos, ‘‘Vertical land motion as a key to understanding sea level change and variability,’’ *Reviews of Geophysics*, vol. 54, no. 1, pp. 64--92, 2016. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015RG000502

[7] C. Zervas, S. Gill, and W. Sweet, ‘‘Estimating vertical land motion from long-term tide gauge records,’’ 2013. [Online]. Available: https://repository.library.noaa.gov/view/noaa/14751

[8] W. C. Hammond, G. Blewitt, *et al.*, ‘‘Gps imaging of global vertical land motion for studies of sea level rise,’’ *Journal of Geophysical Research: Solid Earth*, vol. 126, no. 7, p. e2021JB022355, 2021, e2021JB022355 2021JB022355. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JB022355

[9] S. Sithara, S. K. Pramada, and S. G. Thampi, ‘‘Sea level prediction using climatic variables: a comparative study of svm and hybrid wavelet svm approaches,’’ *Acta Geophysica*, vol. 68, no. 6, pp. 1779--1790, 2020. [Online]. Available: https://link.springer.com/article/10.1007/s11600-020-00484-3

[10] A.-L. Balogun and N. Adebisi, ‘‘Sea level prediction using arima, svr and lstm neural network: assessing the impact of ensemble ocean-atmospheric processes on models' accuracy,’’ *Geomatics, Natural Hazards and Risk*, 2021.

[11] A. Altunkaynak and E. Kartal, ‘‘Transfer sea level learning in the bosphorus strait by wavelet based machine learning methods,’’ *Ocean Engineering*, vol. 233, 05 2021.

[12] W. Huang, C. Murray, *et al.*, ‘‘Development of a regional neural network for coastal water level predictions,’’ *Ocean Engineering*, vol. 30, no. 17, pp. 2275--2295, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801803000830

[13] V. Nieves, C. Radin, and G. Camps-Valls, ‘‘Predicting regional coastal sea level changes with machine learning,’’ *Scientific Reports*, vol. 11, 2021.

[14] B. Hamlington, D. Chambers, *et al.*, ‘‘Observation-based trajectory of future sea level for the coastal united states tracks near high-end model projections,’’ *Communications Earth Environment*, vol. 3, p. 230, 10 2022.

[15] B. Harrison and et. al, ‘‘Future sea-level rise projections for tide gauge locations in south asia,’’ *Environmental Research Communications*, vol. 3, 10 2021.

[16] K. Poppe and J. Rybczyk, ‘‘Assessing the future of an intertidal seagrass meadow in response to sea level rise with a hybrid ecogeomorphic model of elevation change,’’ *Ecological Modelling*, vol. 469, p. 109975, 04 2022.

[17] J. A. Carton and B. S. Giese, ‘‘A reanalysis of ocean climate using simple ocean data assimilation (soda),’’ *Monthly Weather Review*, vol. 136, no. 8, pp. 2999 -- 3017, 2008. [Online]. Available: https://journals.ametsoc.org/view/journals/mwre/136/8/2007mwr1978.1.xml

[18] A. Köhl and D. Stammer, ‘‘Decadal sea level changes in the 50-year gecco ocean synthesis,’’ *Journal of Climate*, vol. 21, no. 9, pp. 1876 -- 1890, 2008. [Online]. Available: https://journals.ametsoc.org/view/journals/clim/21/9/2007jcli2081.1.xml

[19] J. M. Price and L. Magaard, ‘‘Interannual baroclinic rossby waves in the midlatitude north atlantic,’’ *Journal of Physical Oceanography*, vol. 16, no. 12, pp. 2061 -- 2070, 1986. [Online]. Available: https://journals.ametsoc.org/view/journals/phoc/16/12/1520-0485_1986_016_2061_ibrwit_2_0_co_2.xml

[20] B. D. Hamlington, A. S. Gardner, *et al.*, ‘‘Understanding of contemporary regional sea-level change and the implications for the future,’’ *Reviews of Geophysics*, vol. 58, no. 3, p. e2019RG000672, 2020, e2019RG000672 2019RG000672. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019RG000672

[21] P. D. Bromirski, A. J. Miller, *et al.*, ‘‘Dynamical suppression of sea level rise along the pacific coast of north america: Indications for imminent acceleration,’’ *Journal of Geophysical Research: Oceans*, vol. 116, no. C7, 2011. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JC006759

[22] M. A. Merrifield and M. E. Maltrud, ‘‘Regional sea level trends due to a pacific trade wind intensification,’’ *Geophysical Research Letters*, vol. 38, no. 21, 2011. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011GL049576

[23] A. Vecchio, M. Anzidei, *et al.*, ‘‘Natural variability and vertical land motion contributions in the mediterranean sea-level records over the last two centuries and projections for 2100,’’ *Water*, vol. 11, no. 7, 2019. [Online]. Available: https://www.mdpi.com/2073-4441/11/7/1480

[24] D. Chelton and R. E. Davis, ‘‘Monthly mean sea-level variability along the west coast of north america,’’ *Journal of Physical Oceanography*, vol. 12, pp. 757--784, 1982. [Online]. Available: https://api.semanticscholar.org/CorpusID:128745976

