DOCUMENT CONTROL SHEET

Earth System Science

Organization (ESSO)Ministry of

Earth Sciences (MoES)

Indian National Centre for Ocean Information Services (INCOIS)

ESSO Document Number: ESSO-INCOIS-ODM-TR-01(2024)
Title of the report: INCOIS- Delayed Mode Tide Gauge Data Processing and Quality Control Procedure Over North Indian Ocean.
Author(s) [Last name, First name]:
P.Suneeta, TVS Udaya Bhaskar, E. Pattabhi Rama Rao, K.Srinivas and V. Ravichandra
Originating unit
Ocean Data Management Division (ODM)
Type of Document:
Technical Report (TR)
Number of pages and figures: 29, 21
Number of references: 4
Keywords:
Tide gauge observations, Quality Control, Time series, SLP64
Security classification:
Open
Distribution:
Open
Date of publication:
January, 2024

Abstract (150 words):
Observation is a method of data collection through the systematic monitoring of behavior, events, or physical characteristics in their natural setting. It plays a crucial role in environmental and oceanographic studies. INCOIS has established a network of 36 state-of-the-art tide gauge stations at strategic locations along the Indian coastline to monitor tsunami wave propagation and validate oceanographic model outputs. This report focuses on the data processing and delayed-mode Quality Control (QC) of hourly tide gauge data, particularly from Chennai, Cochin, Nagapattinam, Visakhapatnam, and Kakinada stations. A standardized QC procedure was applied across all stations to ensure consistency and reliability. The QC process involves the detection and correction of anomalous values, gap identification, and data flagging. This document details the QC methodology and emphasizes the significance of accurate, high-quality tide gauge data for operational ocean services, model validation, and coastal hazard monitoring.

TECHNICAL REPORT Report No.: ESSO-INCOIS-ODM-TR-01(2024)

INCOIS- Delayed Mode Tide Gauge Data Processing and Quality Control Procedure Over North Indian Ocean.

by

P.Suneeta, TVS Udaya Bhaskar, E. Pattabhi Rama Rao, K.Srinivas and V. Ravichandra

Indian National Centre for Ocean Information Services (INCOIS)

(Earth System Science Organization (ESSO), Ministry of Earth Sciences (MoES))HYDERABAD, INDIA

www.incois.gov.in

January, 2024

Contents

List of Figures

List of Tables

Abstract

Observation is a method of data collection through the systematic monitoring of behavior, events, or physical characteristics in their natural setting. It plays a crucial role in environmental and oceanographic studies. INCOIS has established a network of 36 state-of-the-art tide gauge stations at strategic locations along the Indian coastline to monitor tsunami wave propagation and validate oceanographic model outputs. This report focuses on the data processing and delayed-mode Quality Control (QC) of hourly tide gauge data, particularly from Chennai, Cochin, Nagapattinam, Visakhapatnam, and Kakinada stations. A standardized QC procedure was applied across all stations to ensure consistency and reliability. The QC process involves the detection and correction of anomalous values, gap identification, and data flagging. This document details the QC methodology and emphasizes the significance of accurate, high-quality tide gauge data for operational ocean services, model validation, and coastal hazard monitoring.

1. Introduction:

The global sea level record from tide gauges is an important indicator of global climate change evolution and impact. Tide gauge data also captures a wide range of local and regional phenomena such as decadal climate variability, tides, storm surges, tsunamis, swells, and other coastal processes. Data from tide gauges is used to validate ocean models and detect errors and drifts in satellite altimetry. In comparison to satellite data, tide gauge data has a longer record, a finer temporal resolution, but a coarser spatial resolution. Due to a number of factors, including local and regional changes in winds and ocean circulation that affect sea level, the impact of atmospheric pressure changes on sea level, the relative lack of long, continuous records, and the lack of a common datum across tide gauge sites, calculating global mean sea level from tide gauges is not simple. However, you will typically need to conduct overt observations because of ethical problems related to concealing your observation. Observations can also be either direct or indirect. Direct observation is when you watch interactions, processes, or behaviours as they occur; for example, observing a teacher teaching a lesson from a written curriculum to determine whether they are delivering it with fidelity. Indirect observations are when you watch the results of interactions, processes, or behaviours. Tide gauge is a device for measuring sea level changes and detecting tsunamis. Tide gauges are very critical to monitor the tsunami progress and the coastal sea level changes. INCOIS installed three types of tide gauge sensors (RAD, PRS, ENC) at 21 locations (established in 2010-11) and one sensor (RAD) at 15 locations (established in 2015-16) to measure the water level heights. Figure 1 shows the network of Tide gauge locations as maintained by the INCOIS and red circles with in the figure shows the present tide gauge locations what we have discussed in this technical report. Table 1 show the list of Tide gauges stations established by INCOIS.

Radar (RAD) gauge: The radar gauge emits microwave pulse and measures the distance between its reference point and the water surface with an averaging of the single high-rate measurements over a few seconds. The gauge is usually mounted on top of the tide gauge cabin with a U-profile steel girder. The maintenance of these sensors is very low when compared with other kind of sensors.

Pressure (PRS) gauge: This gauge measures the hydrostatic pressure of the water column and converts it in to height based on knowledge of the water density and local acceleration due to gravity. The gauge is mounted in the sea water within a metal construction.

Shaft encoder (ENC) gauge: The shaft encoder records changes in the water level with a float in a stilling well that rises and falls with the tide. A pulley attached to the float causes the encoder shaft to rotate in response to vertical motion of the float. The position of the rotating disk is determined by single or dual optical or magnetic sensors to provide an electrical output.

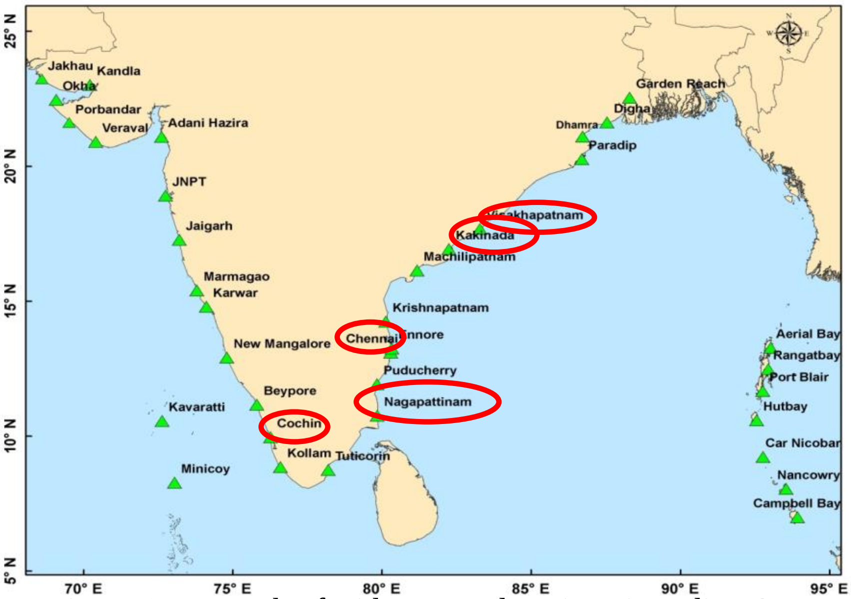


Figure 1: Network of Tide gauges locations in Indian Ocean

S.No.	Station Name	Latitude (N)	Longitude (E)	Establishment in
1	Chennai	13.1	80.3	Established in July 2010
2	Cochin	9.96	76.26	Established in Mar 2011
3	Nagapattinam	10.76	79.85	Established in Jul 2010
4	Kakinada	16.93	82.25	Established in Oct 2011
5	Visakhapatnam	17.68	83.28	Established in Jul 2010

Table 1: List of Tide gauges stations established by INCOIS

All sensors are connected to a data logger which provides data logging, storage and control functions for the sensors as well as the data transmission to tsunami warning centre. Each tide gauge measures the sea level by sampling for every one minute and transmits it for every 5 minutes (Islands stations every 3 minutes). The data are transmitted in real time through different modes of communication like INSAT, GPRS simultaneously for processing and interpretation. INCOIS has equipped with INSAT satellite communication hub (Harikumar et al., 2013, state-of-the-art computing hardware, real-time data processing and visualization facilities for real-time reception, display and archiving of tide gauge data.

