

Predicting Saltwater Intrusion In Coastal Aquifers By Data-Driven Modeling

A Major Thesis Report submitted in partial fulfillment of the degree of

MASTER of TECHNOLOGY
in
GEOINFORMATICS AND EARTH OBSERVATION

Submitted by

Pralay Sankar Maitra
AM.EN.P2GEO21015



**AMRITA CENTER FOR WIRELESS NETWORKS AND
APPLICATIONS**

AMRITA VISHWA VIDYAPEETHAM ,
(Estd. U/S 3 of the UGC Act 1956) Amritapuri Campus
Kollam - 690525

June 2023

**AMRITA CENTER FOR WIRELESS NETWORKS AND
APPLICATIONS**

**AMRITA VISHWA VIDYAPEETHAM ,
(Estd. U/S 3 of the UGC Act 1956) Amritapuri Campus
Kollam -690525**



BONAFIDE CERTIFICATE

This is to certify that the Major Thesis Report entitled "**Predicting Salt-water Intrusion in Coastal Aquifers by Data-Driven Modeling**" submitted by Mr Pralay Sankar Maitra (AM.EN.P2GEO21015), in partial fulfilment of the requirements for the award of Degree of Master of Technology in GeoInformatics and Earth Observation, is a bonafide record of the work carried out by him under my guidance and supervision at Amrita School of Engineering, Amritapuri during Semester 4 of the academic year 2021-2023.

Dr. Dipankar Dwivedi
Thesis Adviser

Dr. Maneesha V. Ramesh
Head of the Department

Dr. Martin Julian Montag
Thesis Co-Adviser

Examiner 1

Examiner 2

Place : Amritapuri
Date : June , 2023

**AMRITA CENTER FOR WIRELESS NETWORKS AND
APPLICATIONS**

**AMRITA VISHWA VIDYAPEETHAM ,
(Estd. U/S 3 of the UGC Act 1956) Amritapuri Campus
Kollam -690525**



DECLARATION

I, Mr. Pralay Sankar Maitra, Reg No: AM.EN.P2GEO21015, hereby declare that this project entitled “Predicting Saltwater Intrusion in Coastal Aquifers by Data-Driven Modeling” is a record of the original work done by me under the guidance of Dr. Dipankar Dwivedi, Department of Earth and Environmental Sciences, Lawrence Berkeley National Laboratory and Dr. Martin Julian Montag, Department of Wireless Networks And Applications, Amrita Vishwa Vidyapeetham, that this work has not formed the basis for any degree/diploma/association ship/fellowship or similar awards to any candidate in any university to the best of my knowledge.

Place : Amritapuri

Date : June 30, 2023

Signature of the Student

Signature of the Thesis Adviser



I would like to dedicate this thesis to AMMA, Sri Mata
Amritanandamayi Devi and my loving mother.

”Surrender and living in the present are one and the same ” - Amma

Acknowledgements

"Be like the honeybee, who gathers only nectar wherever it goes. Seek the goodness that is found in everyone" — Amma

First, I offer my sincere gratitude to the God Almighty whose abundant grace and mercy enabled me to successfully complete my major thesis.

My mother has always inspired me and given me full freedom to choose the path I decided upon. I owe my deepest gratitude to their unflagging support and care.

I would also like to express my sincere thanks to my advisors, Dr. Dipankar Dwivedi and Dr. Martin Julian Montag. They have been exuberant in their ways and have always corrected me, guided me, shown me the truth and forced me to come up with impending solutions to problems I faced while completing this project.

I wish to express my sincere thanks to our Head of the Department (Wireless Networks and Applications), Dr. Maneesha V. Ramesh for giving me the support and encouragement that was necessary for the completion of this project.

I also wish my deep sense of gratitude to Project Coordinators Prof. K. A. Unnikrishnan Menon, Mrs. Meenu L of Wireless Networks and Applications for their constant support and valuable guidance in carrying out this major thesis. I also thank my co-advisor Dr. Martin Julian Montag and Dr. Alka Singh for their guidance and support throughout the period of my study and their presence was overall felt during the course of this study.

I would also like to express my gratitude to all the faculty members and staff who have helped in my project in one way or another.

Last but not least I would like to thank our Beloved Chancellor, Sri (Dr.) Mata Amritanandhamayi Devi for providing me with guidance and a friendly environment.

I also wish to thank my internal reviewers like Dr. Shivaprata for their priceless contribution to this project. Without their guidance this project would not have been successfully completed in time.

Abstract

Saltwater intrusion in coastal aquifers has become an increasing concern globally, due to both anthropogenic influence and climate change. Saltwater intrusion negatively impacts water quality and freshwater availability in coastal areas, and thus presents substantial problems to coastal communities. Mitigation efforts require predictions of saltwater intrusions at different sites. We develop ML models for the prediction of variables related to saltwater intrusion at two sites on the east coast of the US. Saltwater intrusion is primarily caused by climate change, directly via rising sea levels, and indirectly via changing precipitation patterns, and a rise in the frequency of extreme weather events. As sea levels rise, saltwater seeps into freshwater aquifers, limiting the amount of potable water available and deteriorating its quality. The rate at which freshwater aquifers recharge is also impacted by changing precipitation patterns, which worsens saltwater intrusion.

Saltwater intrusion affects communities and ecosystems in several different ways: First, the intrusion jeopardizes freshwater supplies, necessary for domestic and industrial use. Second, groundwater with increasing salinity becomes unusable for irrigation or degrades the soil, lowering crop yields. Thirdly, the intrusion affects flora, fauna of ecosystems and reduces biodiversity in coastal areas. Last but not least, infrastructure threats from saline water in freshwater aquifers include foundation damage and pipe corrosion. To predict future vulnerability of coastal aquifers, several input factors are relevant. In addition to time-invariant site specific factors and inputs available from global climate models, one important model input is the water-level in the coastal aquifer.

As future change of water-levels in coastal aquifers are itself unknown, we follow a two step approach in modeling: In step 1, we predict the groundwater level from precipitation, temperature and tidal level using a SARIMAX (Seasonal Auto-Regressive Integrated Moving Average eXogenous) model. In step 2 we add the predicted water-level

to the input variables and use a Facebook Prophet model for specific conductivity prediction. As both models are entirely data-driven, they can be transferred to a large number of coastal sites and serve as starting model wherever past time series of precipitation, temperature and tidal level data are available. We thus contribute to vulnerability assessment of coastal aquifers that enables the planning and prioritization of mitigation efforts, and thus forms an important step in the protection of coastal communities.

Contents

1	Introduction	1
1.1	Research Motivation and Challenges	1
1.2	Problem Definition	1
1.3	Objectives	2
1.4	Ghyben-Herzberg Relation	2
1.5	Literature Review	3
1.5.1	Susceptibility mapping of groundwater salinity using machine learning models [1]	3
1.5.2	A comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer [2]	3
1.5.3	Advance prediction of coastal groundwater levels with temporal convolutional and long short-term memory networks [3]	4
1.5.4	Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas [4]	4
2	Data Collection from various sources	5
2.1	Data Collection	5

CONTENTS

2.2	Data Visualisation	7
2.3	Conclusion	8
3	Introduction To ARIMA and Prophet	12
3.1	Auto-ARIMA	12
3.2	ARIMA	12
3.3	SARIMAX model	14
3.4	Data Resampling for Auto-ARIMA	15
3.5	Facebook Prophet	15
3.6	SARIMAX and Facebook Prophet: A Comparative Study	18
4	Numerical Modeling - Results and Discussions	23
4.1	GW level forecasting using SARIMA	23
4.2	GW level prediction using SARIMAX	25
4.3	GW Specific Conductivity (SC) prediction using Facebook Prophet	26
5	Conclusion	35
	Publications	36
	Appendix A	43
	Appendix B	44
	Appendix C	49
	Appendix D	52

CONTENTS

Appendix E	61
-------------------	-----------

List of Figures

1.1	Saltwater Intrusion [5]	3
2.1	NOAA Tidal Stations near GW Sites	6
2.2	Tidal levels for site 1	7
2.3	Tidal levels for site 2	7
2.4	Precipitation for site 1	8
2.5	Precipitation for site 2	8
2.6	Temperature for site 1	9
2.7	Temperature for site 2	9
2.8	Groundwater levels for site 1	10
2.9	Groundwater levels for site 2	10
2.10	Groundwater Specific Conductivity for site 1	11
2.11	Groundwater Specific Conductivity for site 2	11
3.1	Site 1 re-sampled monthly data	21
3.2	Site 2 re-sampled monthly data	22
4.1	Site 1 Forecasting of GW levels	24

LIST OF FIGURES

4.2	Site 2 Forecasting of GW levels	24
4.3	Site 1 GW Level Predictions	25
4.4	Site 2 GW Level Predictions	26
4.5	Model diagnostics for Site 1 GW Level	27
4.6	Model summary for Site 1 GW Level	28
4.7	Model diagnostics for Site 2 GW Level	29
4.8	Model summary for Site 2 GW Level	30
4.9	Fitting of GW SC data (Site - 1) into the prophet model.	31
4.10	Ovservation vs prediction of test data set for Site 1 GW SC.	32
4.11	Fitting of GW SC data (Site - 2) into the prophet model.	33
4.12	Ovservation vs prediction of test data set for Site 2 GW SC.	34

Chapter 1

Introduction

1.1 Research Motivation and Challenges

Saltwater intrusion is a major problem in coastal areas where well pumping is a major source of freshwater [6]. The difference in density combined with the relative changes between the mean sea level and the groundwater table leads to the incoming of denser saltwater from the sea into the groundwater table of the mainland. This causes major problems in the overall ecosystem both biological and physical and basically disturbs the balance between nature and artificial constructs. Hence, it needs to be studied so as to take proper steps to mitigate or cease the whole process before it actually affects the rich in minerals freshwater found in the aquifers which have undergone so many filtration steps to finally become the freshwater it is.

Coastal aquifers are dynamic because of compounding events like storm surges etc. It is not easy to predict Groundwater level and salinity evolution. We will address these using data-driven models and analysis of various variable datasets like temperature, precipitation and surface water levels.

1.2 Problem Definition

The problem definition in our major thesis is to take 3 watersheds and 3 groundwater sites and find a correlation between the boundary conditions and

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

internal conditions of the well. The watershed/boundary conditions include temperature, precipitation and surface water levels. The internal conditions include GW levels and GW salinity measured by specific conductivity.

Seawater intrusion can harm groundwater quality in a variety of places, both coastal and inland, throughout the USA. Along the coast, seawater intrusion into aquifers is connected to the over-drafting of groundwater. Additionally, groundwater pumping can draw up salty water from seawater isolated in subsurface sediments. As pumps draw up fresh water from the aquifer, seawater flows inland into the aquifer by the force of the ocean. The flow of salty water can in turn create water quality problems. A chief concern for the future is sea level rise as a result of climate change and the discerning effects of GW usage increase over the years as more and more people rely on GW for their daily activities.

1.3 Objectives

- To produce a set of data for building a model.
- To model the downloaded and cleaned data using the SARIMAX approach.
- To validate the model using specific parameters like RMSE, and R2 score.

1.4 Ghyben-Herzberg Relation

The Ghyben-Herzberg relation gives a relationship between the groundwater level and the mean sea level. It is written as $z=40h$, where z and h parameters are as given in the following figure 1.1. z is the distance between saltwater-freshwater interface and the mean sea level. h is the distance between groundwater table level and the mean sea level.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

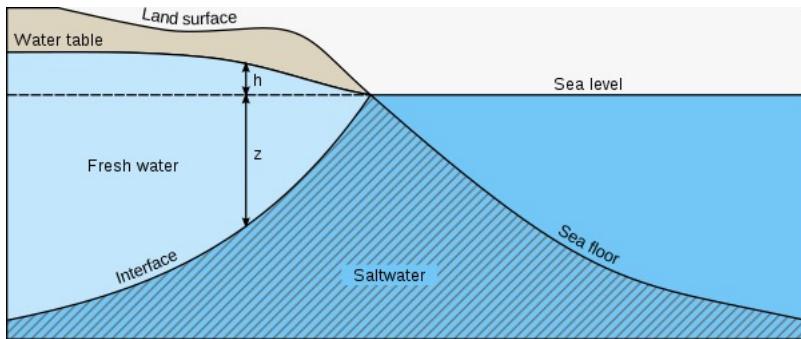


Figure 1.1: Saltwater Intrusion [5]

1.5 Literature Review

1.5.1 Susceptibility mapping of groundwater salinity using machine learning models [1]

Increasing groundwater salinity is a major problem these days. This happens due to climate change and mean sea level rise. GW has many minerals of which salinity must be balanced as GW provides freshwater which is used on a day to day basis. In this paper the authors have predicted groundwater salinity from various input features like soil order, GW withdrawal, precipitation, land use and elevation using six type of machine learning models of which support vector machine worked the best. Results highlighted that the southern parts of the region and some parts in the north, northeast, and west have a high groundwater salinity.

1.5.2 A comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer [2]

Groundwater salinity in an aquifer system is typically measured through field studies (e.g., groundwater sampling, and direct current resistivity method). In this paper hydrogeology and hydrometeorology data were collected and the dataset was divided into 3 parts - training, testing and verification. Thre models were used to predict GW salinity from the input features - extreme gradient boost-

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

ing (EGB), deep neural network (DNN) and multiple linear regression (MLR). The models were evaluated using R-squared, Nash-Sutcliffe efficiency (NSE) and normalized root-mean square deviation (NRMSD). Results showed that consideration should be given to the EGB method, considering its higher performance on the testing subset (R-squared 0.89, NSE 0.87, NRMSD 0.45). An in-depth analysis of the variables showed that aquifer transmissivity is the most crucial parameter affecting groundwater salinity in the region. The adopted approach could potentially be used for groundwater management purposes in the study area and similar settings elsewhere.

1.5.3 Advance prediction of coastal groundwater levels with temporal convolutional and long short-term memory networks [3]

Prediction of groundwater level is of immense importance and challenges coastal aquifer management with rapidly increasing climatic change. With the development of artificial intelligence, data-driven models have been widely adopted in hydrological process management. However, due to the limitation of network framework and construction, they are mostly adopted to produce only 1-time step in advance. Here, the temporal convolutional network (TCN) and models based on long short-term memory (LSTM) were developed to predict groundwater levels with different leading periods in a coastal aquifer.

1.5.4 Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas [4]

Groundwater level prediction is of major importance. In this paper the authors have predicted GW levels from input features like irrigation, precipitation, temperature and evapotranspiration. In this study, the proposed model was applied and evaluated in five sub-areas of Hetao Irrigation District in arid northwestern China using data of 14 years (2000–2013). 14 years of data are separated into two sets: training set (2000–2011) and validation set (2012–2013) in the experiment.

Chapter 2

Data Collection from various sources

2.1 Data Collection

We collected daily meteorological (DAYMET) data from a single pixel extraction tool in the DAYMET website <https://daymet.ornl.gov/single-pixel/>. We got daily data from this website as the name suggests and we took the data for all the 4 GW sites.

From National Oceanic and Atmospheric Administration (NOAA), we received an API which could get us the site's package and its underlying data and information regarding multiple parametric categories. So, we downloaded the station information (ID, names, latitudes, longitudes), plotted it in ArcGIS pro and selected those sites only which were close to each of the groundwater sites we selected. There was a separate API for the information we wanted (i.e. water level) for all the 3 selected sites which was used to download the hourly height of water for the sites. Then these hourly spaced data were averaged for a day and a daily dataset was formed.

We collected the GW information (levels and quality) from USGS using dataRetrieval package in R using site information and parameter codes, 00065 for gage height/levels and 00095 for specific conductivity or GW quality. The code for this data retrieval is given in Appendix F.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

Site No	GW Site ID	NOAA Site ID	DAYMET data
Site-1	300722089150100	8747437	✓
Site-2	301001089442600	8761927	✓

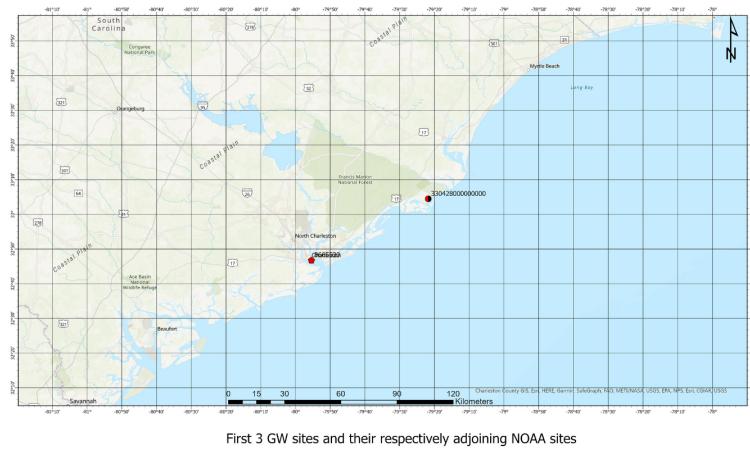
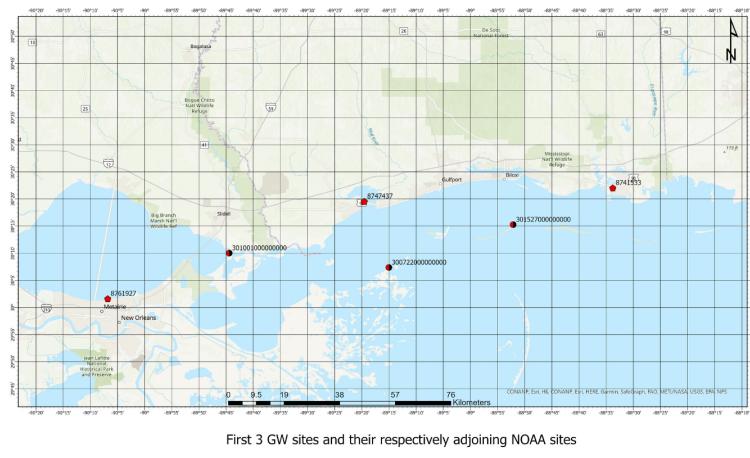


Figure 2.1: NOAA Tidal Stations near GW Sites

The four GW sites have been plotted on ArcGIS Pro along with their nearby NOAA Stations (Figure: 2.1) which are used for plotting data and model.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