[25] K. M. Zubier and L. S. Eyouni, ‘‘Investigating the role of atmospheric variables on sea level variations in the eastern central red sea using an artificial neural network approach,’’ *Oceanologia*, vol. 62, no. 3, pp. 267--290, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0078323420300087

[26] W. Sturges and B. C. Douglas, ‘‘Wind effects on estimates of sea level rise,’’ *Journal of Geophysical Research: Oceans*, vol. 116, no. C6, 2011. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JC006492

[27] T. O. Muslim, A. N. Ahmed, *et al.*, ‘‘Investigating the influence of meteorological parameters on the accuracy of sea-level prediction models in sabah, malaysia,’’ *Sustainability*, vol. 12, no. 3, 2020. [Online]. Available: https://www.mdpi.com/2071-1050/12/3/1193

[28] A. J. Clarke and A. Lebedev, ‘‘Remotely driven decadal and longer changes in the coastal pacific waters of the americas,’’ *Journal of Physical Oceanography*, vol. 29, no. 4, pp. 828 -- 835, 1999. [Online]. Available: https://journals.ametsoc.org/view/journals/phoc/29/4/1520-0485_1999_029_0828_rddalc_2.0.co_2.xml

[29] E. KO, ‘‘Relative sea levels from tide-gauge records.’’ *Proceedings of the National Academy of Sciences of the United States of America*, vol. 77, p. 6968–6972, 1980.

[30] V. Gornitz, S. Lebedeff, and J. Hansen, ‘‘Global sea level trend in the past century,’’ *Science*, vol. 215, pp. 1611--1614, 1982.

[31] G. Wöppelmann and M. Marcos, ''Vertical land motion as a key to understanding sea level change and variability,'' *Reviews of Geophysics*, vol. 54, no. 1, pp. 64--92, 2016. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015RG000502

[32] C. Zervas, S. Gill, and W. Sweet, ''Estimating vertical land motion from long-term tide gauge records,'' Center for Operational Oceanographic Products and Services, Tech. Rep., 2013. [Online]. Available: https://repository.library.noaa.gov/view/noaa/14751

[33] R. C. Holleman and M. T. Stacey, ''Coupling of sea level rise, tidal amplification, and inundation,'' *Journal of Physical Oceanography*, vol. 44, no. 5, pp. 1439 -- 1455, 2014. [Online]. Available: https://journals.ametsoc.org/view/journals/phoc/44/5/jpo-d-13-0214.1.xml

[34] I. Pardoe, *Applied Regression Modeling*, 2020.

[35] R. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice,*, 2nd ed. OTexts: Melbourne, Australia, 2018.

[36] A. Geron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. O'Reilly Media, Inc., 2019.

[37] C. Molnar, *Interpretable Machine Learning*, 1st ed., 2022. [Online]. Available: https://christophm.github.io/interpretable-ml-book

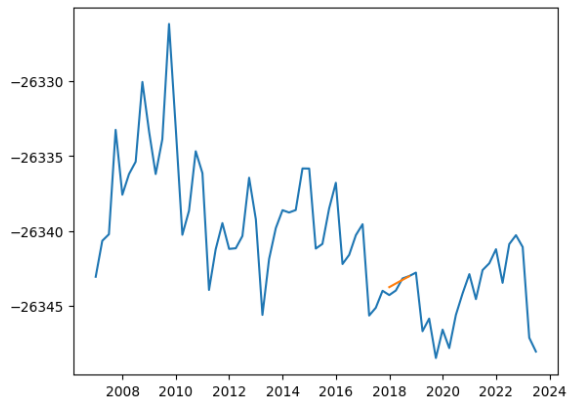
[38] N. Oceanic and A. Administration. ''Air pressure.'' [Online]. Available: https://www.noaa.gov/jetstream/atmosphere/air-pressure

[39] N. Inc., *An Introduction to GNSS: A primer in using Global Navigation Satellite Systems for positioning and autonomy*, 3rd ed. NovAtel Inc., 2010.

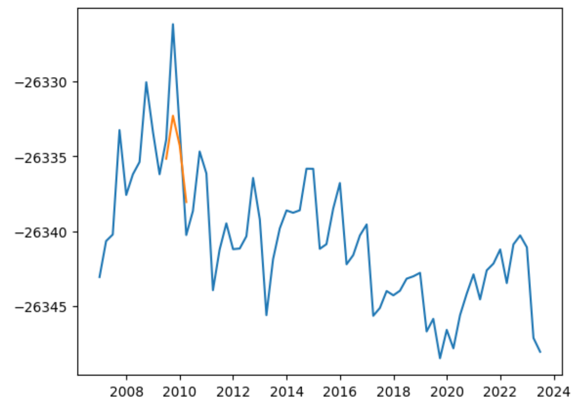
[40] G. Blewitt, W. Hammond, and C. Kreemer, ''Harnessing the gps data explosion for interdisciplinary science,'' Tech. Rep., 2018.