2. Primary data processing and QC for Five Tide Gauge Stations:

Based on the delivery timelines, each associated with a different level of quality control and / or data processing, and related to different applications, data processing can be applied to: real time data (RT), near-real time data (NRT) and delayed mode data (DM) (Pouliquen et al. 2011). In this paper, the term delayed mode data (DM) is applied to latencies of 1 min, and is needed mainly in the context of harbour operation and storm surge and tsunami warning systems.

Otherwise, any Quality Control (QC) system produced with any level of improvement will always fail to identify data that appears to be correct but is actually incorrect. QC techniques regularly vary depending on the type of data, i.e., the parameter, the duration of the data observed, the frequency of the observations, the type of platform used to measure the observations, and so on. The present dataset is a time series dataset with variable frequency; 1 min observations translated it in to 1 hour observations. The present study focuses on five tide gauge stations located along the northern Indian Ocean, representing an initial case study on tide gauge data processing and quality control using standardized codes and validation against predicted tidal data. For this analysis, 14 years of data (2010–2023) from Chennai, Cochin, Nagapattinam, Kakinada, and Visakhapatnam stations were utilized.

2.1 Delayed Mode Data Processing:

The entire QC technique is separated into two parts: a primary QC procedure and a secondary QC procedure with more complexity. Table 2 shows the lists the various QC procedures under both the categories. We assign the QC flags, QC flag which provides the details of the result of each QC test that the datum has undergone. The primary QC flag takes only three QC flags, where ‘1′ denotes that all QC tests are passed by the datum, and it is good data. ‘9′ denotes that the datum has failed in missing values. Table 3 provides the details of the same. Table 3 lists the QC tests that are included in the primary QC flag. The starred ones are those which give a QC flag ‘2′ when passed outliers, The remaining primary QC tests result in QC flags ‘3′ and ‘4′ when QC test is passed spike and out of range, respectively. Finally, the QC tests of time of observation, location, and time sequence will be directly assigned to the primary QC flag, I.e., if any of the four tests pass, then the primary QC flag is assigned ‘1′.

Primary QC Tests
Range check
Spike test
Outliers check
Out of Range
Missing values check Buddy or Neighbor test
Time sequence check

Table 2. Primary QC procedures.

Binary number	Detail
1	Good data
9	Missing values
2	Outliers
3	Spikes
4	Out of control

Table 3. Binary numbers used in defining the secondary QC flag.

Missing Values: Form the observed data it checks for missing values in the variable. After executing these missing values, the variable quality will contain a flag of 1 for the positions where the variable has missing values and 9 for the positions where the data is present. This approach allows easily identifying and handling missing data in variables.

Range Test: Calculates the variable's range the function range calculates the difference between the array's maximum and minimum values. Overall, 5 station range tests were carried out from 2010 to 2023. Table 4 displays the range test results for five tide gauge locations.

S. No.	Station Name	Latitude (°N)	Longitude (°E)	State/UT	Minimum (cm)	Maximum (cm)
1	Chennai	13.1	80.3	Tamil Nadu	0.01	661.1
2	Cochin	9.96	76.26	Kerala	0	349.9
3	Nagapattinam	10.76	79.85	Tamil Nadu	0	197.0
4	Kakinada	16.93	82.25	Andhra Pradesh	0	294.5
5	Visakhapatnam	17.68	83.28	Andhra Pradesh	0	349.0

Table 4. Five stations range test from the period 2010-2023

Spike test: Identifies spike in the variable by comparing the absolute difference between consecutive elements to threshold of 3 times the standard deviation. The quality flags will be updated to 3 for the elements identified as spikes.

Out of control: With these test any data points in variable that fall below the lower limit or above the upper limit will be flagged with a quality value of 4, while the quality flags for the remaining data points (missing values, outliers, etc.) will remain unchanged.

Outliers: Outliers calculate the z-score for each element in the variable extending the functionally to identify outliers in addition to handling missing values. The quality flags will now have the value 2 for the outliers. 0 for the rest of the data position in variables.

$$Z = (x - \mu)/ \sigma$$

Where μ is the mean of the data and σ is the standard deviation of the data. The data with Z-values beyond 3 are considered as outliers.

Buddy or Neighbor test: where the data from one station would be compared with either a second sensor at the same location or possibly with data from a nearby station to see if events propagate along coastline.

Time Sequence Check: A time sequence test for tide gauge data is a statistical or analytical approach used to assess the temporal patterns or trends in sea level data recorded by tide gauges. Tide gauges are instruments that measure and record the changes in sea level over time at a specific location. Analysing time sequences helps understand the long-term variations, trends, and patterns in sea level.

3. Result and Discussions:

3.1 Data processing and quality control by using observed data for Chennai station:

In the initial phase of data processing and quality control, the tide gauge data were systematically evaluated through multiple stages, as illustrated in Figure 2. The blue line represents the observed water level data spanning from 2010 to 2023. A noticeable shift in the data was detected in 2013, with additional anomalies observed in 2021. These irregularities were corrected using appropriate filtering techniques, and the corrected results are presented in the Level-2 plot of Figure 2. Figure 3 shows the outcome of quality checks applied to the observed data, where the blue line represents the raw observations, red dots indicate 2,568 out-of-range values, and the green dotted line marks the mean water level. The recorded water levels ranged from 0 to 661.0 cm. A numerical quality flag system was employed: 1 denotes good data, 2 indicates outliers, 3 represents spikes, 4 marks data out of control, and 9 indicates missing values. These systematic checks help ensure data reliability for subsequent analyses.