2.2 Data Visualisation

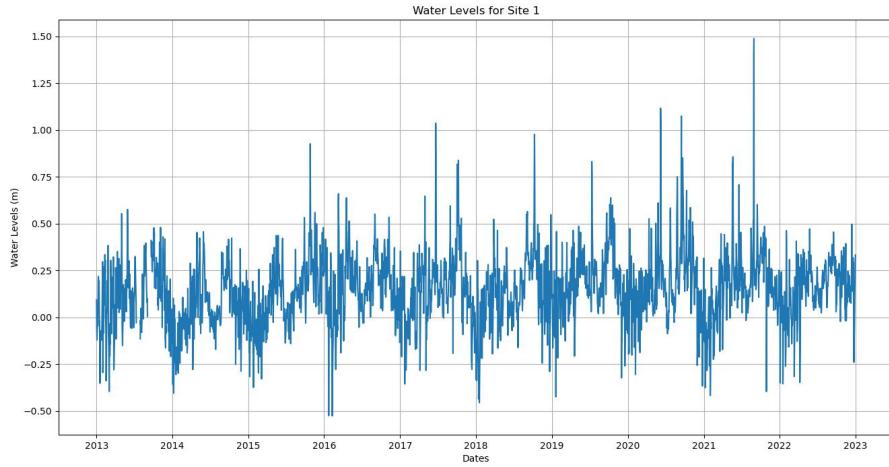


Figure 2.2: Tidal levels for site 1

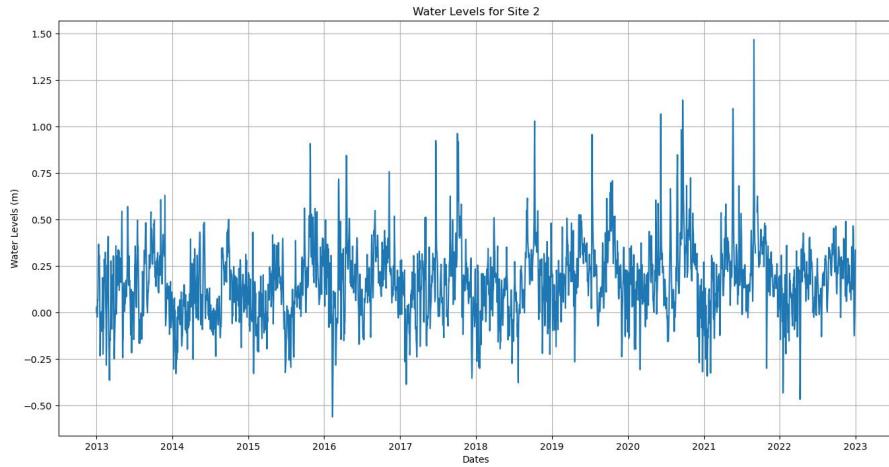


Figure 2.3: Tidal levels for site 2

The figures describe the distribution of daily data with the length of the time series being 3526 from 2013 January 1 to 2022 December 31. The GW levels and GW quality (specific conductivity) have been interpolated using KNN-imputation where the number of neighbours taken was 2. Also, a moving average of 7 days was calculated and plotted from the mean daily temperatures dataset.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

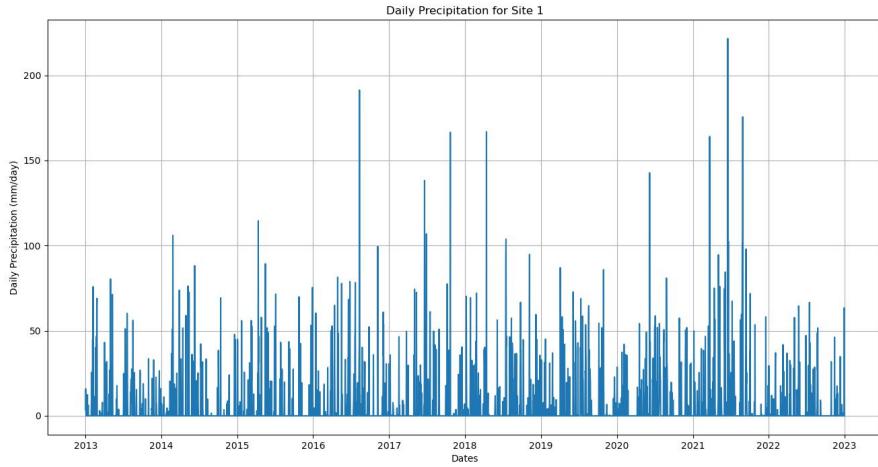


Figure 2.4: Precipitation for site 1

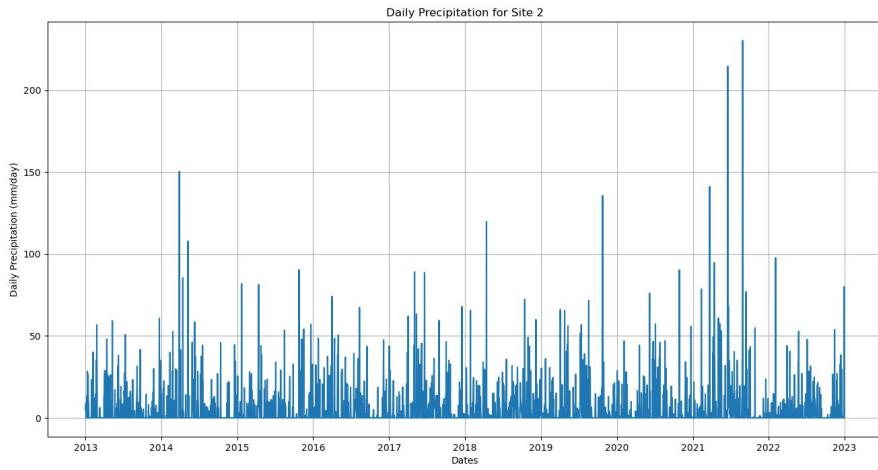


Figure 2.5: Precipitation for site 2

2.3 Conclusion

The interpolation used for interpolating GW levels and quality was K-Nearest Neighbour (KNN) imputer where n neighbours were selected to be 2. The GW data was collected from R with dataRetrieval package from USGS. Tidal level data were collected from National Oceanic and Atmospheric Administration

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

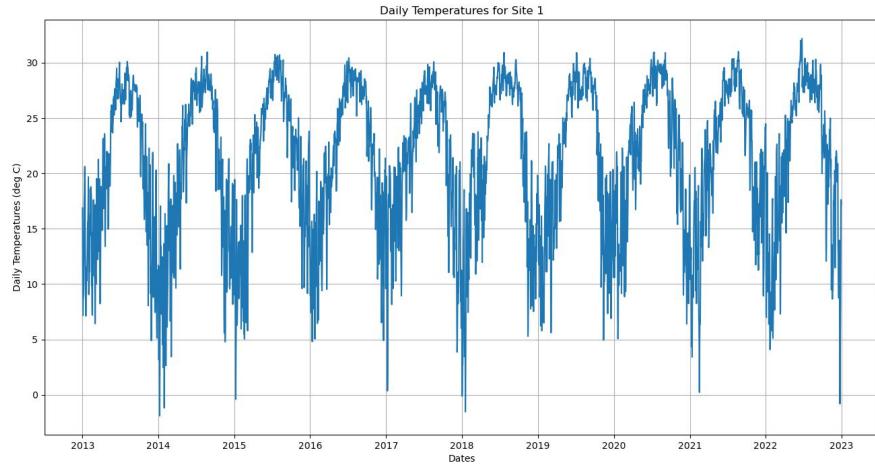


Figure 2.6: Temperature for site 1

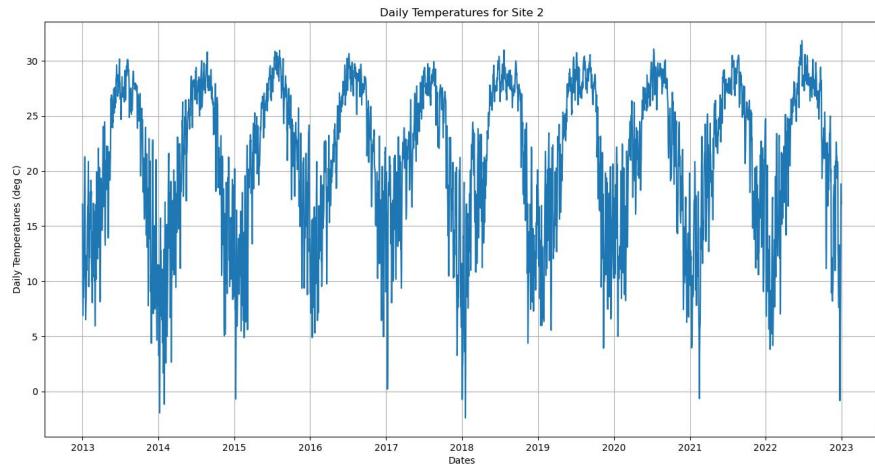


Figure 2.7: Temperature for site 2

(NOAA). Daily Meteorological data was collected from single-pixel extraction tool of daymet website. The datasets show high seasonality and non-stationarity. Imputation does not yield good results as there are gaps filled with constant values at some points in the time-series.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

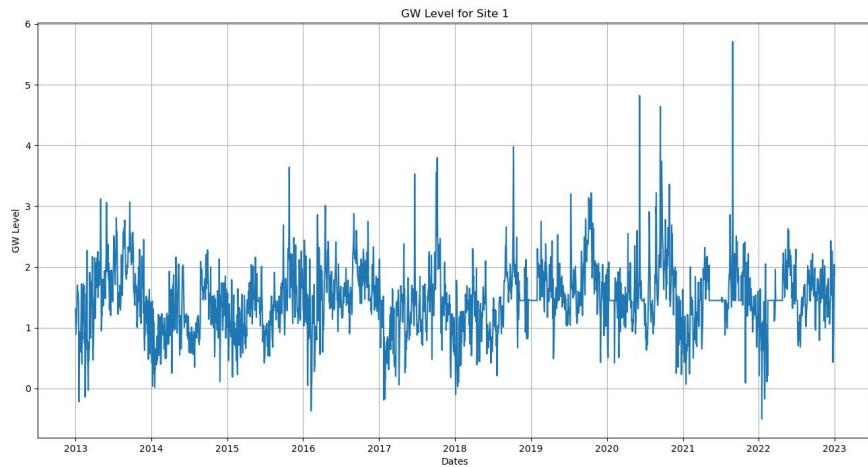


Figure 2.8: Groundwater levels for site 1

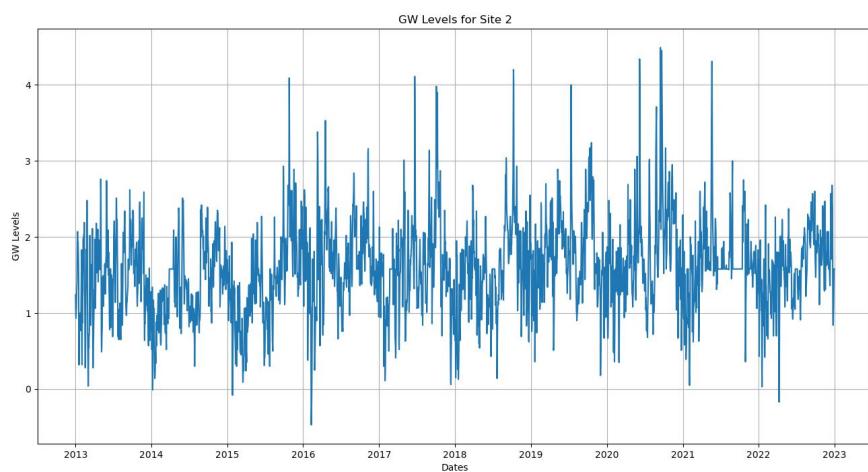


Figure 2.9: Groundwater levels for site 2

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

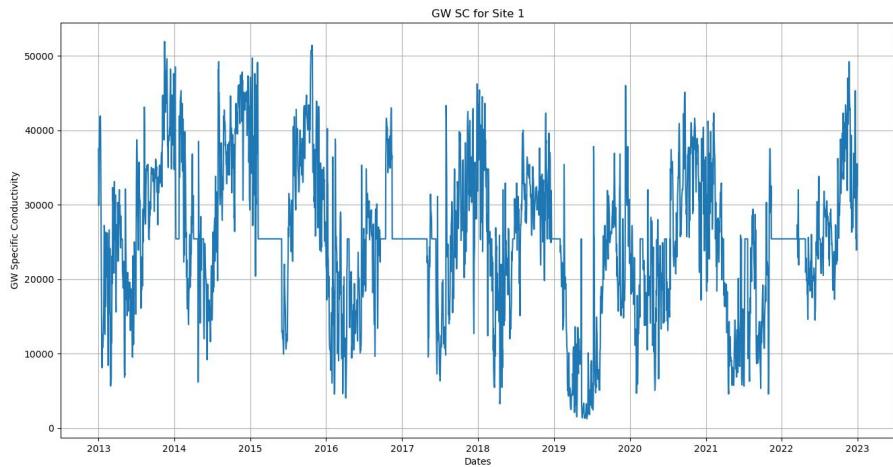


Figure 2.10: Groundwater Specific Conductivity for site 1

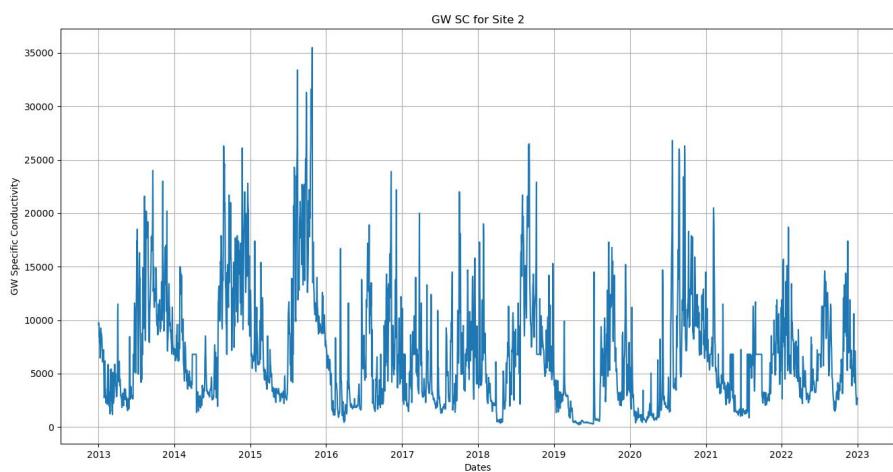


Figure 2.11: Groundwater Specific Conductivity for site 2

Chapter 3

Introduction To ARIMA and Prophet

3.1 Auto-ARIMA

The following figures show the re-sampled data for all the data sets such as temperature, precipitation, tidal change values, groundwater level values and groundwater specific conductivity values. These data have been re-sampled from daily to monthly for better calculation in the Auto-ARIMA model as the repetitive state occurs every 12 iterations while in the daily data the repetition occurs every 365 iterations which costs a lot of computational power in the Auto-ARIMA functioning. The Auto-ARIMA model in our case takes three input features while fitting on the training set and predicting on the testing set. Otherwise the Auto-ARIMA model is not beneficial at most.

3.2 ARIMA [7]

An Autoregressive Integrated Moving Average (ARIMA) model is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. An ARIMA model is composed of three parts mainly:

- **AutoRegression (AR):** refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

- **Integrated (I):** represents the differencing of raw observations to allow the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
- **Moving Average (MA):** incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations. A moving average process of order q is a linear combination of the current white noise term and the q most recent past white noise terms and is defined by

$$x_t = \omega_t + \beta_1\omega_{t-1} + \dots + \beta_q\omega_{t-q}$$

where ω_t is the white noise with zero mean and standard deviation σ_ω^2 [8].

Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA with p , d , and q , where integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

- p : the number of lag observations in the model, also known as the lag order.
- d : the number of times the raw observations are differenced; also known as the degree of differencing.
- q : the size of the moving average window, also known as the order of the moving average.

ARIMA(1,0,0) = first-order autoregressive model: if the series is stationary and autocorrelated, perhaps it can be predicted as a multiple of its own previous value, plus a constant. The forecasting equation in this case is

$$\hat{Y}_t = \mu + \phi_1 Y_{t-1}$$

which is Y regressed on itself lagged by one period. This is an “ARIMA(1,0,0)+constant” model. If the mean of Y is zero, then the constant term would not be included.

The model depicts mean-reverting behaviour in which the value for the following period should be projected to be ϕ_1 times as far from the mean as the value for this period if the slope coefficient ϕ_1 is positive and less than 1 in magnitude (it must be less than 1 in magnitude if Y is stationary). If ϕ_1 is negative, mean-reverting behaviour with a change in sign is predicted, which means that if Y is above the mean this period, it will be below the mean the following period [9].

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

In a second-order autoregressive model (ARIMA(2,0,0)), there would be a Y_{t-2} term on the right as well, and so on. Depending on the signs and magnitudes of the coefficients, an ARIMA(2,0,0) model could describe a system whose mean reversion takes place in a sinusoidally oscillating fashion, like the motion of a mass on a spring that is subjected to random shocks.

3.3 SARIMAX model

A model subclass of the ARIMA family is called SARIMAX, or Seasonal Auto-Regressive Integrated Moving Average with eXogenous Factors. In ARIMA models, the moving-average term (MA) and the autoregressive term (AR) are naturally split into two portions. The former considers a value to be just the weighted sum of prior values at any given moment. The identical value is modelled as a weighted sum of prior residuals in the later method (confer. time series decomposition). To distinguish between multiple time series, there is also an integrated term (I), which is discussed in greater detail below. Given how complicated this subject is and how much arithmetic is involved, we strongly suggest that you read extra materials [10].

Overall, ARIMA is a very decent type of models. However, the problem with this vanilla version is that it cannot handle seasonality — a big weakness. Comes SARIMA — the predecessor of SARIMAX. One shorthand notation for SARIMA models is:

$$SARIMA(p, d, q) \times (P, D, Q, S)$$

where:

- p = non-seasonal autoregressive (AR) order
- d = non-seasonal differencing
- q = non-seasonal moving average (MA) order
- P = seasonal AR order
- D = seasonal differencing
- Q = seasonal MA order
- S = length of repeating seasonal pattern.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

We will use this notation from now on. By adding those seasonal AR and seasonal MA components, SARIMA solves the seasonality problem.

The difference between the ARIMA and the SARIMAX model is that ARIMA only derives its predictions on time whereas the SARIMAX approach gives results based on several exogenous factors in our case temperature, precipitation and GW levels. Also the SARIMAX model takes care of the seasonality embedded in our systems.

3.4 Data Resampling for Auto-ARIMA

The following figures denote the re-sampled data plots of Site 1 information and Site 2 information. The data has been resampled from daily to monthly. The daily raw data has been represented in the last chapter. The monthly re-sampled data is being shown in the following figures in the form of subplots (Figures [3.1](#) and [3.2](#)).