APPENDIX

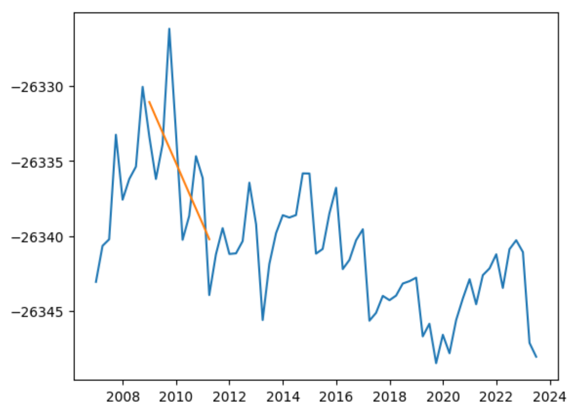
VLM Interpolation Plots



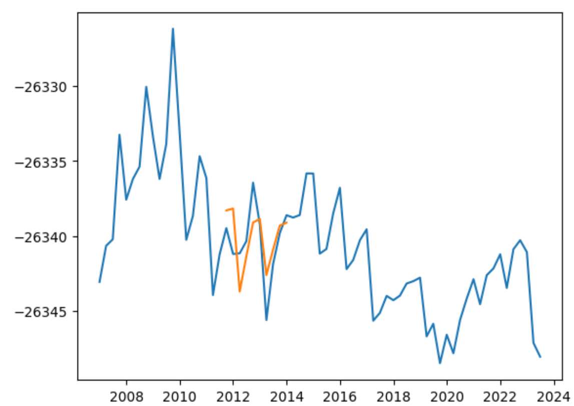
(a) 4 quarter linear interpolation



(b) 4 quarter SARIMA interpolation

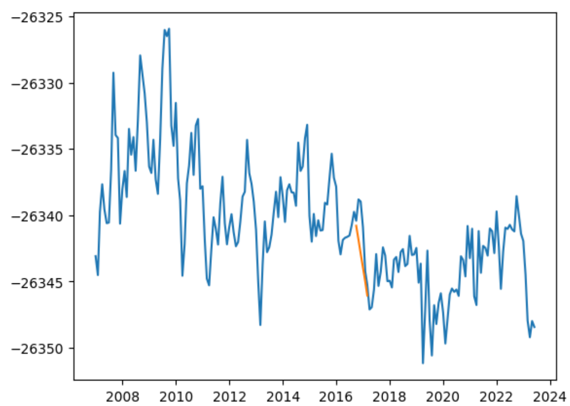


(c) 10 quarter linear interpolation

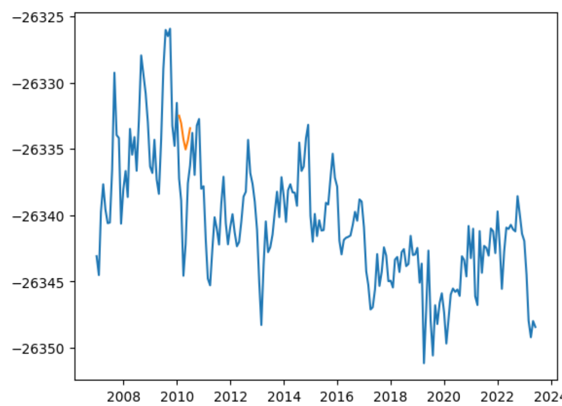


(d) 10 quarter SARIMA interpolation

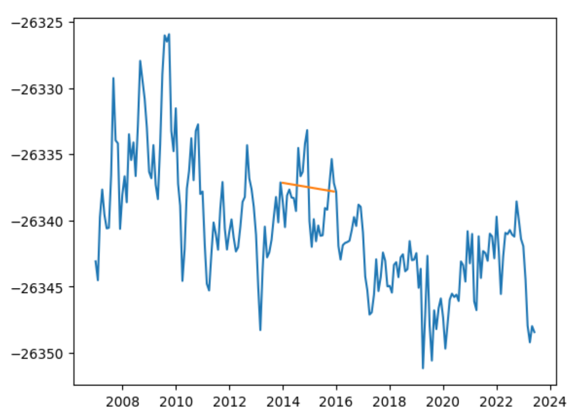
Figure A.10: Quarterly Linear and SARIMA interpolation results



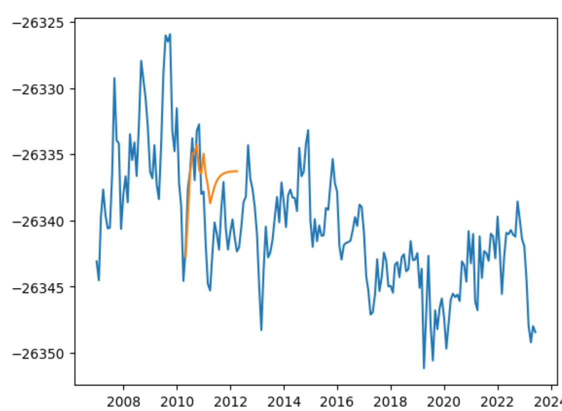
(a) 6 month linear interpolation



(b) 6 month SARIMA interpolation



(c) 24 month linear interpolation



(d) 24 month SARIMA interpolation

Figure A.11: Monthly Linear and SARIMA interpolation results