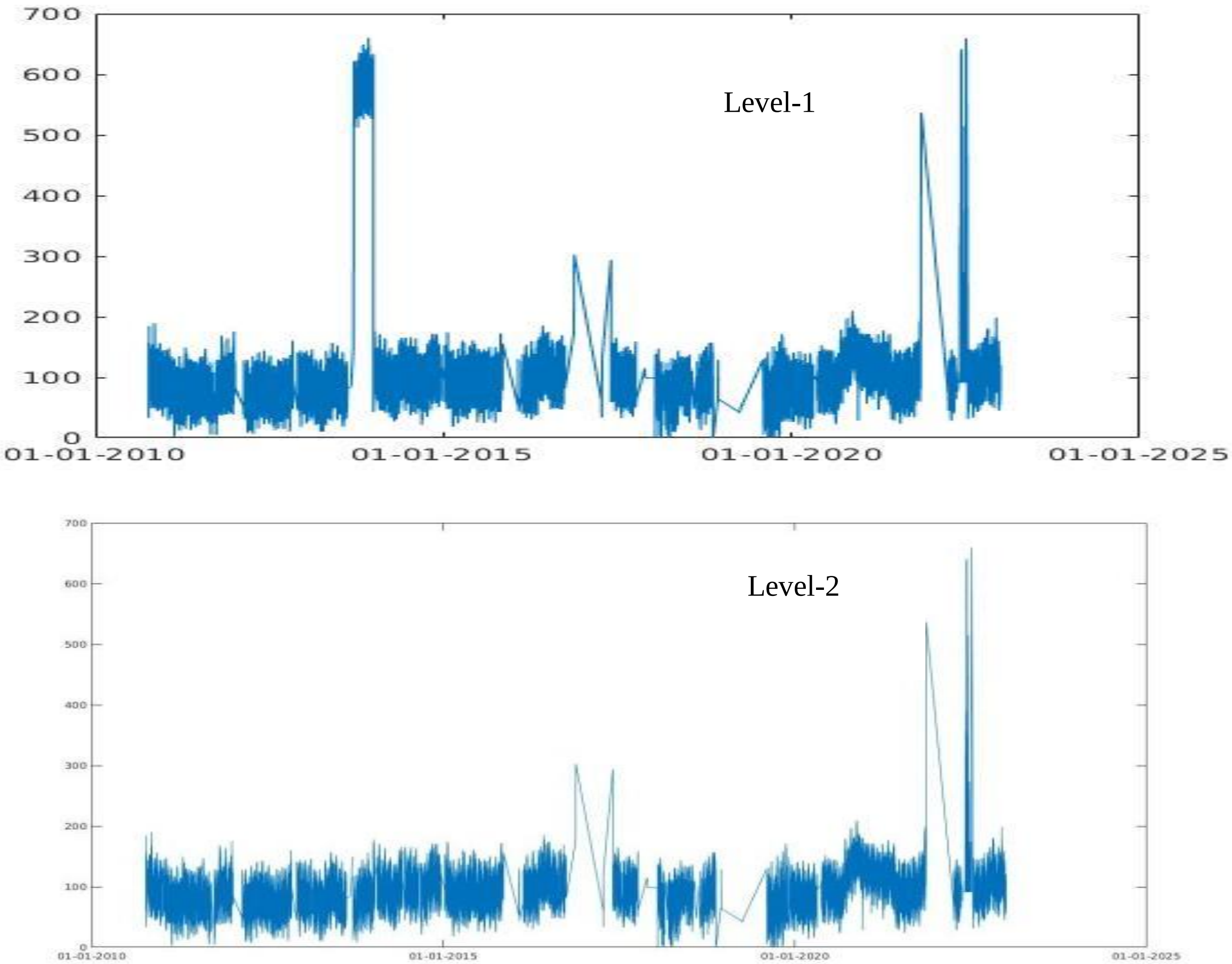


Figure 2: Primary data processing at different levels for Chennai station

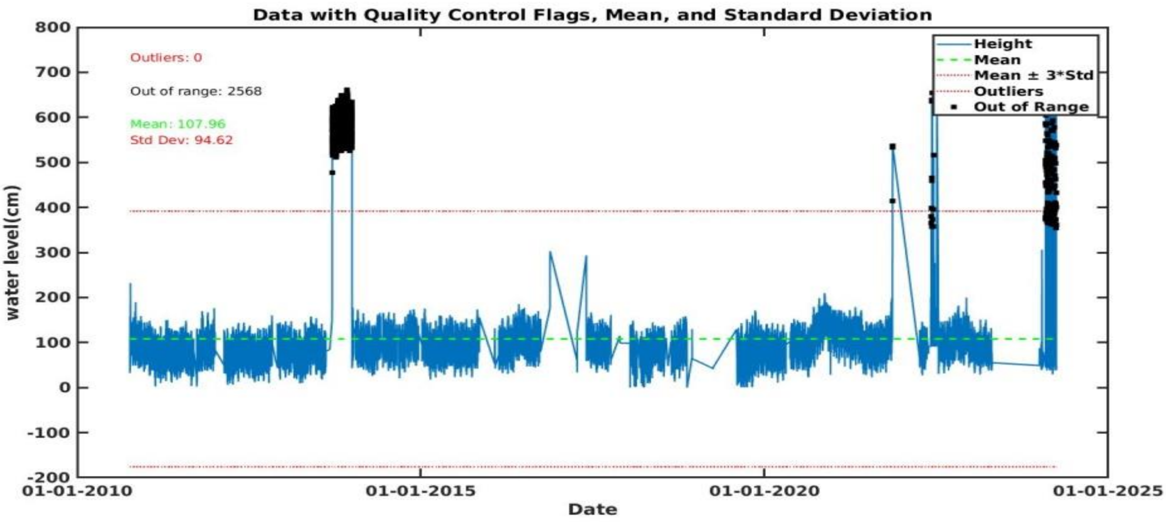
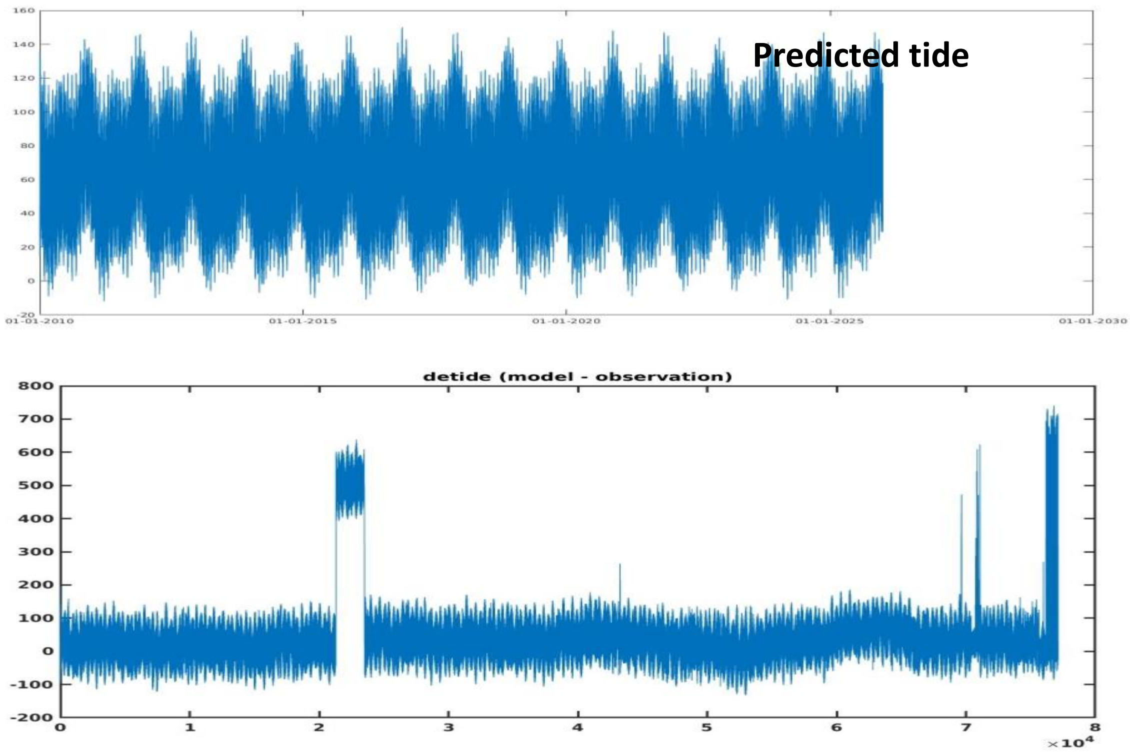


Figure 3: Quality control tests for Chennai station from the period 2010-2023.

3.2 Data processing and quality control by using observed and predicted tide data:

Hourly data processing was conducted using both observed and predicted tidal data for the period 2010–2023. Figure 4 presents the merged predicted tidal data for the Chennai station, illustrating the results of the tidal analysis. The detided values were calculated by subtracting the predicted tide from the observed data (Observation – Predicted). The Root Mean Square Error (RMSE) of 61 cm and a standard deviation of 55 cm indicate a good agreement between the observed and model-predicted data. In Figure 4, the black line represents the observed values, while the blue line denotes the predicted tidal heights.



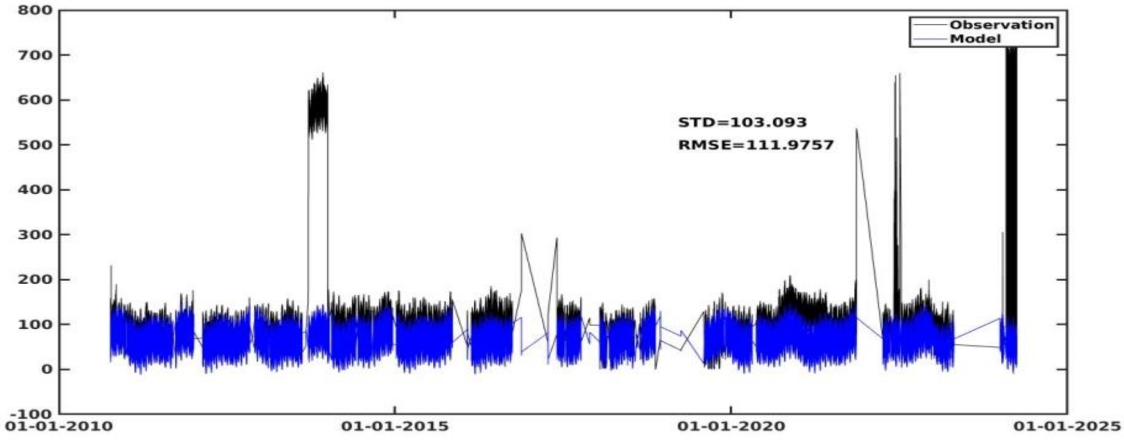


Figure 4: Predicted and observed data processing for Chennai station from the period 2010-2023.