3.5 Facebook Prophet

The Facebook Core Data Science team created Facebook Prophet, an open-source forecasting tool. It was created to make time series forecasting easier to understand for both data scientists and laypeople. Prophet is used frequently because of its adaptability, simplicity, and capacity to handle different time series forecasting workloads. Here are some of Facebook Prophet's main attributes and features:

- Automatic Seasonality Detection: Prophet automatically recognises and models a variety of seasonal patterns in time series data, including daily, weekly, and yearly patterns. Missing values and data with irregular spacing are both supported.
- Trend Modeling: Both linear and non-linear trends in the data can be captured by Prophet's adaptable trend modelling component. It is suitable for analysing time series with sudden variations because it can handle abrupt changes and shifts in the trend.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

- Holiday Effects: The user-defined holiday impacts that can affect the time series being forecasted can be included in Prophet. This function is very helpful when anticipating data that is affected by regular occasions, such as public holidays or marketing activities.
- Customizable Parameters: The Prophet model's different parameters can be altered by users to precisely tailor the predicting outcomes to meet their unique needs. This involves establishing previous assumptions about trend flexibility, seasonality, and handling outliers.
- Scalability: Prophet is made to efficiently handle huge time series datasets. It uses a decomposable approach to support parallel processing and scalability, making it appropriate for forecasting and analysing massive amounts of data.
- Implementation: Python is used to develop Prophet, and it works well with the well-known data science environment, which includes tools like Pandas and NumPy. It offers a straightforward and understandable API for fitting models, making predictions, and visualising the outcomes.
- Open-Source and Community Support: Prophet is an open-source project, which means that anybody can use and contribute to the code and documentation. A thriving user community makes it possible for users to get access to tools, guides, and examples that can help them use Prophet efficiently.

Facebook Prophet has gained popularity in various domains, including retail, finance, energy, and demand forecasting. Its user-friendly interface, automated seasonality detection, and customizable parameters make it a valuable tool for both time series experts and practitioners without extensive forecasting experience. We are going to use this technique to predict and evaluate GW specific conductivity (SC) for the two sites we have selected in the state of Louisiana.

Facebook Prophet is a time series forecasting method developed by Facebook's Core Data Science team. It is based on an additive model that combines trend, seasonality, and holiday effects. The governing equations of the Prophet method can be summarized as follows:

1. Trend equation: The trend component in Prophet is modeled as a piecewise linear function. It consists of a linear growth term and a set of change points where the growth rate can change. The trend equation is defined as:

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

$$y(t) = g(t) + s(t) + h(t) + \epsilon$$

where:

- $g(t)$ represents the trend component.
- $s(t)$ represents the seasonality component.
- $h(t)$ represents the holiday component.
- ϵ is the error term.

2. Trend component: The trend component, $g(t)$, is modeled as a piecewise linear function using changepoints. It can be written as:

$$g(t) = k + \sum_{i=1}^N (a_i(t - t_i)_+)$$

where:

- k is the initial trend value.
- N is the number of changepoints.
- a_i represents the rate of change at each changepoint.
- t_i represents the location of each changepoint.
- $(t - t_i)_+$ is a positive part function that evaluates to $t - t_i$ when $t > t_i$ and 0 otherwise.

3. Seasonality component: The seasonality component, $s(t)$, captures periodic patterns in the data. It is modeled using Fourier series with a specified number of harmonics (Fourier terms). The equation for the seasonality component is given by:

$$s(t) = \sum_{j=1}^M (a_j \cos(2\pi jt/P) + b_j \sin(2\pi jt/P))$$

where:

- M is the number of Fourier terms.
- a_j and b_j are the coefficients of the j -th Fourier term.
- P is the specified seasonality period.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

4. Holiday component: The holiday component, $h(t)$, is used to model the effects of holidays or other significant events. It is a user-defined input that specifies the dates and associated effects. The equation for the holiday component is:

$$h(t) = \sum_k g_k(t)$$

where:

- k represents the different holiday events.
- $g_k(t)$ is a piecewise linear function with changepoints specific to each holiday event.

The Prophet method estimates the model parameters by optimizing the maximum likelihood function using a variant of the stochastic gradient descent algorithm. Once the parameters are estimated, the forecast can be generated by combining the trend, seasonality, and holiday components.

It's important to note that the Prophet method incorporates various additional components and considerations, such as trend flexibility, trend changepoint detection, seasonality prior, and holiday effects. The above equations provide a simplified overview of the key components in the Prophet method. For a more detailed understanding, it is recommended to refer to the original research paper or the official documentation provided by Facebook [11].

3.6 SARIMAX and Facebook Prophet: A Comparative Study

Facebook Prophet and SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) are both widely used time series forecasting techniques, although they differ in their basic assumptions and features. Here is a comparison between Facebook Prophet and SARIMAX:

1. Model Structure:

- SARIMAX: The Box-Jenkins approach serves as the foundation for SARIMAX models. To represent the temporal dependencies and stationarity of the time series, they integrate autoregressive (AR), moving

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

average (MA), and differencing (I) components. Exogenous variables can also be included in SARIMAX models to account for outside influences on the time series.

- Prophet: Prophet is a generalised additive model (GAM) that adds together a number of elements, such as trend, seasonality, and vacations. Prophet uses a Fourier series for seasonality, a piecewise linear model for the trend, and programmable holiday effects. It doesn't specifically deal with autoregressive or differencing components.

2. Seasonality Modeling:

- SARIMAX: By incorporating seasonal differencing (D), seasonal AR, and seasonal MA components, SARIMAX models are able to detect seasonal patterns. It is necessary to provide the seasonal times.
- Prophet: Prophet models seasonality using Fourier series. It automatically recognises and takes into account a variety of seasonalities with various time frames, giving flexibility in capturing intricate seasonal patterns.

3. Trend Modeling:

- SARIMAX: SARIMAX models use autoregressive (AR) and differencing (I) components to capture trend. Statistical tests and diagnostics are often used to identify the order of these elements.
- Prophet: Prophet use a piecewise linear function with changepoints to model trends. The flexibility in capturing trend changes is provided by the automatic detection and selection of the changepoints and the fitting of a linear regression model to each segment.

4. Exogenous Variables:

- SARIMAX: Exogenous variables, or outside variables that could affect the time series, are a feature of SARIMAX models. These factors can be included in the model as additional regressors.
- Prophet: Exogenous variables are also supported by Prophet. Exogenous variables can be included in a dataframe provided by the user, and Prophet will include them into the model to capture their influence on the time series.

5. Ease of Use and Interpretability:

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

- SARIMAX: SARIMAX models necessitate precise model order and parameter selection, which can be difficult for users without a background in time series analysis. Due to the mix of AR, MA, and differencing components, SARIMAX models may potentially be more difficult to comprehend.
- Prophet: Prophet offers a more straightforward interface for time series forecasting and is made to be user-friendly. Many components of the modelling process are automated, including trend recognition and seasonality modelling. Prophet models' additive modelling technique often makes them easier to interpret.

6. Performance and Applicability:

- SARIMAX: SARIMAX models have a strong track record and are frequently employed in time series forecasting. A wide variety of time series with linear or nonlinear temporal relationships can be accommodated by them.
- Prophet: Prophet has grown in prominence as a result of its simplicity and capacity for handling complex forecasting issues. It works well with time series that have various seasonalities, strong seasonality, and irregular patterns, which makes it very valuable in industries like retail, e-commerce, and social media.

The decision between Prophet and SARIMAX ultimately comes down to the details of the time series data, the level of interpretability that is sought, the accessibility of exogenous variables, and the user's experience with modelling approaches. Before making a decision, it is frequently advantageous to test out both approaches and evaluate how well they work on a particular dataset.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

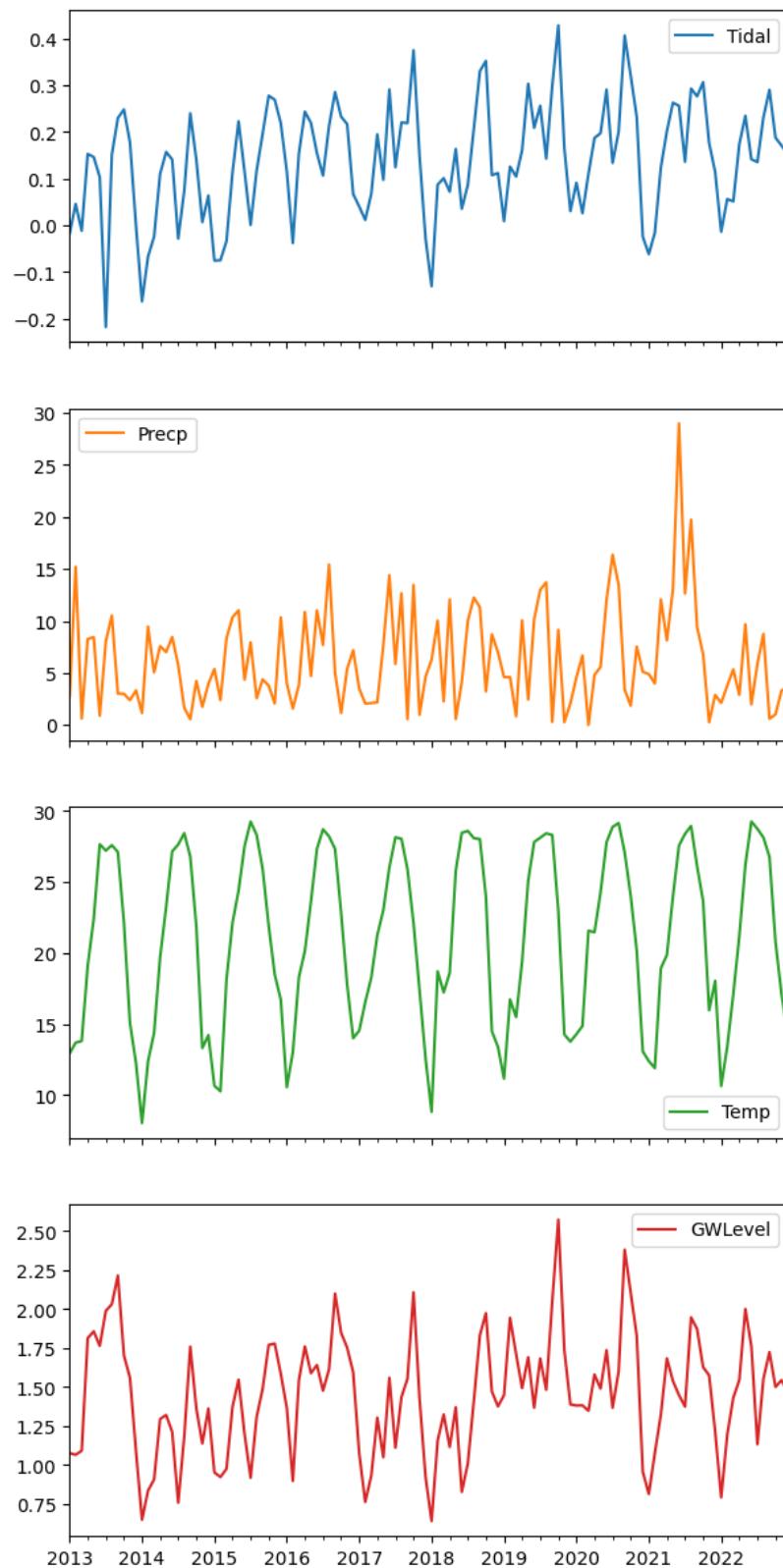


Figure 3.1: Site 1 re-sampled monthly data

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

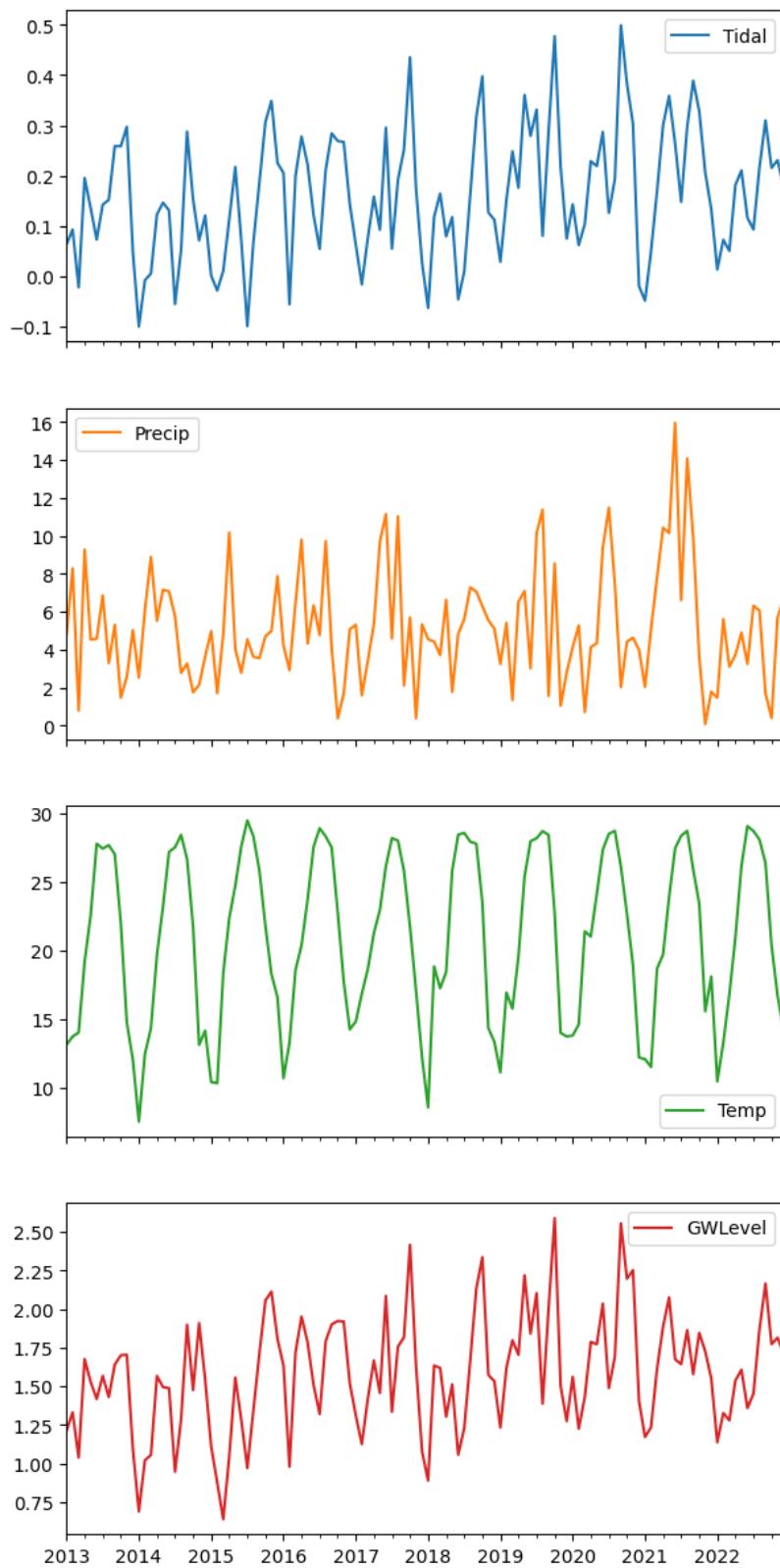


Figure 3.2: Site 2 re-sampled monthly data

Chapter 4

Numerical Modeling - Results and Discussions

Usually, in the basic ARIMA model, we need to provide the p,d, and q values which are essential. We use statistical techniques to generate these values by performing the difference to eliminate the non-stationarity and plotting ACF and PACF graphs. In Auto ARIMA, the model itself will generate the optimal p, d, and q values which would be suitable for the data set to provide better forecasting.

4.1 GW level forecasting using SARIMA

We first tried to forecast for the year 2023. Since there was no data available from the dataset we collected we had to go with the SARIMA approach without the exogenous input variables. The results are quite a bit defying as it moves to the mean of all the deviations over the years.

Hence, we accept the SARIMAX approach to predict for the year 2022.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

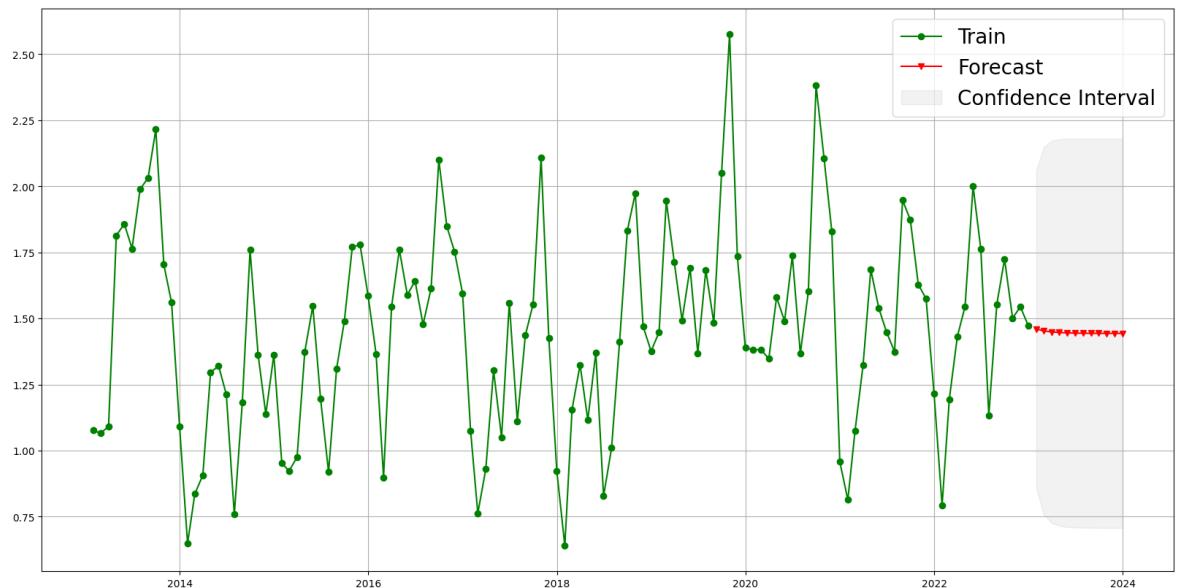


Figure 4.1: Site 1 Forecasting of GW levels

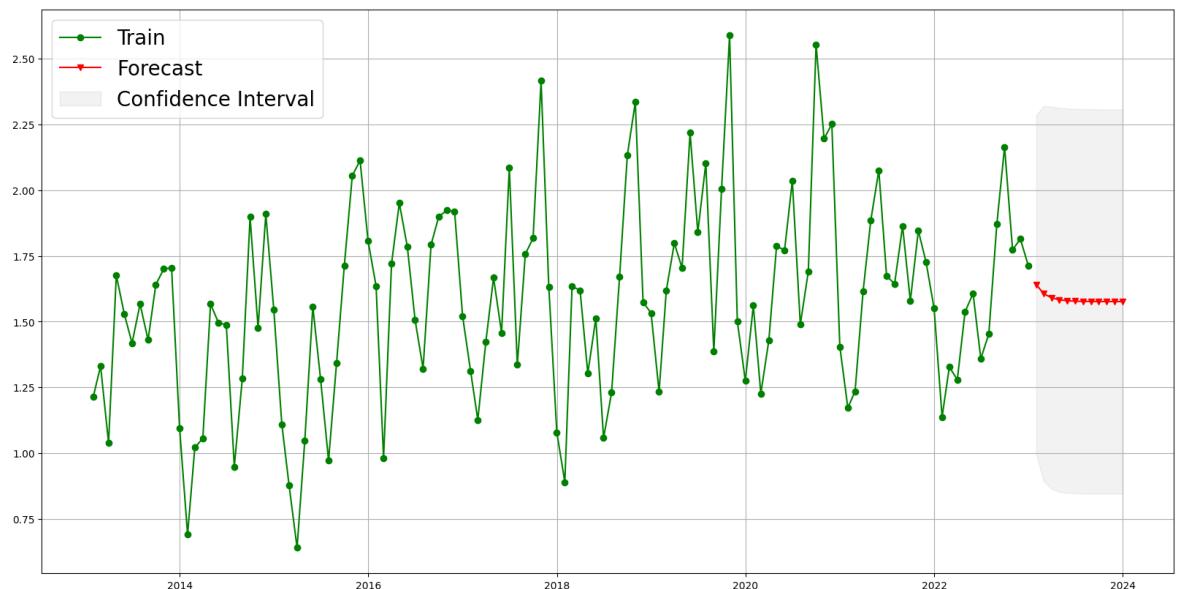


Figure 4.2: Site 2 Forecasting of GW levels

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

4.2 GW level prediction using SARIMAX

SARIMAX Test Set Model Validation		
Site No	R-squared	RMSE
300722089150100	0.61	0.19 = 7.6 %
301001089442600	0.89	0.09 = 3.6 %

The SARIMAX model for the prediction of GW level data from temperature, precipitation and tidal level from MSL show good results in case of site 1 while it is slightly better in case of site 2. The SARIMAX model has successfully predicted the GW levels which we are going to see in the following figures. The SARIMAX model can be successfully used to forecast into the future GW levels from past data. Although daily data could not be used to forecast/predict GW levels because of lack of time and computational power, monthly data was successfully used to train 9 years of historical data and predict for last 1 year or 12 monthly data points. The train-test split was done in this manner wherein there were 120 data points in total for 10 years time span of which first 108 (9 years) data points in the time series was used for training the SARIMAX model and last 12 data points (1 year) was used for testing the model.