3.3 Tide Gauge location in Cochin station:

3.3.1 By using the observed data:

Tide gauge data from the Cochin station, covering the period from 2011 to 2023, were processed and evaluated for quality. Figure 5 illustrates the various stages of data processing. Based on hourly resolution, a total of 113,952 data points were expected; however, only 85,799 data points were available, indicating 28,153 missing values. Figure 6 presents the quality control results, where 20 out-of-range values and 362 outliers were identified. In the figure, the blue line represents the time series of observed water levels, the green dotted line indicates the mean water level, red dots denote outliers, and black dots mark out-of-range values. These quality assessments provide a clear understanding of the data integrity at the Cochin station.

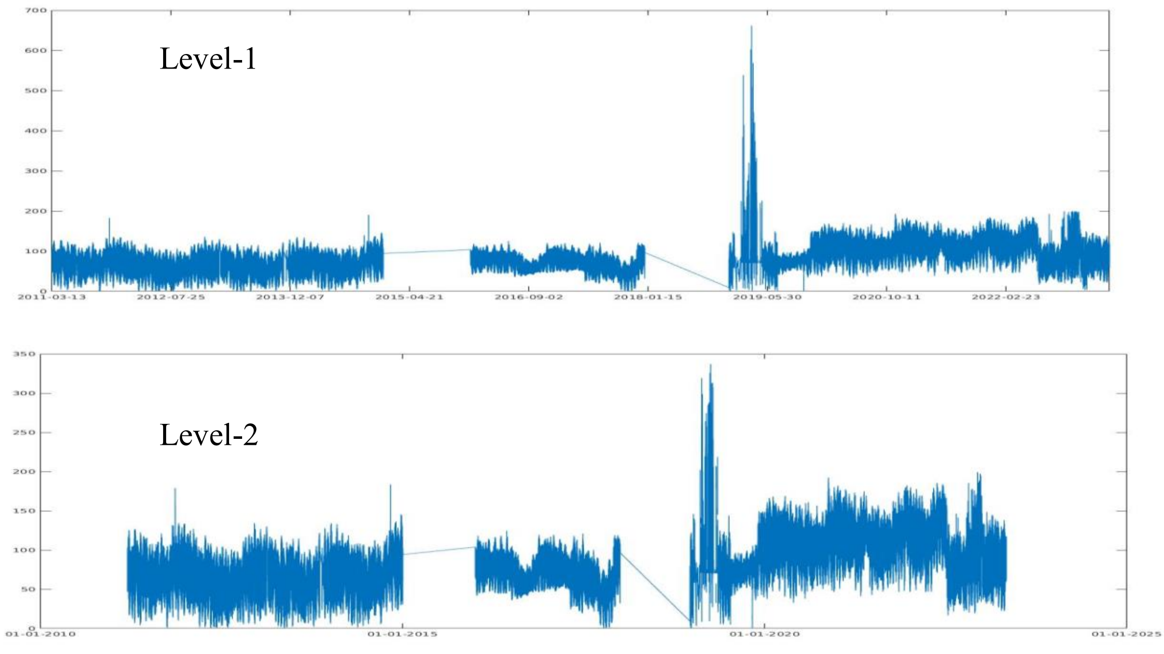


Figure 5: Primary data processing at different levels for Cochin station from the period 2011-2023

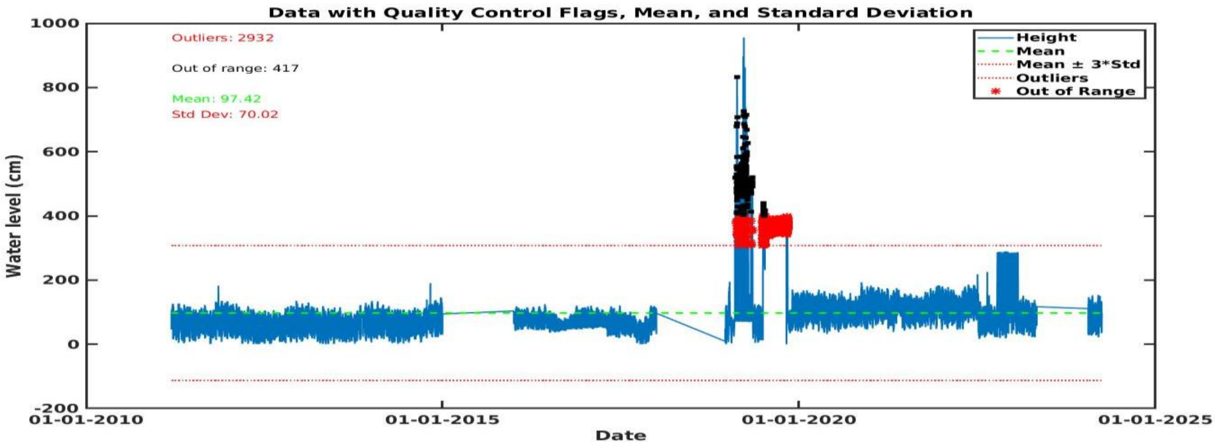


Figure 6: Quality control tests for Cochin station from the period 2011-2023

3.3.2 By using observed and predicted tide data:

Hourly observed and predicted tidal data from 2011 to 2023 were processed for the Cochin station. Figure 7 shows the merged predicted tidal data and detided values (Observation – Predicted). An RMSE of 43 cm and standard deviation of 48 cm indicate good agreement. The black line represents observed values, and the blue line shows predicted tides.

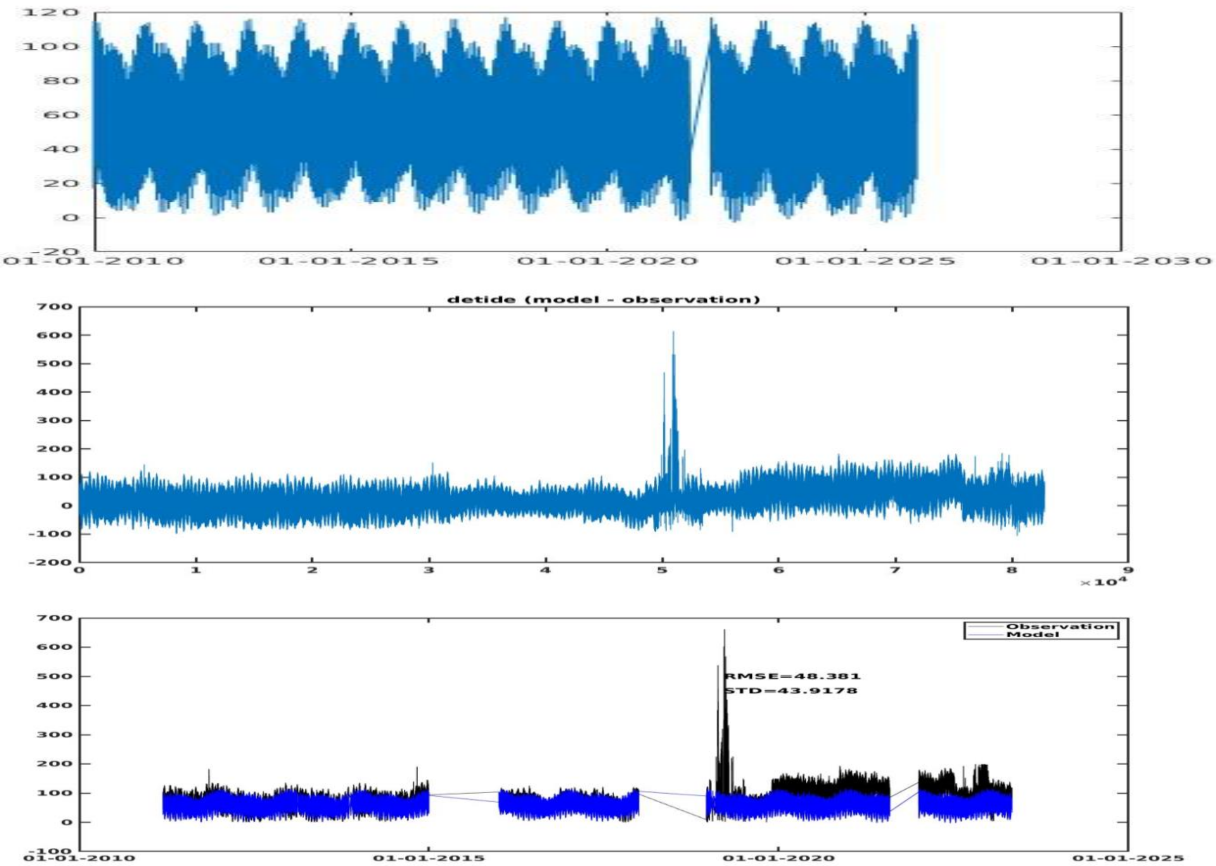


Figure 7: data processing by using predicted data Cochin station from the period 2011-2023

3.4 Tide Gauge location in Nagapattnam station:

3.4.1 By using the observed data:

Tide gauge data from the Nagapattinam station (2011–2023) were processed for quality assessment. Figure 8 shows the stages of data processing. Of the expected 113,952 hourly data points, only 99,317 were available, with 14,635 missing. As shown in Figure 9, the dataset contains 17 out-of-range values and 434 outliers. The blue line represents the water level time series, the green dotted line indicates the mean, red dots mark outliers, and black dots denote out-of-range values.

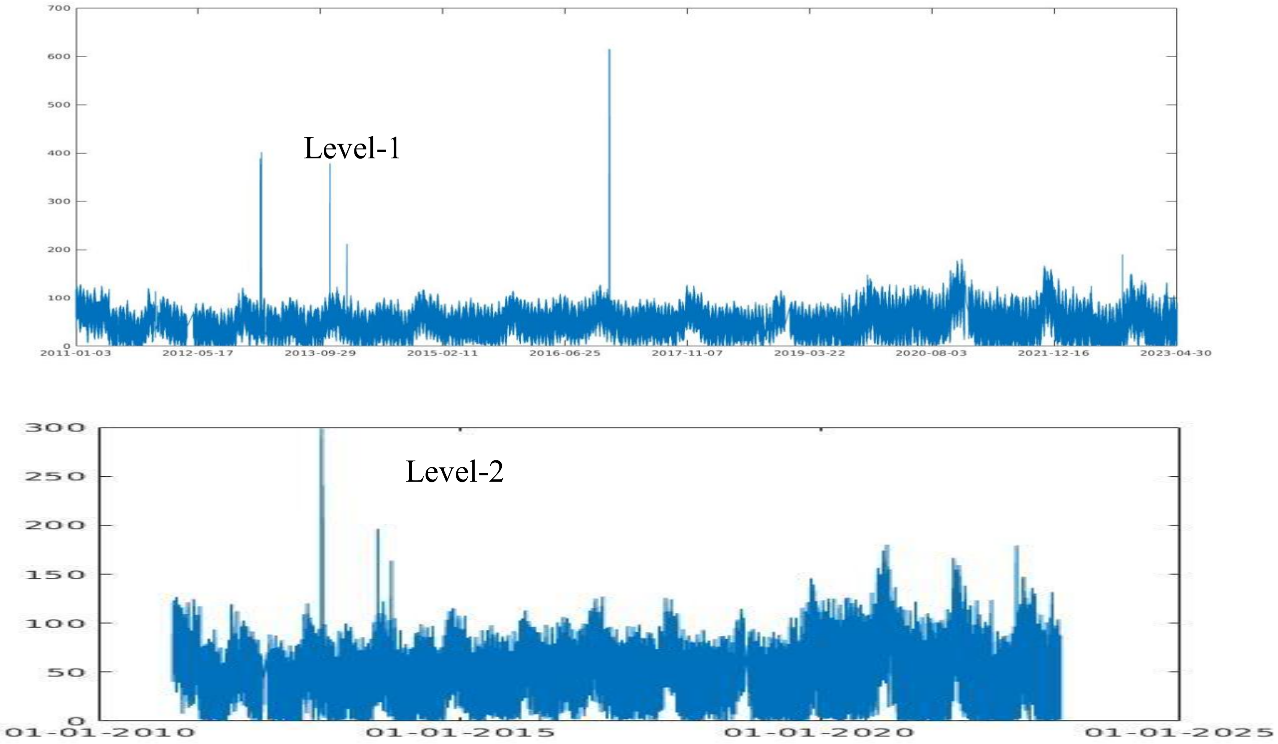


Figure 8: Primary data processing at different levels for Nagapattnam station from the period 2011-2023

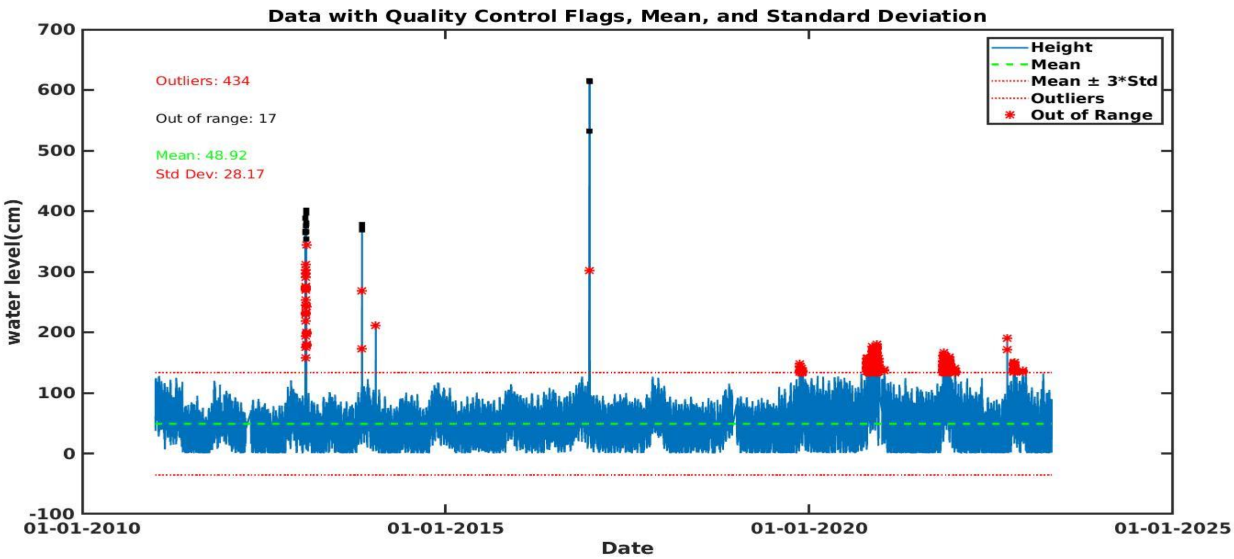


Figure 9: Quality control tests for Nagapattnam station from the period 2011-2023

3.4.2 By Using Observed and Predicted tide data:

Hourly observed and predicted tidal data from 2011 to 2023 were processed for the Nagapattinam station. Figure 10 presents the merged predicted tidal data and detided values, calculated as (Observation – Predicted). The analysis yielded an RMSE of 46 cm and a standard deviation of 43 cm, indicating excellent agreement with the model data. In the figure, the blue line shows predicted values, while the black line represents observed values.