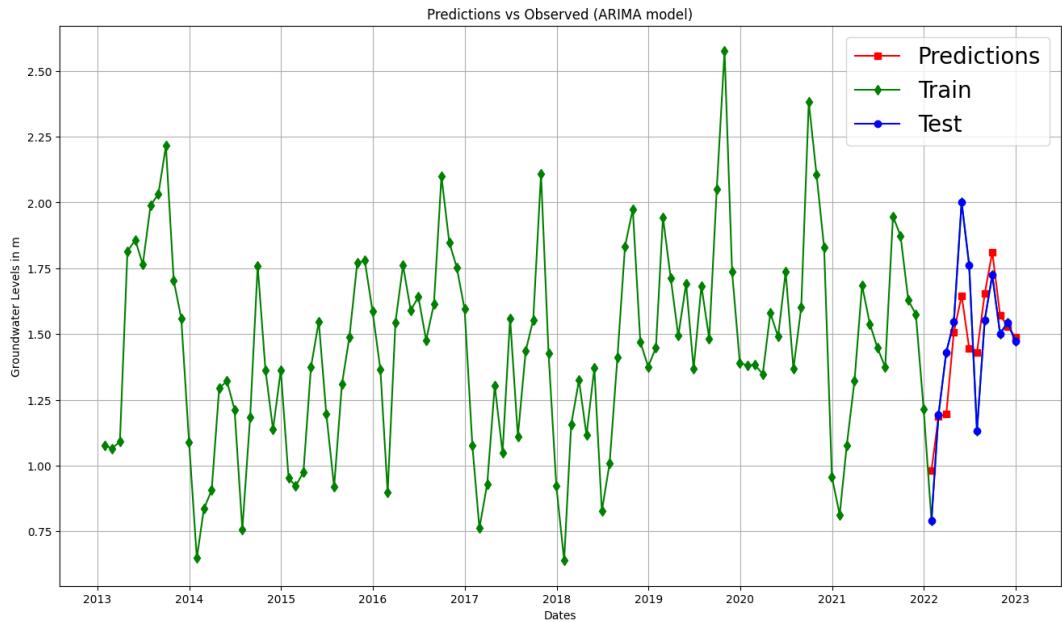


Figure 4.3: Site 1 GW Level Predictions

The model diagnostics and summary are given below (Figures:

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

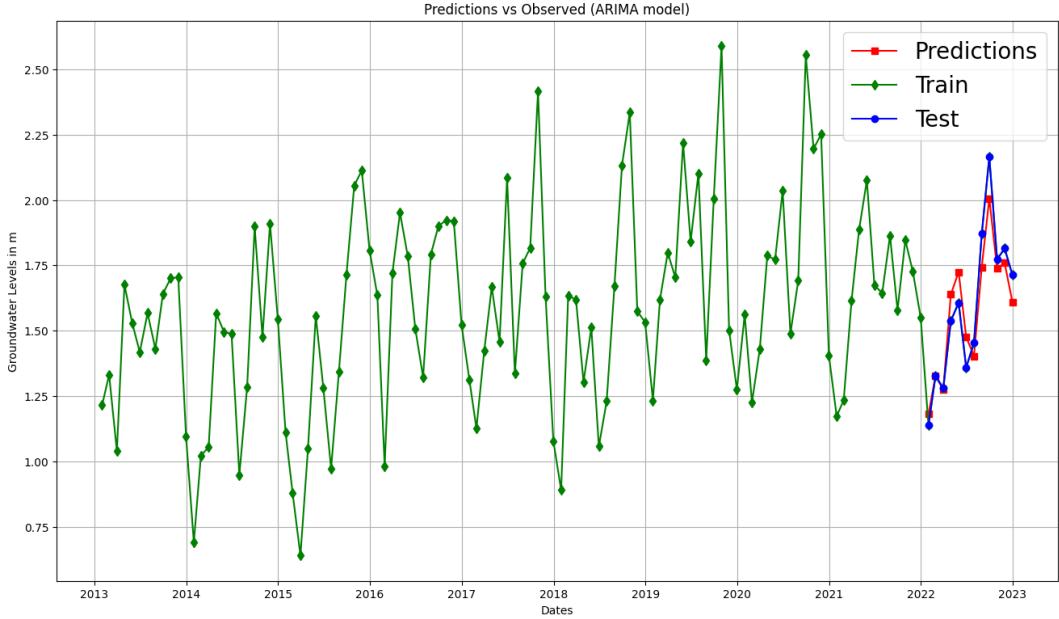


Figure 4.4: Site 2 GW Level Predictions

4.5, 4.6, 4.7, 4.8). The residuals in the models are normally distributed and the residuals also have a zero mean which shows our model is aptly accurate. The model summary show that our model generated a SARIMAX model which is also known as Seasonal ARIMA Exogenous.

4.3 GW Specific Conductivity (SC) prediction using Facebook Prophet

The code for predicting GW SC was done in R using the facebook prophet package. Although the predictions are a bit incompetent as compared to the SARIMAX in case of predicting GW levels which may have happened due to data unavailability or improper fitting, etc. The testing data set was taken as the last 365 days in the daily dataset. ALthough the SARIMAX approach was computed on monthly datasets, the Prophet approach was computed on daily format of the same data taking GW levels as one of the inputs. So, the inputs for Prophet are Tidal levels, Temperature, Precipitation and GW Levels. The target variable being GW Specific Conductivity. Figures 4.9 and 4.11 show the plot of the forecast of the target variable whereas Figures 4.10 and 4.12 shows the comparison between the Observed values and Predicted values in the test

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

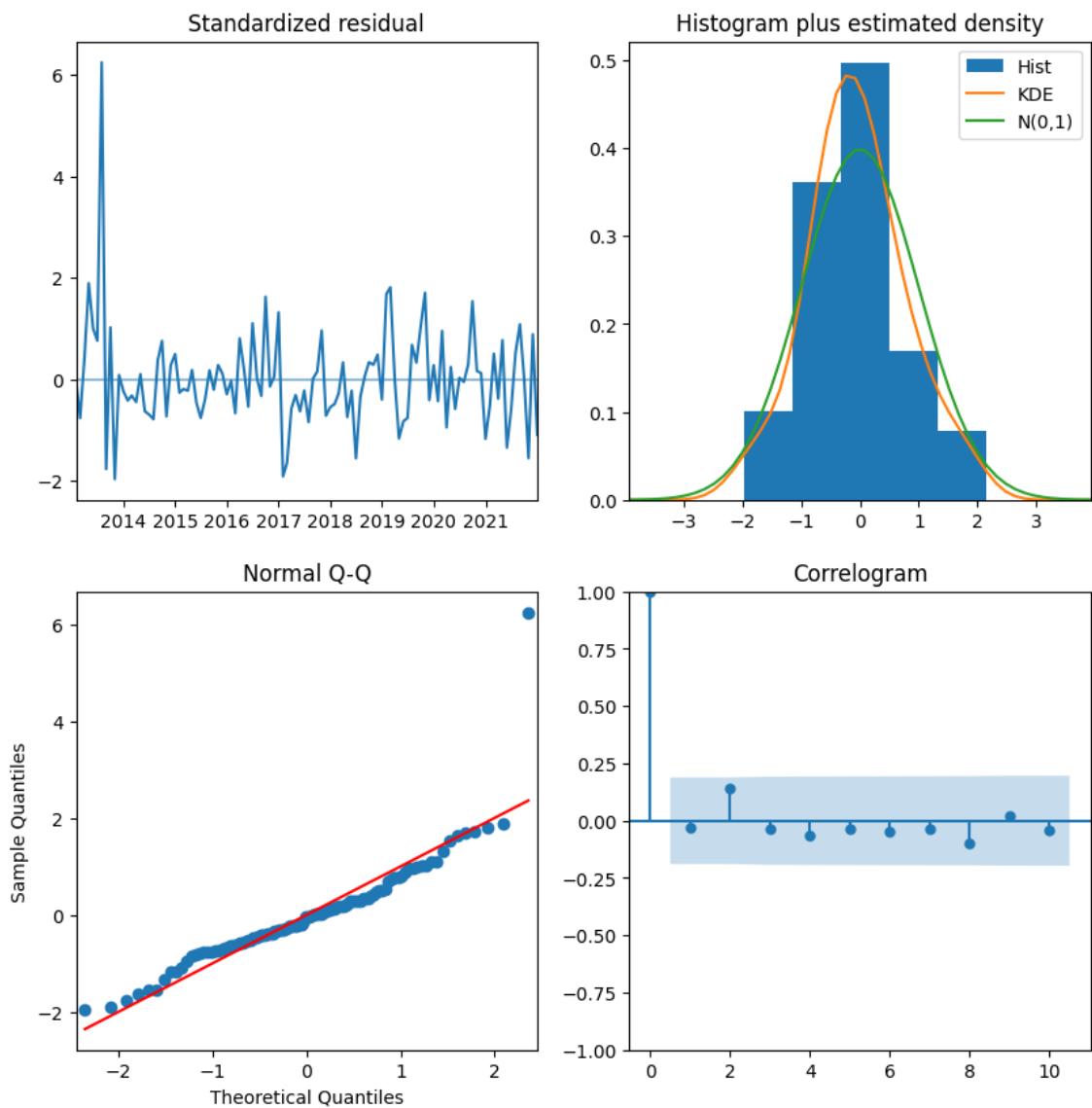


Figure 4.5: Model diagnostics for Site 1 GW Level

dataframe.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

SARIMAX Results

Dep. Variable: y **No. Observations:** 108

Model:	SARIMAX(1, 0, 0)	Log Likelihood	30.120
Date:	Mon, 26 Jun 2023	AIC	-48.240
Time:	06:36:32	BIC	-32.147
Sample:	01-31-2013 - 12-31-2021	HQIC	-41.715

Covariance Type: opg

	coef	std err	z	P> z	[0.025 0.975]
Intercept	0.4117	0.098	4.205	0.000	0.220 0.604
Tidal	2.3675	0.167	14.186	0.000	2.040 2.695
Precp	-0.0020	0.004	-0.452	0.651	-0.011 0.007
Temp	-0.0004	0.007	-0.056	0.956	-0.014 0.013
ar.L1	0.6386	0.062	10.292	0.000	0.517 0.760
sigma2	0.0334	0.004	7.599	0.000	0.025 0.042

Ljung-Box (L1) (Q): 0.11 **Jarque-Bera (JB):** 772.87

Prob(Q):	0.74	Prob(JB):	0.00
-----------------	------	------------------	------

Heteroskedasticity (H): 0.50 **Skew:** 2.22

Prob(H) (two-sided): 0.04 **Kurtosis:** 15.33

Figure 4.6: Model summary for Site 1 GW Level

Prophet Model Train Set Validation		
Site No	R-squared	RMSE
300722089150100	0.58	6,588 = 20 %
301001089442600	0.56	3,617 = 28 %

Table 4.1: Train set validation for GW SC prediction

Prophet Model Test Set Validation		
Site No	R-squared	RMSE
300722089150100	0.48	10,359 = 40 %
301001089442600	0.08	4,554 = 33.33 %