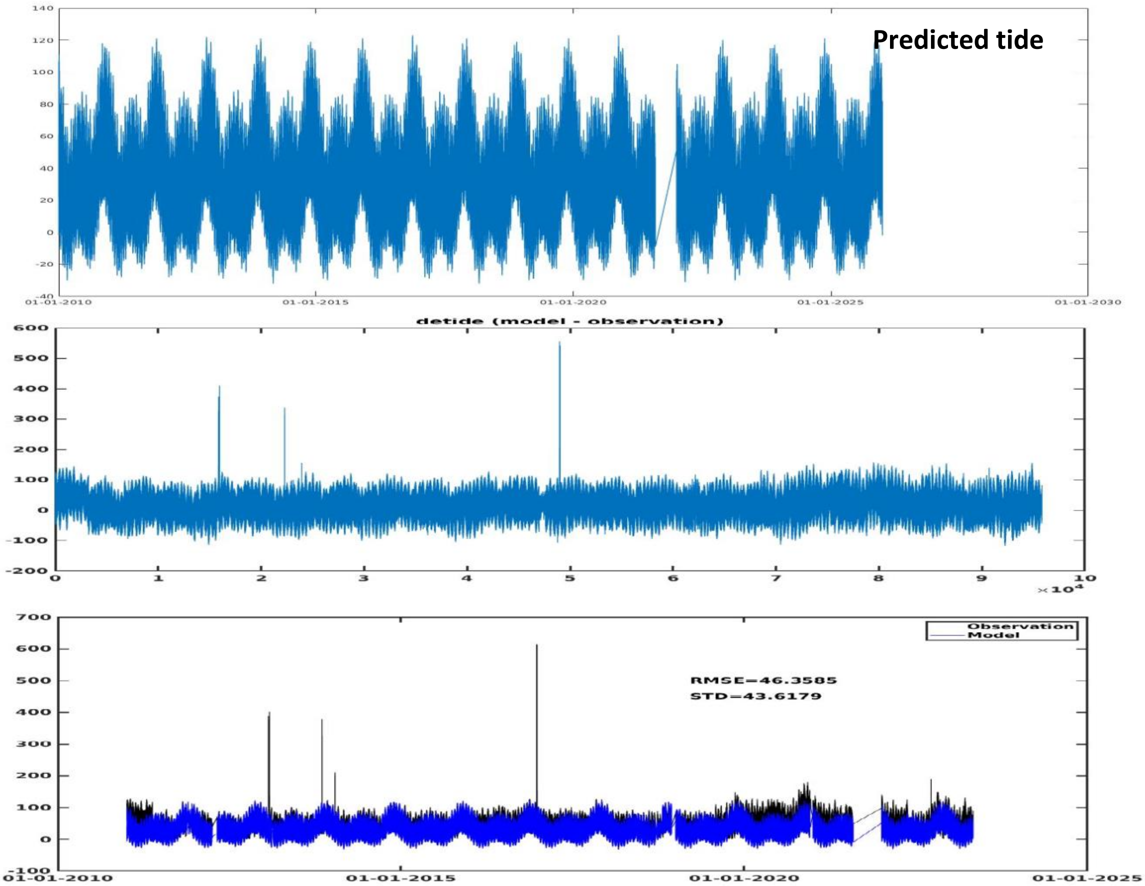


Figure 10: Data processing by using predicted data Nagapattnam station from the period 2011-2023

3.5 Tide Gauge location in Visakhapatnam station:

3.5.1 By using the observed data:

Tide gauge data from the Visakhapatnam station (2011–2023) were processed for quality assessment. Figure 11 illustrates the various stages of data processing. Of the expected 113,952 hourly data points, only 97,247 were available, resulting in 16,705 missing values. As shown in Figure 12, the dataset includes 13 out-of-range values and 127 outliers. In the figure, the blue line represents the water level time series, the green dotted line indicates the mean, red dots mark outliers, and black dots denote out-of-range values.

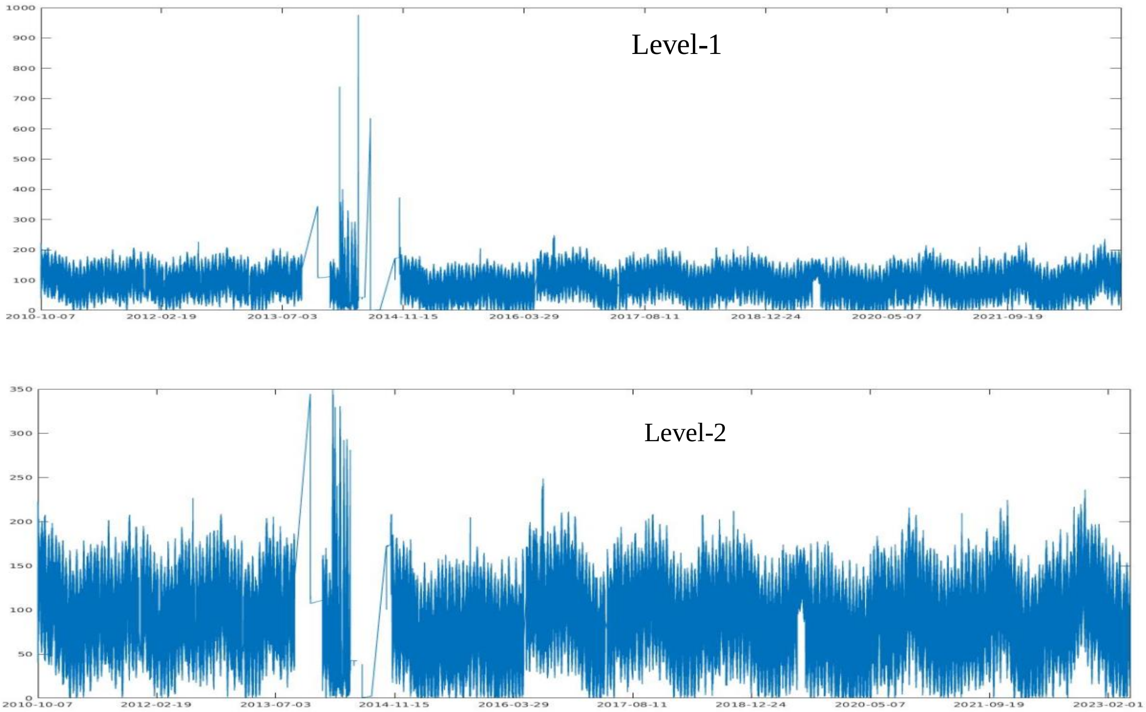


Figure 11: Primary data processing at different levels for Visakhapatnam station from the period 2011-2023

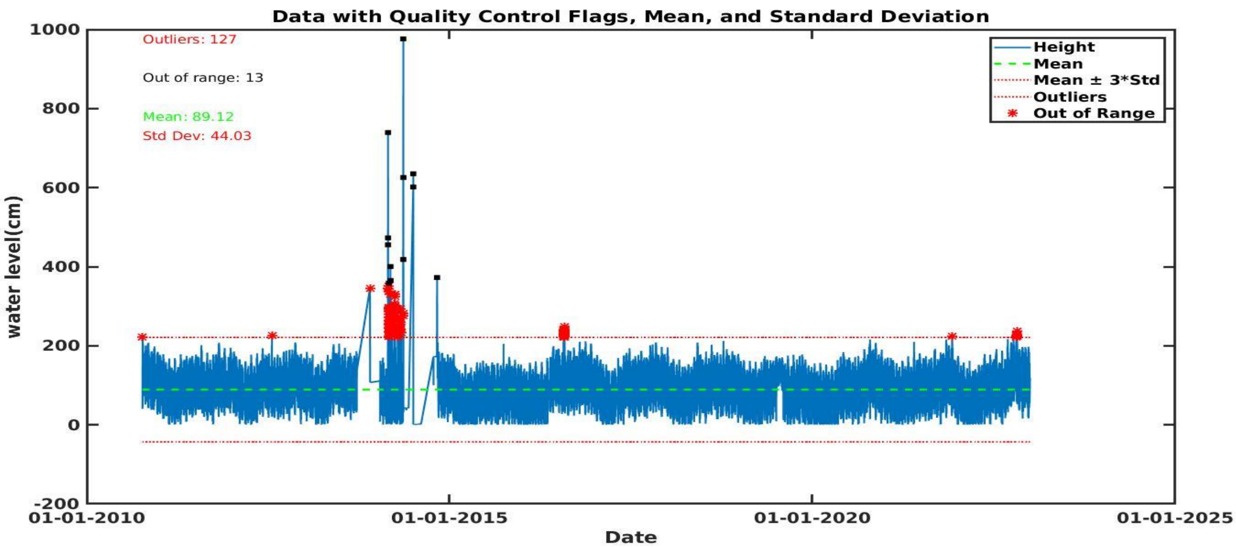


Figure 12: Quality control tests for Visakhapatnam station from the period 2011-2023

3.5.2 By using Observed and Predicted tide data:

Hourly observed and predicted tidal data from 2011 to 2023 were processed for the Visakhapatnam station. Figure 13 shows the merged predicted tidal data and detided values, calculated as (Observation – Predicted). The RMSE of 87 cm and standard deviation of 79 cm indicate good agreement with the model data. In the figure, the blue line represents predicted values, while the black line shows observed values.