Table 4.2: Test set validation for GW SC prediction

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

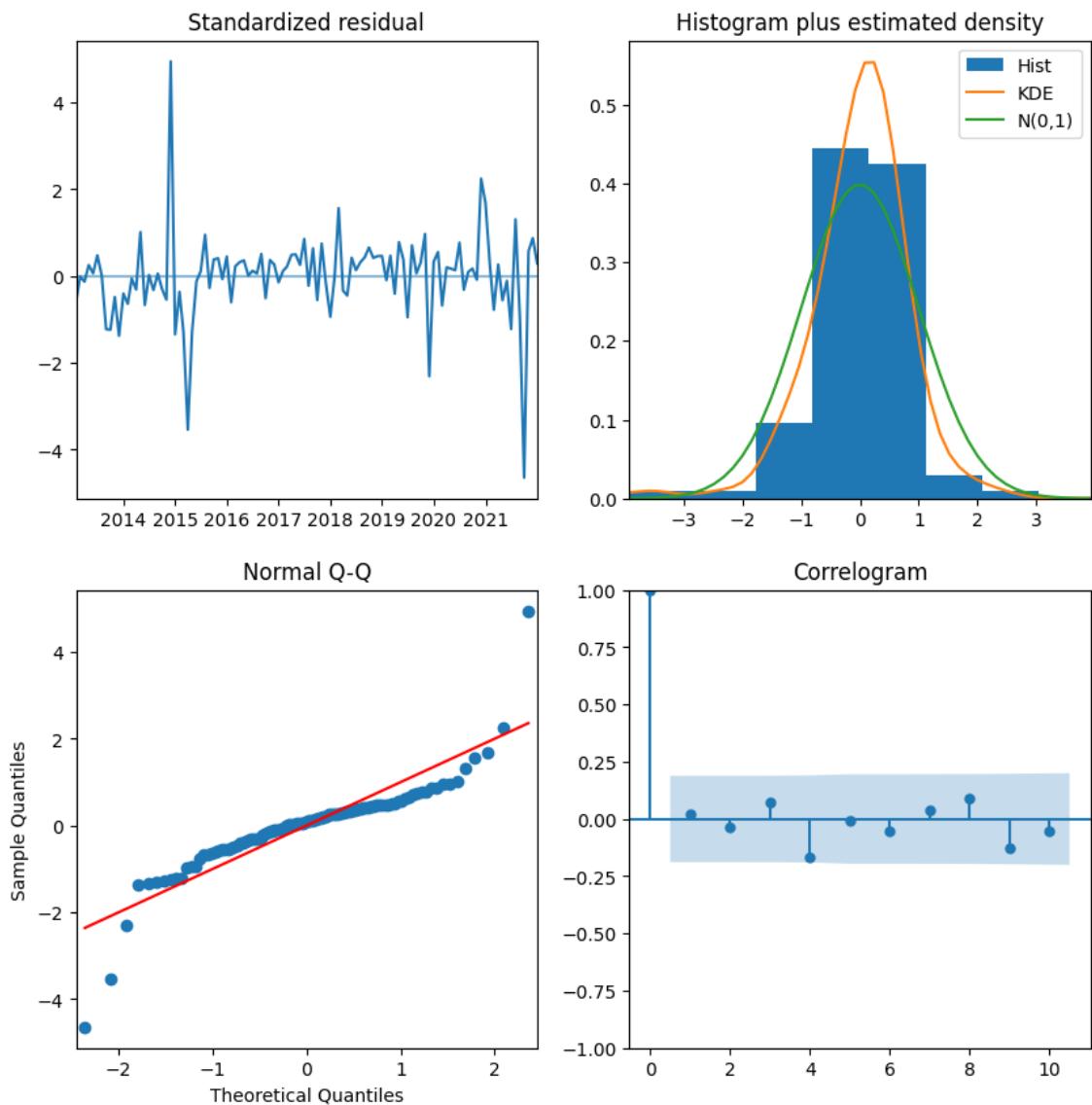


Figure 4.7: Model diagnostics for Site 2 GW Level

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
SARIMAX Results
Dep. Variable: y No. Observations: 108
Model: SARIMAX(1, 0, 0) Log Likelihood: 70.058
Date: Thu, 01 Jun 2023 AIC: -128.115
Time: 14:39:46 BIC: -112.023
Sample: 01-31-2013 HQIC: -121.590
                    - 12-31-2021
Covariance Type: opg
            coef  std err      z   P>|z| [0.025 0.975]
intercept 0.6199  0.105  5.879  0.000 0.413  0.827
Tidal     2.7508  0.135 20.402 0.000 2.487  3.015
Precip    -0.0025  0.006 -0.420 0.674 -0.014 0.009
Temp      0.0020  0.004  0.488  0.625 -0.006 0.010
ar.L1      0.4390  0.086  5.085  0.000 0.270  0.608
sigma2    0.0160  0.001 11.811  0.000 0.013  0.019
Ljung-Box (L1) (Q): 0.05 Jarque-Bera (JB): 386.13
Prob(Q):        0.83 Prob(JB):    0.00
Heteroskedasticity (H): 0.89 Skew:       -0.24
Prob(H) (two-sided): 0.73 Kurtosis:    12.25
```

Figure 4.8: Model summary for Site 2 GW Level

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

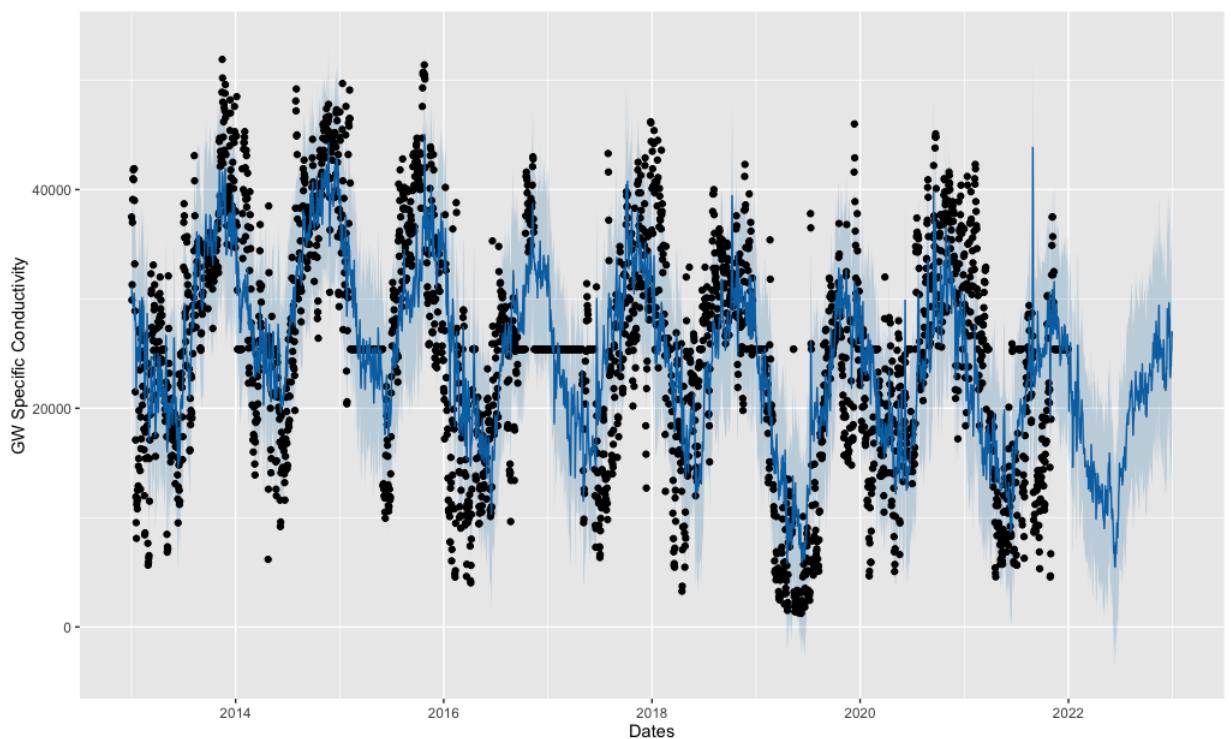


Figure 4.9: Fitting of GW SC data (Site - 1) into the prophet model.

**PREDICTING SALTWATER INTRUSION IN COASTAL
AQUIFERS BY DATA-DRIVEN MODELING**

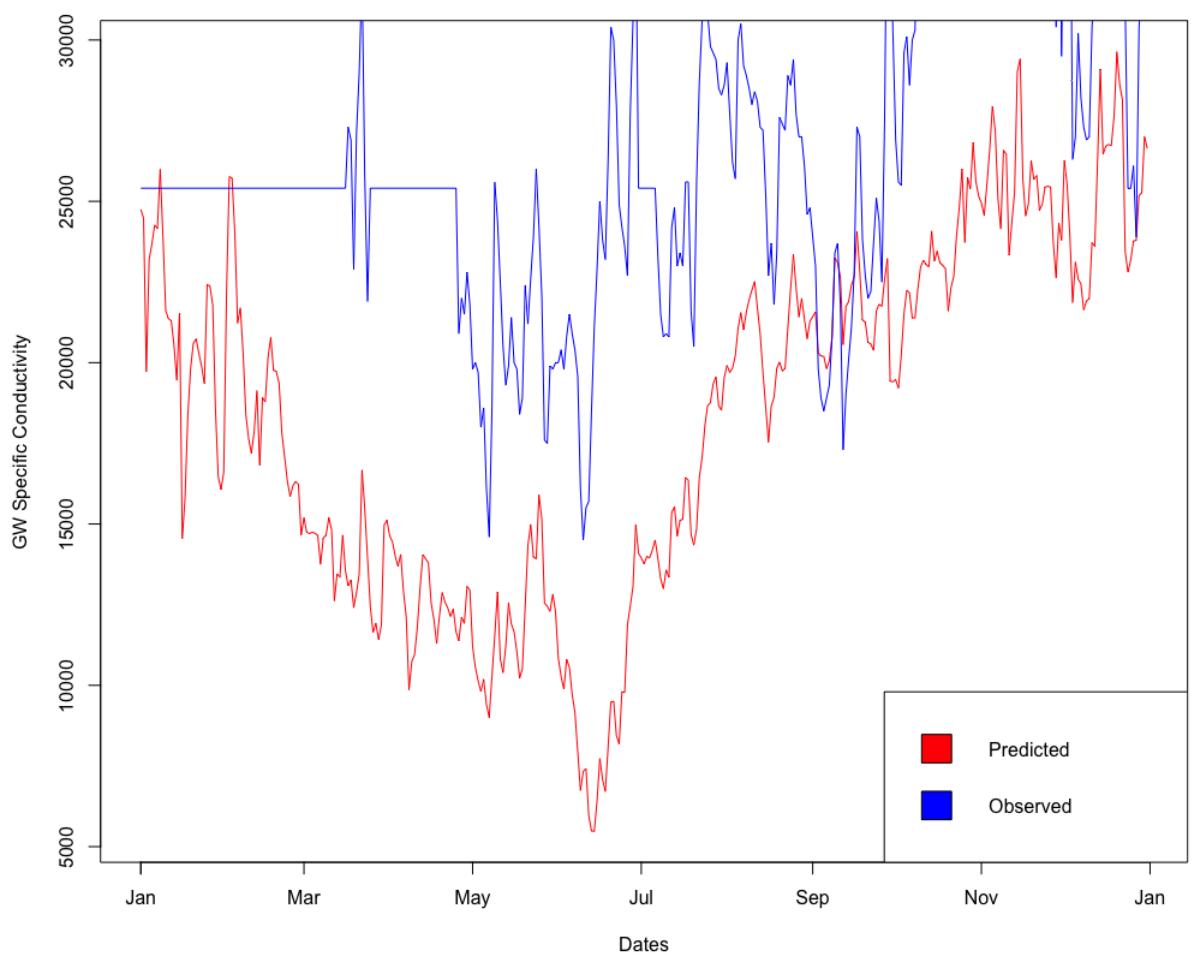


Figure 4.10: Ovservation vs prediction of test data set for Site 1 GW SC.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

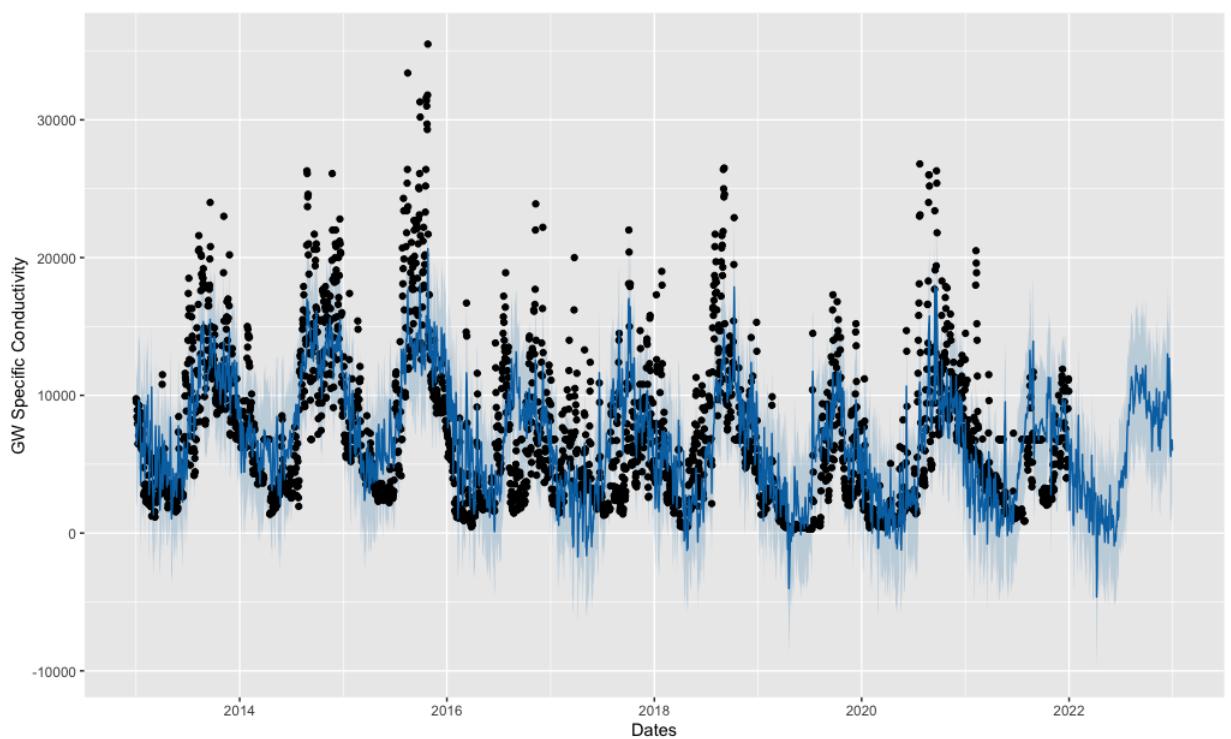


Figure 4.11: Fitting of GW SC data (Site - 2) into the prophet model.

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

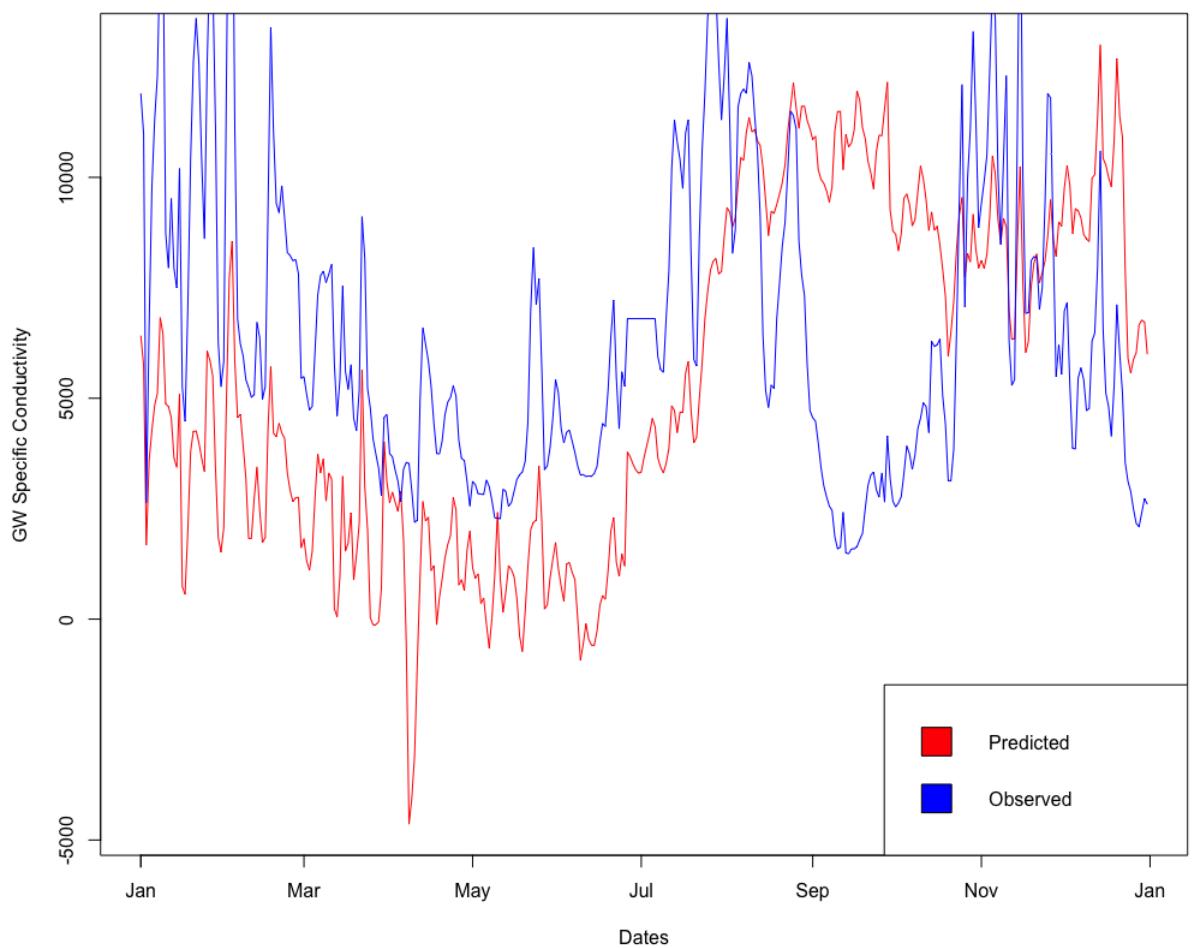


Figure 4.12: Ovservation vs prediction of test data set for Site 2 GW SC.

Chapter 5

Conclusion

The ARIMA approach has been beneficial for predicting monthly GW levels and similar models can be trained for other sites where the three input parameters are available. With the help of forecasts in which the input variables are taken from global climate models, estimates of aquifer GW levels can be obtained. This benefits stakeholders and policy makers.

Also, the GW Specific Conductivity has been predicted using Prophet. Although the predictions at one of the sites are not very accurate, the prediction on the other sites performed well. Saltwater intrusions also depend on the topography of the aquifer. Our model with Prophet is only a starting point in the modelling and future models should incorporate further variables. We note that the Prophet model outperformed ARIMA for the prediction of GW quality. With Groundwater Specific Conductivity (GW SC) we used a surrogate variable for GW quality, which indicates saltwater intrusion. Higher conductivity and salinity content accelerate corrosion of subsurface infrastructures, hence our model also informs estimates of expected durability and future replacement costs.

To the vulnerability assessment of coastal aquifers that enables the planning and prioritization of mitigation efforts, we have contributed a modelling approach for two important parameters related to saltwater intrusion. Our modelling approach includes commonly measured variables and can be transferred and extended to aid planning and inform policies. It thus forms one puzzle piece in the protection of coastal communities.

Publications

Pralay Sankar Maitra, Remya R and Dr. Alka Singh "*Groundwater variability along with the varying coastline of Thiruvananthapuram city*", International Conference for Advancement in Technology, ICONAT IEEE Goa Chapter, India. [Presented]

References

- [1] A. Mosavi, F. Sajedi Hosseini, B. Choubin, F. Taromideh, M. Ghodsi, B. Nazari, and A. A. Dineva, “Susceptibility mapping of groundwater salinity using machine learning models,” *Environmental Science and Pollution Research*, vol. 28, pp. 10804–10817, 2021. [vi, 3](#)
- [2] H. Sahour, V. Gholami, and M. Vazifedan, “A comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer,” *Journal of Hydrology*, vol. 591, p. 125321, 2020. [vi, 3](#)
- [3] X. Zhang, F. Dong, G. Chen, and Z. Dai, “Advance prediction of coastal groundwater levels with temporal convolutional and long short-term memory networks,” *Hydrology and Earth System Sciences*, vol. 27, no. 1, pp. 83–96, 2023. [vi, 4](#)
- [4] J. Zhang, Y. Zhu, X. Zhang, M. Ye, and J. Yang, “Developing a long short-term memory (lstm) based model for predicting water table depth in agricultural areas,” *Journal of hydrology*, vol. 561, pp. 918–929, 2018. [vi, 4](#)
- [5] W. contributors, “File:Saltwater intrusion en.svg - Wikipedia,” 11 2020. [ix, 3](#)
- [6] M. A. Marfai and L. King, “Potential vulnerability implications of coastal inundation due to sea level rise for the coastal zone of Semarang city, Indonesia,” *Environmental Geology*, vol. 54, pp. 1235–1245, May 2008. [1](#)
- [7] A. Hayes, “Autoregressive Integrated Moving Average (ARIMA) Prediction Model,” *Investopedia*, 12 2022. [12](#)
- [8] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting*. Springer, 2002. [13](#)
- [9] F. S. of Business, “Introduction to ARIMA models.” [13](#)

REFERENCES

- [10] B. Nguyen, “End-to-End Time Series Analysis and Forecasting: a Trio of SARIMAX, LSTM and Prophet (Part 1),” 12 2021. [14](#)
- [11] S. J. Taylor and B. Letham, “Forecasting at scale,” *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018. [18](#)
- [12] S. W. Chang, T. P. Clement, M. J. Simpson, and K.-K. Lee, “Does sea-level rise have an impact on saltwater intrusion?,” *Advances in Water Resources*, vol. 34, no. 10, pp. 1283–1291, 2011.
- [13] Q. He and B. R. Silliman, “Climate Change, Human Impacts, and Coastal Ecosystems in the Anthropocene,” *Current Biology*, vol. 29, no. 19, pp. R1021–R1035, 2019.
- [14] J. Mielińczuk and P. Wojdyłło, “Estimation of hurst exponent revisited,” *Computational Statistics Data Analysis*, vol. 51, no. 9, pp. 4510–4525, 2007.
- [15] B. Arora, H. M. Wainwright, D. Dwivedi, L. J. Vaughn, J. B. Curtis, M. S. Torn, B. Dafflon, and S. S. Hubbard, “Evaluating temporal controls on greenhouse gas (ghg) fluxes in an arctic tundra environment: An entropy-based approach,” *Science of The Total Environment*, vol. 649, pp. 284–299, 2019.
- [16] J. Hinkel, C. Jaeger, R. J. Nicholls, J. Lowe, O. Renn, and S. Peijun, “Sea-level rise scenarios and coastal risk management,” *Nature Clim Change*, vol. 5, pp. 188–190, Mar. 2015.
- [17] S. A. Kulp and B. H. Strauss, “New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding,” *Nature Communications*, vol. 10, p. 4844, Oct. 2019.
- [18] S. Jasechko, D. Perrone, H. Seybold, Y. Fan, and J. W. Kirchner, “Groundwater level observations in 250,000 coastal US wells reveal scope of potential seawater intrusion,” *Nature communications*, vol. 11, no. 1, pp. 1–9, 2020. ISBN: 2041-1723 Publisher: Nature Publishing Group.
- [19] S. Huq, H. Reid, M. Konate, A. Rahman, Y. Sokona, and F. Crick, “Mainstreaming adaptation to climate change in least developed countries (LDCs),” *Climate Policy*, vol. 4, no. 1, pp. 25–43, 2004. ISBN: 1469-3062 Publisher: Taylor & Francis.
- [20] A. D. Werner, M. Bakker, V. E. Post, A. Vandenbohede, C. Lu, B. Ataie-Ashtiani, C. T. Simmons, and D. A. Barry, “Seawater intrusion processes, investigation and management: recent advances and future challenges,” *Advances in water resources*, vol. 51, pp. 3–26, 2013. ISBN: 0309-1708 Publisher: Elsevier.

REFERENCES

- [21] S. Halder, M. B. Roy, and P. K. Roy, “Analysis of groundwater level trend and groundwater drought using Standard Groundwater Level Index: A case study of an eastern river basin of West Bengal, India,” *SN Applied Sciences*, vol. 2, no. 3, pp. 1–24, 2020. ISBN: 2523-3971 Publisher: Springer.
- [22] J. Nanteza, C. R. De Linage, B. F. Thomas, and J. S. Famiglietti, “Monitoring groundwater storage changes in complex basement aquifers: An evaluation of the GRACE satellites over East Africa,” *Water Resources Research*, vol. 52, no. 12, pp. 9542–9564, 2016. ISBN: 0043-1397 Publisher: Wiley Online Library.
- [23] J. Klassen and D. M. Allen, “Assessing the risk of saltwater intrusion in coastal aquifers,” *Journal of Hydrology*, vol. 551, pp. 730–745, 2017. ISBN: 0022-1694 Publisher: Elsevier.
- [24] D. V. Budescu, S. Broome, and H.-H. Por, “Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change,” *Psychological science*, vol. 20, no. 3, pp. 299–308, 2009. ISBN: 0956-7976 Publisher: SAGE Publications Sage CA: Los Angeles, CA.
- [25] C. M. Crain, B. S. Halpern, M. W. Beck, and C. V. Kappel, “Understanding and managing human threats to the coastal marine environment,” *Annals of the New York Academy of Sciences*, vol. 1162, no. 1, pp. 39–62, 2009. ISBN: 0077-8923 Publisher: Wiley Online Library.
- [26] C. Tebaldi, B. H. Strauss, and C. E. Zervas, “Modelling sea level rise impacts on storm surges along US coasts,” *Environmental Research Letters*, vol. 7, no. 1, p. 014032, 2012. ISBN: 1748-9326 Publisher: IOP Publishing.
- [27] H. Xiao, D. Wang, S. C. Medeiros, S. C. Hagen, and C. R. Hall, “Assessing sea-level rise impact on saltwater intrusion into the root zone of a geo-typical area in coastal east-central Florida,” *Science of The Total Environment*, vol. 630, pp. 211–221, 2018. ISBN: 0048-9697 Publisher: Elsevier.
- [28] R. M. Sorensen, R. N. Weisman, and G. P. Lennon, “Control of erosion, inundation, and salinity intrusion caused by sea level rise,” *Barth MC, Titus JG (eds) Greenhouse effect and sea level rise. A challenge for this generation. Van Nostrand Reinhold Company Inc, New York. Pp179*, vol. 214, 1984. Publisher: Citeseer.
- [29] M. S. Raimundo and J. Okamoto Jr, “Application of Hurst Exponent (H) and the R/S Analysis in the Classification of FOREX Securities,” *International Journal of Modeling and Optimization*, vol. 8, no. 2, pp. 116–124, 2018.

REFERENCES

- [30] L. Kristoufek and M. Vosvrda, “Commodity futures and market efficiency,” *Energy Economics*, vol. 42, pp. 50–57, 2014. ISBN: 0140-9883 Publisher: Elsevier.
- [31] M. Li, Y. Zhang, J. Wallace, and E. Campbell, “Estimating annual runoff in response to forest change: a statistical method based on random forest,” *Journal of Hydrology*, vol. 589, p. 125168, 2020. ISBN: 0022-1694 Publisher: Elsevier.
- [32] E. P. Chassignet, H. E. Hurlbert, O. M. Smedstad, G. R. Halliwell, P. J. Hogan, A. J. Wallcraft, R. Baraille, and R. Bleck, “The HYCOM (hybrid coordinate ocean model) data assimilative system,” *Journal of Marine Systems*, vol. 65, no. 1-4, pp. 60–83, 2007. ISBN: 0924-7963 Publisher: Elsevier.
- [33] J. R. Donoghue, “Univariate Screening Measures for Cluster Analysis,” *Multivariate Behavioral Research*, vol. 30, pp. 385–427, July 1995.
- [34] C. Tebaldi, B. H. Strauss, and C. E. Zervas, “Modelling sea level rise impacts on storm surges along US coasts,” *Environmental Research Letters*, vol. 7, p. 014032, Mar. 2012.
- [35] R. Bintanja, G. J. van Oldenborgh, S. S. Drijfhout, B. Wouters, and C. A. Katsman, “Important role for ocean warming and increased ice-shelf melt in Antarctic sea-ice expansion,” *Nature Geoscience*, vol. 6, pp. 376–379, May 2013.
- [36] B. S. Lecavalier, G. A. Milne, M. J. Simpson, L. Wake, P. Huybrechts, L. Tarasov, K. K. Kjeldsen, S. Funder, A. J. Long, S. Woodroffe, A. S. Dyke, and N. K. Larsen, “A model of Greenland ice sheet deglaciation constrained by observations of relative sea level and ice extent,” *Quaternary Science Reviews*, vol. 102, pp. 54–84, Oct. 2014.
- [37] H. Åkesson, M. Morlighem, K. H. Nisancioglu, J. I. Svendsen, and J. Mangerud, “Atmosphere-driven ice sheet mass loss paced by topography: Insights from modelling the south-western Scandinavian Ice Sheet,” *Quaternary Science Reviews*, vol. 195, pp. 32–47, Sept. 2018.
- [38] N. Lin and E. Shullman, “Dealing with hurricane surge flooding in a changing environment: part I. Risk assessment considering storm climatology change, sea level rise, and coastal development,” *Stochastic Environmental Research and Risk Assessment*, vol. 31, pp. 2379–2400, Nov. 2017.
- [39] The PLOS ONE Staff, “Correction: Future Coastal Population Growth and Exposure to Sea-Level Rise and Coastal Flooding - A Global Assessment,” *PLOS ONE*, vol. 10, p. e0131375, June 2015.

REFERENCES

- [40] M. V. Bilskie, S. C. Hagen, S. C. Medeiros, and D. L. Passeri, “Dynamics of sea level rise and coastal flooding on a changing landscape: BILSKIE ET AL.,” *Geophysical Research Letters*, vol. 41, pp. 927–934, Feb. 2014.
- [41] D. T. Resio and J. L. Irish, “Tropical Cyclone Storm Surge Risk,” *Current Climate Change Reports*, vol. 1, pp. 74–84, June 2015.
- [42] M. Moradi, M. H. Kazeminezhad, and K. Kabiri, “Integration of Geographic Information System and system dynamics for assessment of the impacts of storm damage on coastal communities - Case study: Chabahar, Iran,” *International Journal of Disaster Risk Reduction*, vol. 49, p. 101665, 2020.
- [43] H. Sterr, “Assessment of Vulnerability and Adaptation to Sea-Level Rise for the Coastal Zone of Germany,” *Journal of Coastal Research*, vol. 242, pp. 380–393, Mar. 2008.
- [44] X. Chen, X. Zhang, J. A. Church, C. S. Watson, M. A. King, D. Monselesan, B. Legresy, and C. Harig, “The increasing rate of global mean sea-level rise during 1993–2014,” *Nature Climate Change*, vol. 7, no. 7, pp. 492–495, 2017.
- [45] V. Gornitz, “Sea-level rise: A review of recent past and near-future trends,” *Earth surface processes and landforms*, vol. 20, no. 1, pp. 7–20, 1995.
- [46] E. Cho and J. M. Jacobs, “Extreme value snow water equivalent and snowmelt for infrastructure design over the contiguous United States,” *Water Resources Research*, vol. 56, no. 10, p. e2020WR028126, 2020.
- [47] F. Trotta, I. Federico, N. Pinardi, G. Coppini, S. Causio, E. Jansen, D. Iovino, and S. Masina, “A relocatable ocean modeling platform for downscaling to shelf-coastal areas to support disaster risk reduction,” *Frontiers in Marine Science*, vol. 8, p. 642815, 2021.
- [48] A. Singhal and D. E. Seborg, “Clustering multivariate time-series data,” *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 19, no. 8, pp. 427–438, 2005.
- [49] H. Nazarnia, M. Nazarnia, H. Sarmasti, and W. O. Wills, “A Systematic Review of Civil and Environmental Infrastructures for Coastal Adaptation to Sea Level Rise,” *Civil Engineering Journal*, vol. 6, pp. 1375–1399, July 2020.
- [50] J. A. Church, “How Fast Are Sea Levels Rising?,” *Science*, vol. 294, pp. 802–803, Oct. 2001.

REFERENCES

- [51] S. W. Chang, T. P. Clement, M. J. Simpson, and K.-K. Lee, “Does sea-level rise have an impact on saltwater intrusion?,” *Advances in Water Resources*, vol. 34, pp. 1283–1291, Oct. 2011.
- [52] M. M. Sherif and V. P. Singh, “Effect of climate change on sea water intrusion in coastal aquifers,” *Hydrological processes*, vol. 13, no. 8, pp. 1277–1287, 1999.

Appendix A

=====NOAA DATA COLLECTION THROUGH API=====

```
%=====NOAA DATA COLLECTION THROUGH API=====%
df = pd.DataFrame()
url_new = "https://api.tidesandcurrents.noaa.gov/api/prod/datagetter?
begin_date=20220101&end_date=20221231&station=8741533&product=
hourly_height&datum=STND&time_zone=gmt&units=metric&format=csv"
r = requests.get(url_new)
data_new = r.json
print(data_new)
content_new = r.content
df=pd.read_csv(io.StringIO(content_new.decode('utf-8')))
print(df.head())
df.to_csv("MC_TidalData.csv")
```

Appendix B

=====Python code for SARIMAX=====

```
%=====Python code for SARIMAX=====
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

df = pd.read_csv("/content/sample_data/Site1_GWL.csv")

df = df.drop(["Unnamed: 0"], axis = 1)

df

Date = pd.date_range("2013-01-01", "2022-12-31", freq="D")
print(len(Date))

df.index = Date

df = df.resample("M").mean()

df.plot(subplots = True, figsize = (7,15))

train = df.iloc[:-12,:]
test = df.iloc[-12:,:]

print(train.shape, test.shape)

!pip install pmdarima

import pmdarima as pm
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
model = pm.auto_arima(df['GWLevel'], X = df.drop(["GWLevel"], axis = 1)
                      ,
                      m=12, seasonal=True,
                      start_p=0, start_q=0, max_order=4, test='adf',
                      error_action='ignore',
                      suppress_warnings=True,
                      stepwise=True, trace=True)

result = model.fit(train["GWLevel"], X = train.drop(["GWLevel"], axis = 1))

out = model.plot_diagnostics(figsize = (10,10))

model.summary()

pred = model.predict(n_periods = 12, alpha = 0.05, index =
                     forecast_range, return_conf_int=True)

forecast_range = pd.date_range("2023-01-01", "2023-12-31", freq = "M")
print(len(forecast_range))

lower = pd.Series(pred[1][:,0], index = forecast_range)
upper = pd.Series(pred[1][:,1], index = forecast_range)
print(lower)
print(upper)

print(pred[0])

plt.figure(figsize=(16,9))
plt.plot(pred.index, pred, color = "red", marker = "s")
plt.plot(df.index, df["GWLevel"], color = "green", marker = "d")
plt.plot(pred.index, test["GWLevel"], color = "blue", marker = "o")
plt.grid()
plt.title("Predictions vs Observed (ARIMA model)")
plt.xlabel("Dates")
plt.ylabel("Groundwater Levels in m")
plt.legend(["Predictions", "Train", "Test"], prop = { "size": 20 })

plt.figure(figsize = (20,10))
plt.plot(pd.date_range("2013-01-01", "2022-12-31", freq = "M"),df["GWLevel"], color = "green", marker = "o")
plt.plot(pred[0].index, pd.Series(pred[0]), color = "red", marker = "v")
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
plt.fill_between(forecast_range, lower, upper, alpha = 0.05, color = "k")
plt.legend(["Train", "Forecast", "Confidence Interval"], prop = {"size":20})
plt.grid()

from sklearn.metrics import r2_score
print(r2_score(test["GWLevel"], pred))

from sklearn.metrics import mean_squared_error
print(mean_squared_error(pred, test["GWLevel"], squared = False))

df1 = pd.read_csv("/content/sample_data/Site2_GWL.csv")

df1 = df1.drop(["Unnamed: 0"], axis = 1)
df1.index = Date

df1

df1 = df1.resample("M").mean()

df1.plot(subplots = True, figsize = (7,15))

df1

train1 = df1.iloc[:-12,:]
test1 = df1.iloc[-12:,:]

print(train1.shape, test1.shape)

model1 = pm.auto_arima(df1['GWLevel'], X = df1.drop(["GWLevel"], axis = 1),
                       m=12, seasonal=True,
                       start_p=0, start_q=0, max_order=4, test='adf',
                       error_action='ignore',
                       suppress_warnings=True,
                       stepwise=True, trace=True)

model1.fit(train1["GWLevel"], X = train1.drop(["GWLevel"], axis = 1))

out1 = model1.plot_diagnostics(figsize = (10,10))

model1.summary()
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
pred1 = model1.predict(n_periods = 12, index = forecast_range,
    return_conf_int = True, alpha = 0.05)

lower1 = pd.Series(pred1[1][:,0], index = forecast_range)
upper1 = pd.Series(pred1[1][:,1], index = forecast_range)
print(lower1)
print(upper1)

plt.figure(figsize=(16,9))
plt.plot(pred1.index, pred1, color = "red", marker = "s")
plt.plot(df1.index, df1["GWLevel"], color = "green", marker = "d")
plt.plot(pred1.index, test1["GWLevel"], color = "blue", marker = "o")
plt.grid()
plt.title("Predictions vs Observed (ARIMA model)")
plt.xlabel("Dates")
plt.ylabel("Groundwater Levels in m")
plt.legend(["Predictions", "Train", "Test"], prop = { "size": 20 })

plt.figure(figsize = (20,10))
plt.plot(pd.date_range("2013-01-01", "2022-12-31", freq = "M"),df1["GWLevel"], color = "green", marker = "o")
plt.plot(pred1[0].index, pd.Series(pred1[0]), color = "red", marker = "v")
plt.fill_between(forecast_range, lower1, upper1, alpha = 0.05, color = "k")
plt.legend(["Train", "Forecast", "Confidence Interval"], prop = {"size": 20}, loc = "upper left")
plt.grid()

from sklearn.metrics import r2_score
print(r2_score(test1["GWLevel"], pred1))

from sklearn.metrics import mean_squared_error
print(mean_squared_error(pred1, test1["GWLevel"], squared = False))

"""## **SPECIFIC CONDUCTIVITY**"""

dfsc = pd.read_csv("/content/sample_data/Site1_GWSC.csv")
dfsc = dfsc.drop(["Unnamed: 0"], axis = 1)
dfsc.index = Date

dfsc = dfsc.resample("M").mean()
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
dfsc.plot(subplots = True)

train2 = dfsc.iloc[:-12,:]
test2 = dfsc.iloc[-12:,:]

print(train2.shape, test2.shape)

model2 = pm.auto_arima(dfsc['SCon'], X = dfsc.drop(["SCon"], axis = 1),
                      m=12, seasonal=True,
                      start_p=0, start_q=0, max_order=4, test='adf',
                      error_action='ignore',
                      suppress_warnings=True,
                      stepwise=True, trace=True)

model2.fit(train2["SCon"], X = train2.drop(["SCon"], axis = 1))

pred2 = model2.predict(n_periods = 12, X = test2.drop(["SCon"], axis =
1), alpha = 0.05)

plt.figure(figsize=(16,9))
plt.plot(pred2.index, pred2, color = "red")
plt.plot(dfsc.index, dfsc["SCon"], color= "green")
plt.plot(pred2.index, test2["SCon"], color = "blue")
plt.grid()
plt.title("Predictions vs Observed (ARIMA model)")
plt.xlabel("Dates")
plt.ylabel("Groundwater Specific Conductivity in microSiemens/cm")
plt.legend(["Predictions", "Whole", "Observed"])

from sklearn.metrics import r2_score
print(r2_score(test2["SCon"], pred2))

from sklearn.metrics import mean_squared_error
print(mean_squared_error(pred2, test2["SCon"], squared = False))
```

Appendix C

=====R code for Prophet=====

```
%=====
library(prophet)
library(dplyr)
library(Metrics)

#Reading the datasets
scw1<-read.csv("/Users/pranayjyotimaitra/Desktop/Data/GWSC1.csv")
scw2<-read.csv("/Users/pranayjyotimaitra/Desktop/Data/GWSC2.csv")

#Getting date column
scw1$date = as.POSIXct(strptime(scw1$X, "%Y-%m-%d"))
scw2$date = as.POSIXct(strptime(scw2$X, "%Y-%m-%d"))

#Changing the names to standard
names(scw1) = c("X","Tidal","Prcp","Temp","GWL","y", "ds")
names(scw2) = c("X","Tidal","Prcp","Temp","GWL","y", "ds")

#Removing redundant columns
scw1 = subset(scw1, select = -c(X) )
scw2 = subset(scw2, select = -c(X) )

#Train-Test Split
scw1_train<-scw1[1:3287,c(1,2,3,4,5,6)]
scw1_test<-scw1[3288:3652,c(1,2,3,4,5,6)]
scw2_train<-scw2[1:3287,c(1,2,3,4,5,6)]
scw2_test<-scw2[3288:3652,c(1,2,3,4,5,6)]

#Site 1 future predictions and regressors
m1<-prophet(yearly.seasonality = TRUE, weekly.seasonality = TRUE,
              daily.seasonality = TRUE)
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
m1<-add_regressor(m1, "Tidal", mode = "additive")
m1<-add_regressor(m1, "Prcp", mode = "additive")
m1<-add_regressor(m1, "Temp", mode = "additive")
m1<-add_regressor(m1, "GWL", mode = "additive")
m1<-fit.prophet(m1,scw1_train)
future1 <- make_future_dataframe(m1, periods = 365)
future1$Tidal <- scw1$Tidal
future1$Prcp <- scw1$Prcp
future1$Temp<-scw1$Temp
future1$GWL<-scw1$GWL
forecast1 <- predict(m1, future1)

#Site 2 future predictions and regressors
m2<-prophet(yearly.seasonality = TRUE, weekly.seasonality = TRUE, daily
              .seasonality = TRUE)
m2<-add_regressor(m2, "Tidal", mode = "additive")
m2<-add_regressor(m2, "Prcp", mode = "additive")
m2<-add_regressor(m2, "Temp", mode = "additive")
m2<-add_regressor(m2, "GWL", mode = "additive")
m2<-fit.prophet(m2,scw2_train)
future2 <- make_future_dataframe(m2, periods = 365)
future2$Tidal <- scw2$Tidal
future2$Prcp <- scw2$Prcp
future2$Temp<-scw2$Temp
future2$GWL<-scw2$GWL
forecast2 <- predict(m2, future2)

#Plotting predictions
plot(m1, forecast1, xlabel = "Dates", ylabel = "GW Specific
      Conductivity")
plot(m2, forecast2, xlabel = "Dates", ylabel = "GW Specific
      Conductivity")