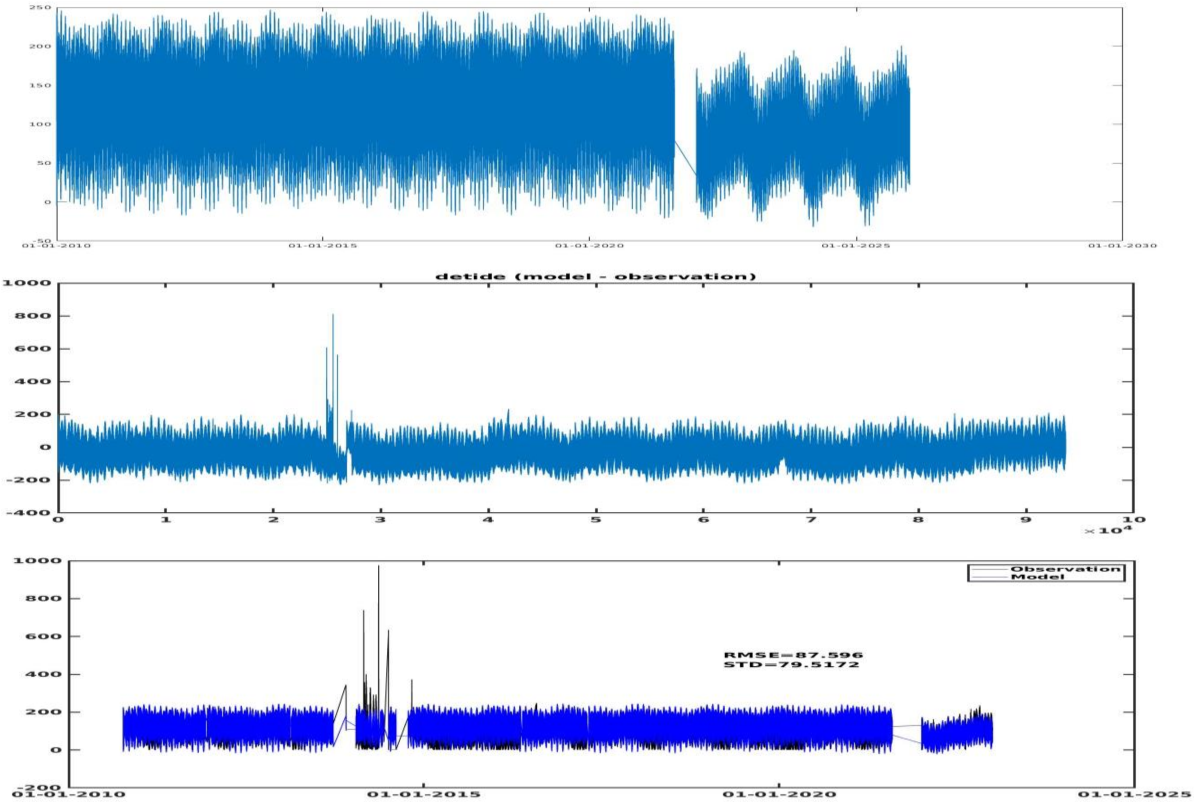


Figure 13: Data processing by using predicted data Visakhapatnam station from the period 2011-2023

3.6 Tide Gauge location in Kakinada station:

3.6.1 By using the observed data:

Tide gauge data from the Kakinada station (2011–2023) were processed for quality assessment. Figure 14 illustrates key stages of data processing. Of the expected 113,952 hourly data points, only 87,273 were available, resulting in 26,673 missing values. As shown in Figure 15, the dataset includes 10 out-of-range values and 32 outliers. In the figure, the blue line represents the water level time series, the green dotted line indicates the mean, red dots mark outliers, and black dots denote out-of-range values.

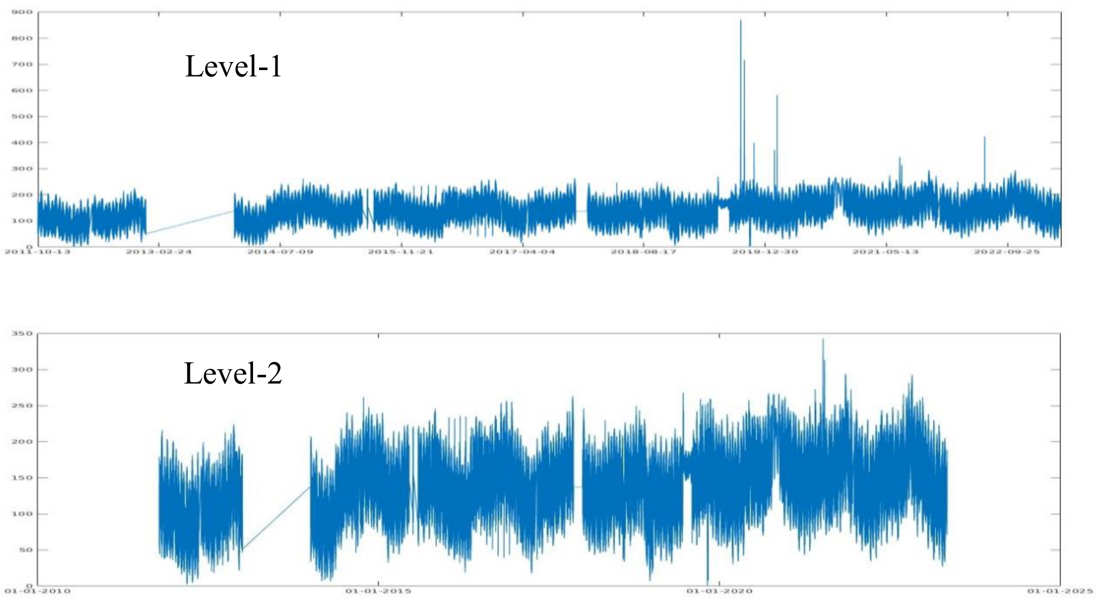


Figure 14: Primary data processing at different levels for Kakinada station from the period 2011-2023

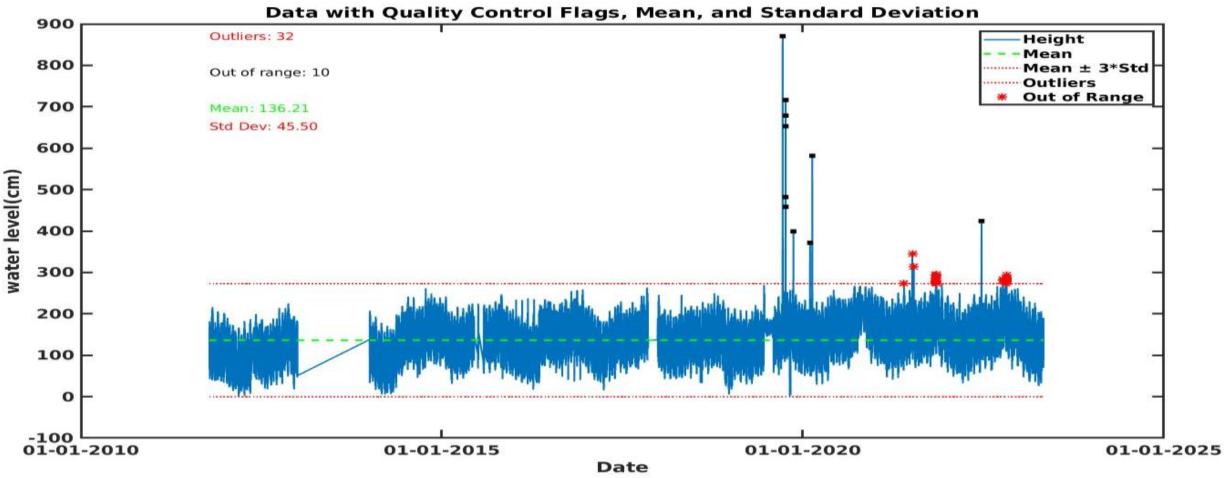


Figure 15: Quality control tests for Kakinada station from the period 2011-2023

3.6.2 By using observed and predicted tide data:

Hourly observed and predicted tidal data from 2011 to 2023 were processed for the Kakinada station. Figure 16 presents the merged predicted tidal data, while detided values were calculated as (Observation – Predicted). The RMSE of 92 cm and standard deviation of 77 cm indicate good agreement with the model data. In the figure, the blue line represents predicted values, and the black line shows observed values.