#Plotting observed vs predicted in the test series
plot(forecast1$ds[3288:3652], forecast1$yhat[3288:3652], type = "l",
      col = "red", xlab = "Dates", ylab = "GW Specific Conductivity")
lines(forecast1$ds[3288:3652], scw1_test$y, col = "blue")
legend(x = "bottomright", legend=c("Predicted", "Observed"), fill = c(
      "red", "blue"))

plot(forecast2$ds[3288:3652], forecast2$yhat[3288:3652], type = "l",
      col = "red", xlab = "Dates", ylab = "GW Specific Conductivity")
lines(forecast2$ds[3288:3652], scw2_test$y, col = "blue")
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
legend(x = "bottomright", legend=c("Predicted", "Observed"), fill = c("red", "blue"))

df1.cv <- cross_validation(m1, initial=365.25, period=60, horizon=120,
    units='days')
df2.cv <- cross_validation(m2, initial=365.25, period=60, horizon=120,
    units='days')
plot_cross_validation_metric(df1.cv, metric='rmse')
plot_cross_validation_metric(df2.cv, metric='rmse')
prophet_plot_components(m1, forecast1)
prophet_plot_components(m2, forecast2)
plot(sc1$y~sc1$ds, type = "l")
plot(sc2$y~sc2$ds, type = "l")
metric_df1<-merge(x = sc1_test, y = forecast1, by = "ds", all = TRUE)
metric_df2<-merge(x = sc2_test, y = forecast2, by = "ds", all = TRUE)
plot(forecast2[3288:3652,"yhat"]~sc2_test$y, type = "l")
r2_df1<-data.frame(forecast1[3288:3652,"yhat"], sc1_test$y)

#Metrics scores
#Testing
rsq <- function (x, y) cor(x, y) ^ 2
R21<-rsq(forecast1[3288:3652,"yhat"],scw1_test$y)
R22<-rsq(forecast2[3288:3652,"yhat"],scw2_test$y)
rmse1<-rmse(forecast1[3288:3652,"yhat"],scw1_test$y)
rmse2<-rmse(forecast2[3288:3652,"yhat"],scw2_test$y)
#Training
R21t<-rsq(forecast1[1:3287,"yhat"],scw1_train$y)
R22t<-rsq(forecast2[1:3287,"yhat"],scw2_train$y)
rmse1t<-rmse(forecast1[1:3287,"yhat"],scw1_train$y)
rmse2t<-rmse(forecast2[1:3287,"yhat"],scw2_train$y)
```

Appendix D

=====Initial Data analysis and visualisation=====

```
=====Initial data analysis and visualisation=====

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

TMAX_ELC = pd.read_csv("TMAX_Daily_2022_ELC.csv", header=None)
TMAX_MC = pd.read_csv("TMAX_Daily_2022_MC.csv", header=None)
TMAX_BB = pd.read_csv("TMAX_Daily_2022_BB.csv", header=None)

Dates = pd.date_range("2022-01-01", "2022-12-31", freq = "1D")

TMIN_ELC = pd.read_csv("TMIN_Daily_2022_ELC.csv", header=None)
TMIN_MC = pd.read_csv("TMIN_Daily_2022_MC.csv", header=None)
TMIN_BB = pd.read_csv("TMIN_Daily_2022_BB.csv", header=None)

Temp_ELC = (TMAX_ELC+TMIN_ELC)/2
Temp_MC = (TMAX_MC+TMIN_MC)/2
Temp_BB = (TMAX_BB+TMIN_BB)/2

plt.plot(Dates, TMAX_ELC)
plt.plot(Dates, Temp_ELC)
plt.plot(Dates, TMIN_ELC)
plt.xlabel("Dates")
plt.ylabel("Daily Temperatures (Kelvin)")
plt.legend(["Maximum Temperature", "Mean Temperature", "Minimum
Temperature"])
plt.title("Eastern Louisiana Coastal Watershed HUC-08090203")
plt.savefig("ELC_TEMP.png")
plt.show()
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
plt.plot(Dates, TMAX_MC)
plt.plot(Dates, Temp_MC)
plt.plot(Dates, TMIN_MC)
plt.xlabel("Dates")
plt.ylabel("Daily Temperatures (Kelvin)")
plt.legend(["Maximum Temperature", "Mean Temperature", "Minimum Temperature"])
plt.title("Mississippi Coastal Watershed HUC-03170009")
plt.savefig("MC_TEMP.png")
plt.show()

plt.plot(Dates, TMAX_BB)
plt.plot(Dates, Temp_BB)
plt.plot(Dates, TMIN_BB)
plt.xlabel("Dates")
plt.ylabel("Daily Temperatures (Kelvin)")
plt.legend(["Maximum Temperature", "Mean Temperature", "Minimum Temperature"])
plt.title("Bulls Bay Watershed HUC-03050209")
plt.savefig("BB_TEMP.png")
plt.show()

Temp_ELC.to_csv("Temperature_Dataset_ELC.csv")
Temp_MC.to_csv("Temperature_Dataset_MC.csv")
Temp_BB.to_csv("Temperature_Dataset_BB.csv")

Prec_ELC = pd.read_csv("Prec_Daily_2022_ELC.csv", header=None)
Prec_MC = pd.read_csv("Prec_Daily_2022_MC.csv", header=None)
Prec_BB = pd.read_csv("Prec_Daily_2022_BB.csv", header=None)

plt.plot(Dates, Prec_ELC[0])
plt.xlabel("Dates")
plt.ylabel("Daily Precipitation (mm)")
plt.title("Eastern Louisiana Coastal Watershed HUC-08090203")
plt.savefig("ELC_PREC.png")
plt.show()

plt.plot(Dates, Prec_MC[0])
plt.xlabel("Dates")
plt.ylabel("Daily Precipitation (mm)")
plt.title("Mississippi Coastal Watershed HUC-03170009")
plt.savefig("MC_PREC.png")
plt.show()
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
plt.plot(Dates, Prec_BB[0])
plt.xlabel("Dates")
plt.ylabel("Daily Precipitation (mm)")
plt.title("Bulls Bay Watershed HUC-03050209")
plt.savefig("BB_PREC.png")
plt.show()

plt.plot(Dates, Prec_ELC[0])
plt.plot(Dates, Prec_MC[0])
plt.plot(Dates, Prec_BB[0])
plt.xlabel("Dates")
plt.ylabel("Daily Precipitation (mm)")
plt.legend(["HUC-08090203", "HUC-03170009", "HUC-03050209"])
plt.title("Comparison of precipitation for the three watersheds")
plt.savefig("ALL_3_PREC.png")
plt.show()

GW_Level = pd.read_csv("GW_Level.csv")
GW_Salinity = pd.read_csv("GW_Quality.csv")

levels_1 = GW_Level[GW_Level["site_no"]==300722089150100]
levels_2 = GW_Level[GW_Level["site_no"]==301001089442600]
levels_3 = GW_Level[GW_Level["site_no"]==301527088521500]
levels_4 = GW_Level[GW_Level["site_no"]==301912088583300]
levels_5 = GW_Level[GW_Level["site_no"]==330428079214800]

from datetime import datetime
dates = []
for x in levels_1.Date:
    dates.append(datetime.strptime(x, "%Y-%m-%d"))
levels_1["Dates_Act"] = dates
levels_1 = levels_1.groupby(pd.Grouper(key='Dates_Act', axis=0, freq='D')).mean()

plt.plot(levels_1.X_00065_00003)
plt.xlabel("Dates")
plt.ylabel("GW_Levels")
plt.title("Site - 300722089150100")
plt.savefig("Levels_1.png")
plt.show()

dates = []
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
for x in levels_2.Date:
    dates.append(datetime.strptime(x, "%Y-%m-%d"))
levels_2["Dates_Act"] = dates
levels_2 = levels_2.groupby(pd.Grouper(key='Dates_Act', axis=0, freq='D')).mean()

plt.plot(levels_2.X_00065_00003)
plt.xlabel("Dates")
plt.ylabel("GW Levels")
plt.title("Site - 301001089442600")
plt.savefig("Levels_2.png")
plt.show()

dates = []
for x in levels_3.Date:
    dates.append(datetime.strptime(x, "%Y-%m-%d"))
levels_3["Dates_Act"] = dates
levels_3 = levels_3.groupby(pd.Grouper(key='Dates_Act', axis=0, freq='D')).mean()

plt.plot(levels_3.X_00065_00003)
plt.xlabel("Dates")
plt.ylabel("GW Levels")
plt.title("Site - 301527088521500")
plt.savefig("Levels_3.png")
plt.show()

dates = []
for x in levels_5.Date:
    dates.append(datetime.strptime(x, "%Y-%m-%d"))
levels_5["Dates_Act"] = dates
levels_5 = levels_5.groupby(pd.Grouper(key='Dates_Act', axis=0, freq='D')).mean()

plt.plot(levels_5.X_00065_00003)
plt.xlabel("Dates")
plt.ylabel("GW Levels")
plt.title("Site - 330428079214800")
plt.savefig("Levels_5.png")
plt.show()

dates = []
for x in sal_1.Date:
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
dates.append(datetime.strptime(x, "%Y-%m-%d"))
sal_1["Dates_Act"] = dates
sal_1 = sal_1.groupby(pd.Grouper(key='Dates_Act', axis=0, freq='D')).mean()

plt.plot(sal_1.X_00095_00003)
plt.xlabel("Dates")
plt.ylabel("Specific Conductivity")
plt.title("Site - 300722089150100")
plt.savefig("Sal_1.png")
plt.show()

dates = []
for x in sal_2.Date:
    dates.append(datetime.strptime(x, "%Y-%m-%d"))
sal_2["Dates_Act"] = dates
sal_2 = sal_2.groupby(pd.Grouper(key='Dates_Act', axis=0, freq='D')).mean()

plt.plot(sal_2.X_00095_00003)
plt.xlabel("Dates")
plt.ylabel("Specific Conductivity")
plt.title("Site - 301001089442600")
plt.savefig("Sal_2.png")
plt.show()

dates = []
for x in sal_3.Date:
    dates.append(datetime.strptime(x, "%Y-%m-%d"))
sal_3["Dates_Act"] = dates
sal_3 = sal_3.groupby(pd.Grouper(key='Dates_Act', axis=0, freq='D')).mean()

plt.plot(sal_3.X_00095_00003)
plt.xlabel("Dates")
plt.ylabel("Specific Conductivity")
plt.title("Site - 301527088521500")
plt.savefig("Sal_3.png")
plt.show()

dates = []
for x in sal_5.Date:
    dates.append(datetime.strptime(x, "%Y-%m-%d"))
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
sal_5["Dates_Act"] = dates
sal_5 = sal_5.groupby(pd.Grouper(key='Dates_Act', axis=0, freq='D')).mean()

plt.plot(sal_5.X_00095_00003)
plt.xlabel("Dates")
plt.ylabel("Specific Conductivity")
plt.title("Site - 330428079214800")
plt.savefig("Sal_5.png")
plt.show()

levels_1 = levels_1.reset_index()
levels_2 = levels_2.reset_index()
levels_3 = levels_3.reset_index()
levels_5 = levels_5.reset_index()

df_levels = pd.merge(levels_1, levels_3, on = "Dates_Act", how = "inner")
df_levels = pd.merge(df_levels, levels_5, on = "Dates_Act", how = "inner")
df_levels = pd.merge(df_levels, levels_2, on = "Dates_Act", how = "left")
DF_Levels = df_levels.iloc[:,[0,3,6,9,12]]
DF_Levels = DF_Levels.set_axis(["Dates", "300722089150100",
                                "301527088521500", "330428079214800", "301001089442600"], axis = 1,
                                inplace = False)

levels_temp = DF_Levels.iloc[:,[1,2,3,4]]
from sklearn.impute import KNNImputer
imputer = KNNImputer(n_neighbors=2)
levels_temp_imputed = imputer.fit_transform(levels_temp)

date = pd.date_range("2022-01-01", "2022-12-31", freq = "1D")

final_gwlevels = pd.DataFrame(levels_temp_imputed)
final_gwlevels["Dates"] = date
final_gwlevels = final_gwlevels.set_axis(["300722089150100",
                                         "301527088521500", "330428079214800", "301001089442600", "Dates"],
                                         axis = 1, inplace = False)

plt.plot(final_gwlevels.Dates, final_gwlevels.iloc[:,0])
plt.plot(final_gwlevels.Dates, final_gwlevels.iloc[:,1])
plt.plot(final_gwlevels.Dates, final_gwlevels.iloc[:,2])
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
plt.plot(final_gwlevels.Dates, final_gwlevels.iloc[:,3])
plt.xlabel("Dates")
plt.ylabel("Groundwater Levels")
plt.title("Groundwater Levels of the 4 sites.")
plt.legend(["300722089150100", "301527088521500", "330428079214800",
           "301001089442600"])
plt.savefig("ALL_4_GW_SITES.png")
plt.show()

final_gwlevels.to_csv("finalGroundwaterLevelsForFourSites.csv")

sal_1 = sal_1.reset_index()
sal_2 = sal_2.reset_index()
sal_3 = sal_3.reset_index()
sal_5 = sal_5.reset_index()

df_sal = pd.merge(sal_2, sal_5, on = "Dates_Act", how = "inner")
df_sal = pd.merge(df_sal, sal_1, on = "Dates_Act", how = "left")
df_sal = pd.merge(df_sal, sal_3, on = "Dates_Act", how = "left")

df_sal = df_sal.iloc[:,[0,3,6,9,12]]
df_sal = df_sal.set_axis(["Dates", "301001089442600",
                         "330428079214800", "300722089150100", "301527088521500"], axis = 1,
                         inplace = False)

sal_temp = df_sal.iloc[:,[1,2,3,4]]

sal_temp_imputed = imputer.fit_transform(sal_temp)

final_conductivity = pd.DataFrame(sal_temp_imputed)
final_conductivity["Dates"] = date
final_conductivity = final_conductivity.set_axis(["301001089442600",
                                                 "330428079214800", "300722089150100", "301527088521500", "Dates"],
                                                 axis = 1, inplace = False)

plt.plot(final_conductivity.Dates, final_conductivity.iloc[:,0])
plt.plot(final_conductivity.Dates, final_conductivity.iloc[:,1])
plt.plot(final_conductivity.Dates, final_conductivity.iloc[:,2])
plt.plot(final_conductivity.Dates, final_conductivity.iloc[:,3])
plt.xlabel("Dates")
plt.ylabel("Specific Conductivity")
plt.title("Salinity Measurements of all 4 sites.")
plt.legend(final_conductivity.columns[0:4])
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
plt.savefig("ALL_4_GWQ_SITES.png")
plt.show()

final_conductivity.to_csv("finalSpecificConductivityOfFourSites.csv")

ELC_Hourly_WL = pd.read_csv("ELC_TidalData.csv")
MC_Hourly_WL = pd.read_csv("MC_TidalData.csv")
BB_Hourly_WL = pd.read_csv("BB_TidalData.csv")

ELC_Hourly_WL['Dates'] = pd.to_datetime(ELC_Hourly_WL['Date Time']).dt.
    date
ELC_Hourly_WL['Time'] = pd.to_datetime(ELC_Hourly_WL['Date Time']).dt.
    time

MC_Hourly_WL['Dates'] = pd.to_datetime(MC_Hourly_WL['Date Time']).dt.
    date
MC_Hourly_WL['Time'] = pd.to_datetime(MC_Hourly_WL['Date Time']).dt.
    time

BB_Hourly_WL['Dates'] = pd.to_datetime(BB_Hourly_WL['Date Time']).dt.
    date
BB_Hourly_WL['Time'] = pd.to_datetime(BB_Hourly_WL['Date Time']).dt.
    time

ELC_Hourly_WL = ELC_Hourly_WL.groupby(ELC_Hourly_WL["Dates"]).mean()
MC_Hourly_WL = MC_Hourly_WL.groupby(MC_Hourly_WL["Dates"]).mean()
BB_Hourly_WL = BB_Hourly_WL.groupby(BB_Hourly_WL["Dates"]).mean()

ELC_Hourly_WL = ELC_Hourly_WL.drop(["Unnamed: 0", "Sigma", "I", "L"],
    axis = 1)
MC_Hourly_WL = MC_Hourly_WL.drop(["Unnamed: 0", "Sigma", "I", "L"],
    axis = 1)
BB_Hourly_WL = BB_Hourly_WL.drop(["Unnamed: 0", "Sigma", "I", "L"],
    axis = 1)

ELC_Hourly_WL = ELC_Hourly_WL.reset_index()
MC_Hourly_WL = MC_Hourly_WL.reset_index()
BB_Hourly_WL = BB_Hourly_WL.reset_index()

plt.plot(ELC_Hourly_WL.Dates, ELC_Hourly_WL[" Water Level"])
plt.plot(MC_Hourly_WL.Dates, MC_Hourly_WL[" Water Level"])
plt.plot(BB_Hourly_WL.Dates, BB_Hourly_WL[" Water Level"])
plt.xlabel("Dates")
```

PREDICTING SALTWATER INTRUSION IN COASTAL AQUIFERS BY DATA-DRIVEN MODELING

```
plt.ylabel("Surface Water Levels")
plt.title("Surface water levels of the three watersheds from nearby
NOAA stations")
plt.legend(["Eastern Louisiana Coastal", "Mississippi Coastal", "Bulls
Bay"])
plt.grid()
plt.savefig("Watersheds'_WL.png")
plt.show()

ELC_Hourly_WL.to_csv("ELC_Daily_2022_WL.csv")
MC_Hourly_WL.to_csv("MC_Daily_2022_WL.csv")
BB_Hourly_WL.to_csv("BB_Daily_2022_WL.csv")
```

Appendix E

```
=====R code for retrieving GW levels and quality=====
=====R code for retrieving GW levels and quality=====

library(dataRetrieval)
vignette("dataRetrieval", package = "dataRetrieval")
sites<-c
  ("300722089150100", "301001089442600", "301527088521500", "301912088583300", "330428079

startDate<-as.Date("2022-01-01")
endDate<-as.Date("2022-12-31")
datagwl<-readNWISdv(sites,parameterCd="00065",startDate = startDate,
  endDate = endDate)
datagwq<-readNWISdv(sites,parameterCd="00095",startDate = startDate,
  endDate = endDate)

write.csv(datagwl, "/Users/pranayjyotimaitra/Desktop/Data/GW_Level.csv
  ")
write.csv(datagwq, "/Users/pranayjyotimaitra/Desktop/Data/GW_Quality.
  csv")
```

GROUNDWATER VARIABILITY ALONG WITH THE VARYING COASTLINE OF THIRUVANATHAPURAM CITY

¹PRALAY SANKAR MAITRA, ²REMYA R, ³ ALKA SINGH

^{1,2,3}GeoInformatics and Earth Observation, Amrita Centre for Wireless Network and Applications, Amritapuri, Kerala, India
Email: ¹pralay.sankar95@gmail.com, ²remya95ravikumar@gmail.com, ²alka228@gmail.com

Abstract: Shoreline change is a constantly evolving phenomenon that threatens people and their livelihoods around the globe. India observes this phenomenon strongly at different locations being a tropical peninsular country with 6635kms of coastline. This study tried to analyze the effect of shoreline change on surrounding ground water reserves along Thiruvananthapuram coast in Kerala district of India. Net changes in coastline positions are statistically calculated and observed using Linear Regression Rate (LRR). The shoreline change rate shows most of the region are undergoing erosion, only few accretions or land formation are observed which was mostly artificially formed due to harbor building. The highest erosion rate in terms of LRR was -7m/year and highest accretion was 28m/year. We compared the shoreline change with groundwater variability along the coast and also groundwater salinity using electrical conductivity as the factor. The study observed decreasing trend of groundwater along the eroding coastline. The study also predicted decadal shoreline change along the Thiruvananthapuram coast was predicted using Kalman filter model.

Keywords: Shoreline change rate, LRR, Kalman filter model, Ground water.

1. INTRODUCTION

India has a coastline of about 6635 kilometers long with the Arabian Sea to the west Bay of Bengal to the east and the Indian Ocean to the south which is dotted with several major ports on both its west and east coastlines. According to a report published by National Centre for Coastal Research (NCCR) as much as 32 percent of the coastline has undergone erosion majority of which occurred between 1990 and 2018.

One of the major hazards which accompanies any coastline is its erosion. In order to study the effect of shoreline change on an urban settlement and its subsequent effect on the ground water in the region we chose Thiruvananthapuram coast in Kerala State which is on the western coast of India as studies on the western coast especially towards the southern part of the country was limited. The study focuses on the shoreline change and its implication on ground water (GW) along the Thiruvananthapuram coast for a period from 2002 to 2022 at 4 years interval using Landsat 7/8 (2002-2014) and Sentinel-2 (2018 and 2022). Digital Shoreline Analysis System tool (DSAS) of ArcGIS Desktop was used to understand shoreline change rates using statistical measure like the Linear Regression Rates (LRR). To classify the shoreline, coastal vulnerability index (CVI) is calculated using the LRR, geomorphology data, mean sea-level, tidal range, mean wave height and slope. The result shows major erosion in two locations and a landmass formation owing to the creation of artificial coast for harbor construction.

The change in shoreline for the year 2032 was forecasted using the Kalman filter model and finally, the effect of shoreline changes on groundwater trend

and salinity near to coast was observed using data obtained from India Water Resources Information System (India WRIS).

2. STUDY AREA

The 78 kilometers long study area stretching along the shore of Arabian Sea includes three taluks (Chirayankizhu, Neyyatinkara and Thiruvananthapuram) of Thiruvananthapuram district of Kerala which is situated between 8.17° – 8.54° latitude and 76.41° – 77.17° longitude. The district is mainly divided into three main geographical regions namely highlands, midlands and low lands, the Chirayankizhu and Thiruvananthapuram taluks fall under the mid to low land category whereas Neyyatinkara stretches among all the three.

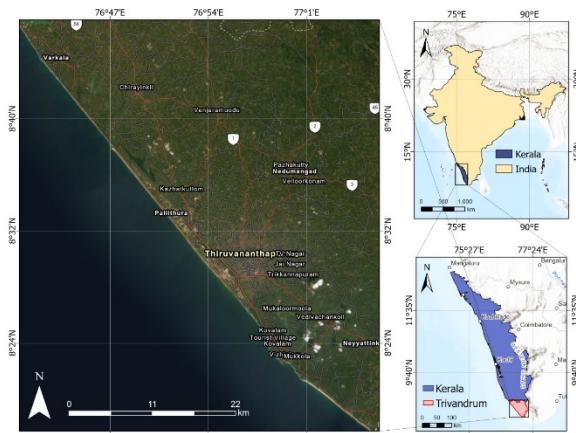


Figure 1: Map showing the study area: Thiruvananthapuram coastline along the Thiruvananthapuram district of Kerala state in India

3. DATA AND METHODOLOGY

3.1. Shoreline extraction

The Satellite images for the region was obtained from Landsat 7 (2002, 2006 and 2010), Landsat 8 (2014) and Sentinel 2 (2018 and 2022), which were collected from USGS Earth Explorer, these where preprocessed as Landsat 7 images had scan line error which was rectified using ‘Landsat gap fill’ tool of ENVI and sentinel data needed to be mosaiced using ‘mosaic to new raster’ tool in ArcGIS Pro. The shorelines were then extracted using various geoprocessing tools in ArcGIS Pro as shown in the figure 2.

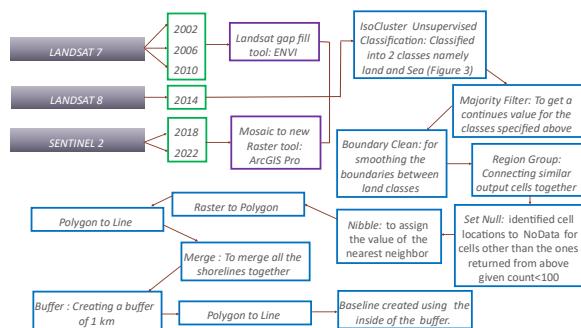


Figure 2: Shoreline extraction process, step by step guide.

All the individual shorelines extracted are merged together using merge tool in ArcGIS Pro and the Baseline created by providing a buffer on the merged shoreline for 1 km distance and taking the inside of the buffer as the baseline.

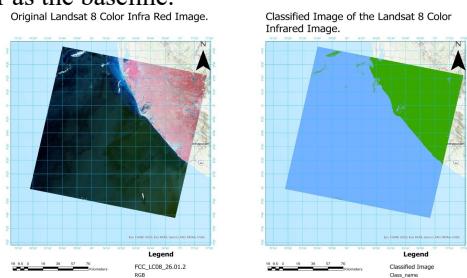


Figure 3: IsoCluster Unsupervised Classification from where Shoreline is extracted in between the two polygons of land and sea. The remaining geo processing tools shown in figure2 is used for further cleaning these two classes and to finally extract the shoreline.

3.2. Coastal Vulnerability Index (CVI)

The CVI index is calculated using DSAS tool in ArcGIS desktop, for the purpose of creating the index, transects were formed from the baseline obtained. In the attribute table for shoreline and baseline more fields (such as ID, DATE, GROUP, SEARCH DISTANCE, UNCERTAINTY, etc.) were added as documented in the DSAS user manual. Following which the LRR, EPR (End point rate), WLR (Weighted Linear Regression), SCE (Shoreline Change Envelope) and NSM (Net Shoreline Movement) were calculated. Among these only LRR

was considered for shoreline change rate calculation for our study.

The equation used for CVI index is:

$$CVI = \frac{\sqrt{a*b*c*d*e*f}}{6} \quad \dots \dots \dots \text{(eq1)}$$

Where a is geomorphology (data obtained from Bhukosh), b is slope (from SRTM DEM 30m), c is mean sea level rise (from Intergovernmental Panel on Climate Change (IPCC) report), d is mean wave height (1.1m), e is tidal range (from WXTide: 0.45-0.65 meters, these values were taken by averaging the tide height between 10-12 AM for each date corresponding to the Landsat/Sentinel image) and f is LRR or shoreline change. All these parameters were individually classified into three classes (high, medium and low risks) and finally CVI found using the equation (eq1)

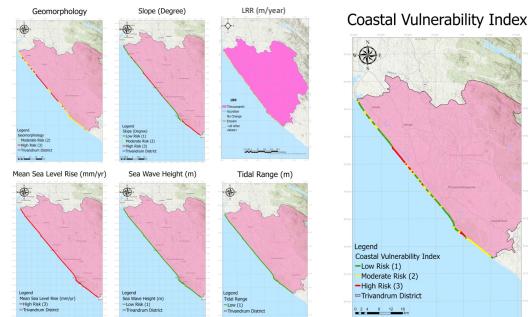


Figure 4: Clock wise from left to right: The geomorphology, slope, LRR, CVI Index, Mean Sea level rise, Sea Wave height and tidal range classified under the three main classes (High, medium and Low Risks).

3.3. Shoreline change prediction

The decadal shoreline change was predicted using a forecasting method based on Kalman Filter model in the DSAS tool.

The formula used for predicting the shoreline is:

where s' is the predicted shoreline for year 2032, s is the shoreline for year 2022, scr is the shoreline change rate (obtained from LRR along the shoreline) and is time gap between the years (10 in this case).

3.4. Ground water Variability along shoreline

Ground water level was obtained from India WRIS for 21 well points (Figure 5) along the shoreline for determining its variability with respect to varying coastline. Out of which some points were omitted as we only took those points which had negative LRR values, the physical implication of which is we just studied groundwater variability with only erosion and not accretion. The water level for these well points were observed for a period of 20 years (2002-2022), the missing data were populated with the average of observations. The new average for each well point and

corresponding LRR of the shoreline were used to create a linear regression model for determining the effect of shoreline change on ground water near the coast.

3.4. Ground water salinity along shoreline

Electrical conductivity of 7 well points (the points were chosen based on the availability of data for the time period) were obtained from India WRIS for studying the effect of shoreline on groundwater salinity in the region (Figure 5). Electrical conductivity was chosen as a parameter for determining the salinity as it is a direct factor which shows for the presence of ions in water. A linear regression model was applied with well points and electrical conductivity as parameters for determining the correlation between the two.

Groundwater Well Points - Electrical Conductivity and Depth

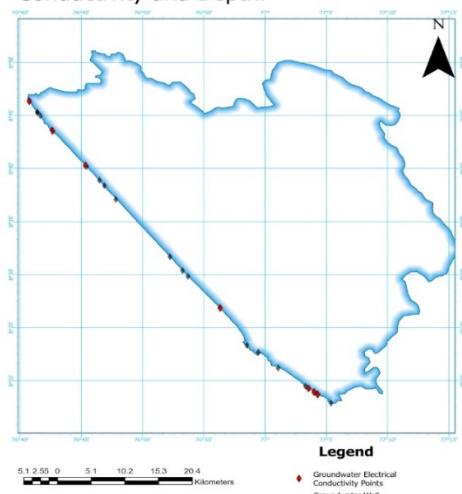


Figure 5: well points along the coastline where the black colored shows the well points considered for ground water level and the red colored ones are considered for electrical conductivity.

4. RESULTS AND DISCUSSION

4.1. Shoreline Change

The shoreline change rate shows most of the region are undergoing erosion only few accretions or land formation were observed. The highest erosion rate in terms of LRR was -7m/year and highest accretion was 28m/year.

From figure 4 it is observed that high rates of erosion are present in Shangumugham and Pozhikkara beaches and also among Varkala and Anchuthengu coasts. Instances of accretion mainly along the Vizhinjam coast due to artificial land mass formation for the upcoming International Sea port construction.

Shoreline Change Rate m/yr

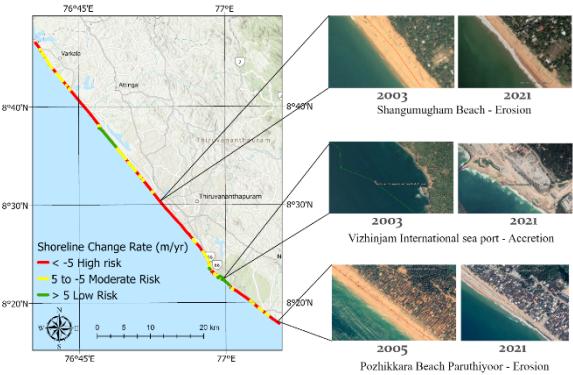


Figure 6: The map shows shoreline change for the past decade and how the shoreline changed in the major regions. Here the red color denotes regions of high risk (erosion), green for low (accretion) and yellow for moderate risk.

4.2. Decadal Prediction

Kalman filter model predicted the decadal shoreline as following the same trend as observed in the previous decade (2002-2022) with regions of high erosion continuing to erode and vice versa.

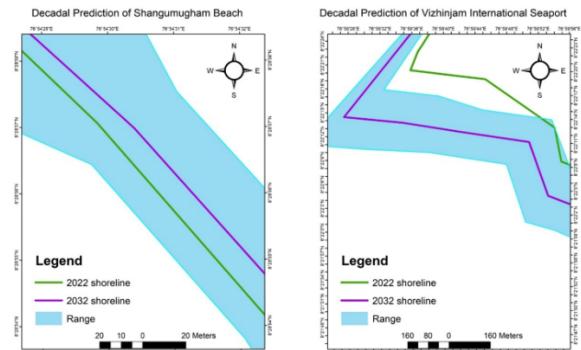


Figure 7: from left to right: Shangumugham beach showing erosion tendencies and next figure shows Vizhinjam coast which shows accretion tendencies. The green color depicts 2022 shoreline and the purple color 2032 shoreline and the cyan color shaded portion shows the range polygon within which the decadal shoreline could lie.

4.3. Groundwater Variability and Salinity

The plot between shoreline change rate and ground water depth (Figure 8) shows a linear relationship which implies that there is an increase in ground water level in areas of erosion and decrease in areas of accretion. This may be due to the sea actually merging with the aquifers deep below the earth.

The plot obtained from shoreline change rate and electrical conductivity (Figure 9) of selected well points were observed to be linear which implies that the conductivity (which is a property of salinity) of ground water increase with increase in shoreline change. This could be due to the intrusion of saline sea water into the GW table during erosion which causes the ion concentration in GW to rise.

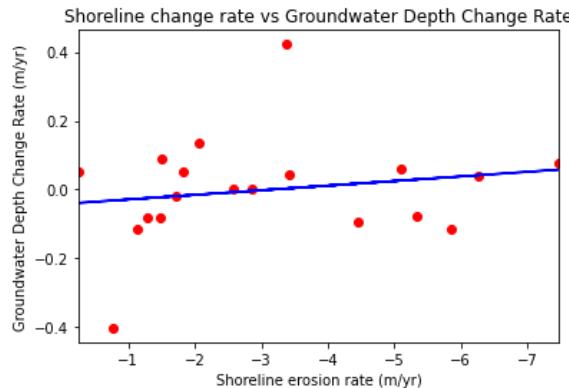


Figure 8: Regression line showing a linear relationship between shoreline change rate and ground water depth.

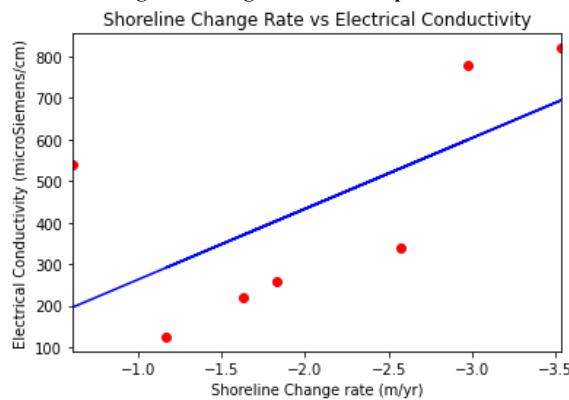


Figure 9: Regression line showing a linear relationship between shoreline change rate and electrical conductivity which is a direct measure of salinity.

CONCLUSIONS

The current study helps us to conclude that the highest shoreline change rate for erosion and accretion are - 7m/year and 28m/year respectively in terms of LRR for the study region. The decadal shoreline prediction showed that the coastline mostly showed erosional tendencies apart from isolated section like Vizhinjam International Sea port region where artificial land mass formation was taking place.

The study further showcased linear relationship between ground water depth and shoreline change rate and also between shoreline change and GW electrical conductivity (salinity). This may be dependent on the population allocation along the shorelines or due to the sea actually merging with the aquifers deep below the earth or a combination of both the parameters.

REFERENCES

1. M. Mondal, S. Haldar, A. Biswas, S. Mandal, S. Bhattacharya, and S. Paul, "Modeling cyclone-induced multi-hazard risk assessment using analytical hierarchical processing and GIS for coastal West Bengal, India," *Reg. Stud. Mar. Sci.*, vol. 44, p. 101779, May 2021, doi: 10.1016/j.rsma.2021.101779.
2. L. R. Vieira, J. G. Vieira, I. M. da Silva, E. Barbieri, and F. Morgado, "GIS Models for Vulnerability of Coastal Erosion Assessment in a Tropical Protected Area," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 9, p. 598, Sep. 2021, doi: 10.3390/ijgi10090598.
3. I. Boateng, "An application of GIS and coastal geomorphology for large scale assessment of coastal erosion and management: a case study of Ghana," *J. Coast. Conserv.*, vol. 16, no. 3, pp. 383–397, Sep. 2012, doi: 10.1007/s11852-012-0209-0.
4. Q. Liu, L. Liang, X. Yuan, X. Mou, and L. Su, "Effects of Groundwater Level Changes Associated with Coastline Changes in Coastal Wetlands," *Wetlands*, vol. 40, no. 5, pp. 1647–1656, Oct. 2020, doi: 10.1007/s13157-019-01253-9.
5. J. Shaji, "Coastal sensitivity assessment for Thiruvananthapuram, west coast of India," *Nat. Hazards*, vol. 73, no. 3, pp. 1369–1392, Sep. 2014, doi: 10.1007/s11069-014-1139-y.
6. T. Kuleli, A. Guneroglu, F. Karsli, and M. Dihkan, "Automatic detection of shoreline change on coastal Ramsar wetlands of Turkey," *Ocean Eng.*, vol. 38, no. 10, pp. 1141–1149, Jul. 2011, doi: 10.1016/j.oceaneng.2011.05.006.
7. Md. R. A. Mullick, K. M. A. Islam, and A. H. Tanim, "Shoreline change assessment using geospatial tools: a study on the Ganges deltaic coast of Bangladesh," *Earth Sci. Inform.*, vol. 13, no. 2, pp. 299–316, Jun. 2020, doi: 10.1007/s12145-019-00423-x.
8. V. Kotinas, N. Evelpidou, and A. Karkani, "Modelling Coastal Erosion," p. 42.
9. L. Natarajan et al., "Shoreline changes over last five decades and predictions for 2030 and 2040: a case study from Cuddalore, southeast coast of India," *Earth Sci. Inform.*, vol. 14, no. 3, pp. 1315–1325, Sep. 2021, doi: 10.1007/s12145-021-00668-5.
10. S. Goswami et al., "Assessment of Shoreline Changes and the Groundwater Quality along the Coast of Kuakata, Patuakhali, Bangladesh," *J. Ecol. Eng.*, vol. 23, no. 7, pp. 323–332, Jul. 2022, doi: 10.12911/22998993/149938.
11. M. M. R. A. M. and V. P., "Coastal vulnerability assessment of Puducherry coast, India, using the analytical hierarchical process," *Nat. Hazards Earth Syst. Sci.*.
12. B. R. Tripathy, S. Kaliraj, K. K. Ramachandran, P. Kumar, and G. S. Pippal, "Assessing relative coastal vulnerability to sea level rise along the Thiruvananthapuram coast, Kerala using GIS-based CVI model," presented at the DISASTER RISK AND VULNERABILITY CONFERENCE 2017, Trivandrum.
13. J. W. Long and N. G. Plant, "Extended Kalman Filter framework for forecasting shoreline evolution: FORECASTING SHORELINE EVOLUTION," *Geophys. Res. Lett.*, vol. 39, no. 13, p. n/a-n/a, Jul. 2012, doi: 10.1029/2012GL052180.
14. D. Ciritci and T. Türk, "Assessment of the Kalman filter-based future shoreline prediction method," *Int. J. Environ. Sci. Technol.*, vol. 17, no. 8, pp. 3801–3816, Aug. 2020, doi: 10.1007/s13762-020-02733-w.