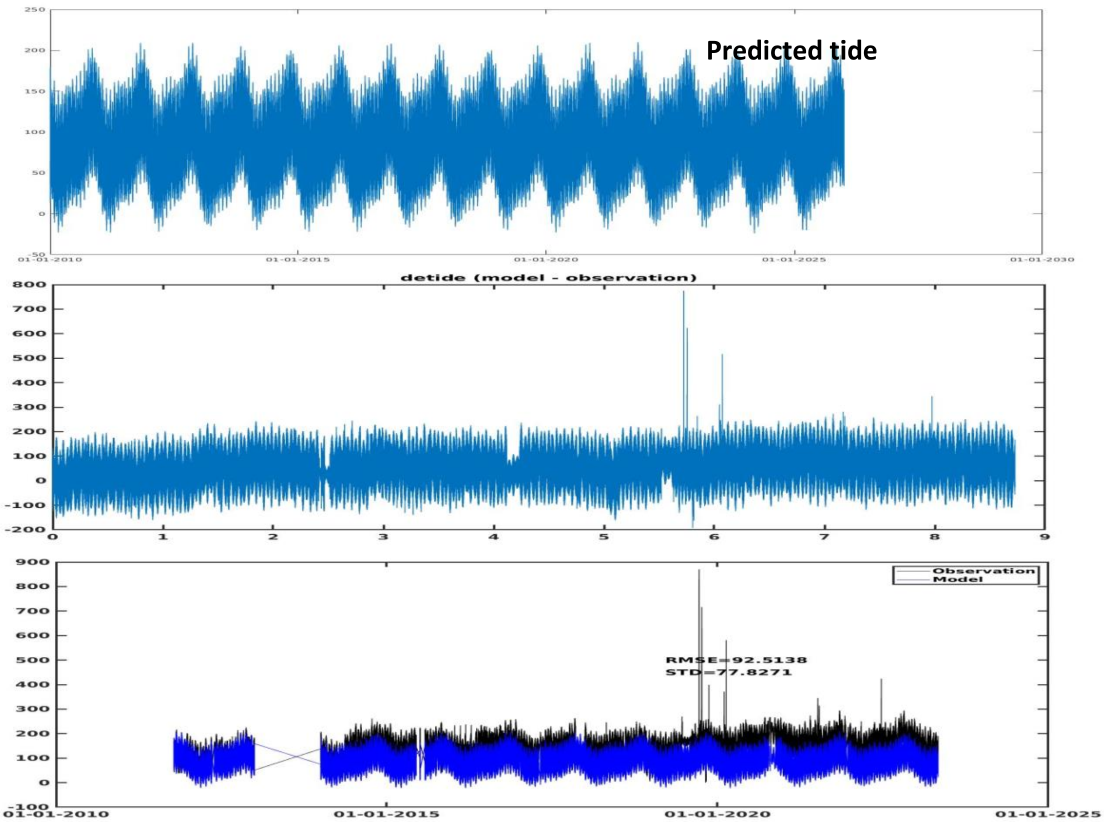


Figure 16: Data processing by using predicted data Kakinada station from the period 2011-2023

4. Data processing and quality control based on SLP64 software:

SLP64 is Hourly Sea Level Data Processing and Quality Control Software the package includes three principle tasks: 1) tidal analysis and prediction, 2) quality control, and 3) filtering hourly into daily and monthly values. The software is geared to those with prior knowledge of computer basics as well as a general understanding of sea level processing. The goal is to share software on a commonly available platform (the personal computer) that will enhance the quality of sea level data sets. This in turn will promote the growth of the international archives of sea level, thus, providing scientists with greater opportunities to understand the ocean. Python, an open source programming language, is required to run some SLP64 routines. The user must ensure Python is properly set up within the MS operating system. Matplotlib can be acquired and set up concurrently with Python, since it is part of the overall Python package.

4.1 Tidal Analysis and Prediction

The goal of this package is to quality control the hourly data files. The techniques have been well documented (Caldwell and Kilonsky, 1992; IOC, 1992). The basis for quality control is the inspection of plots of residuals, defined as observed data minus predicted tides. In the following sections, the tidal analysis and prediction routines, the plotting programs, and the various quality control procedures are outlined. Lastly, a filter routine for the creation of daily and monthly means from the hourly data is described.

4.1.1 Tidal Analysis

This program analyzes the hourly tide gauge data for a given period of time. Amplitudes and Greenwich phase lags are calculated via a least squares fit method coupled with nodal modulation for only those constituents that can be resolved over the length of the input record. The output is a file of harmonic constants which become input for the tidal height prediction program.

4.1.2 Tidal Prediction

Tidal prediction utilizes the harmonic constituents calculated by the tidal analysis to compute predictions in the form of either equally-spaced (in time) hourly values in the SLP64 processing format or high-low heights in the form of a tide table from the results got the 60 tidal constituents. Major tidal constituents contributing to the astronomical tide.

M2 - Principal lunar semidiurnal constituent

S2 – Principal solar semidiurnal constituent

N2 - Larger Lunar elliptic semidiurnal constituent

K1 - Luni-solar declinational diurnal constituent

O1 - Lunar declinational diurnal constituent

4.1.3 Quality control

Quality control ensures the scientific validity of the data. Three main aspects are emphasized: 1) the linking of the data to a reference level (tidal datum), 2) the inspection of thetiming quality, and 3) the replacement of short gaps and spikes. Technical aspects of quality control procedures have been well documented (Caldwell and Kilonsky, 1992; IOC, 1992). Different types quality control steps those were listed below.

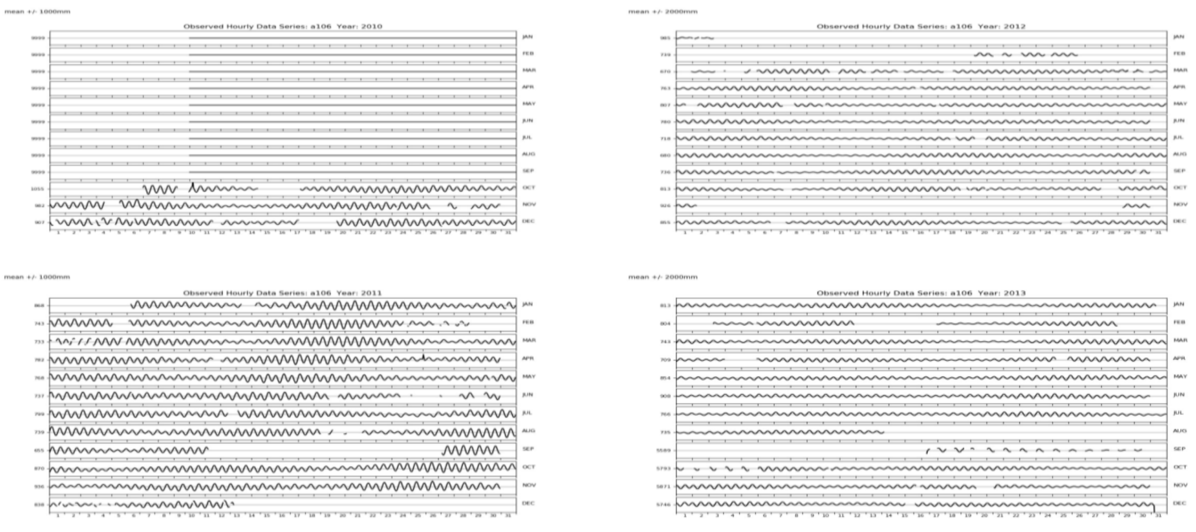
1. Hourly Residuals 2. Reference Level Stability,
3. Timing Errors 4. Short Gaps and Data Spikes

5. SLP64 Results:

Real-time data files from 2010 to 2023 for the five tide gauge stations, which can handle hourly data. Figure 10 depicts the observed hourly data sets and residuals for the water level in millimeters. The slp64 utilities are used to calculate residuals (Observed -Predicted). The residuals are based on the same year as plotted. In this case, the residuals are smoother. Since the goal of the residuals is to identify potential errors in the data, then for some sites with tidal species similar to this example, it would be best to refresh the harmonic constants file using time spans of tidal analysis close to the year in review. In this case, it would be best to recreate a harmonic constants file based on a more suitable time span as input to the tidal analysis. It is good practice to recreate a harmonic constants file after quality control is complete so that one has the best possible set of constituents to base future inspections.

5.1 Chennai station:

The observed hourly data, the predicted tides, and the residuals (defined as observed data minus predicted tides) have a similar SLP64 format and file naming convention. Each file consists of a year of values at an hourly sampling interval. Figure 17 shows the observed and residuals outputs from SLP64 software.



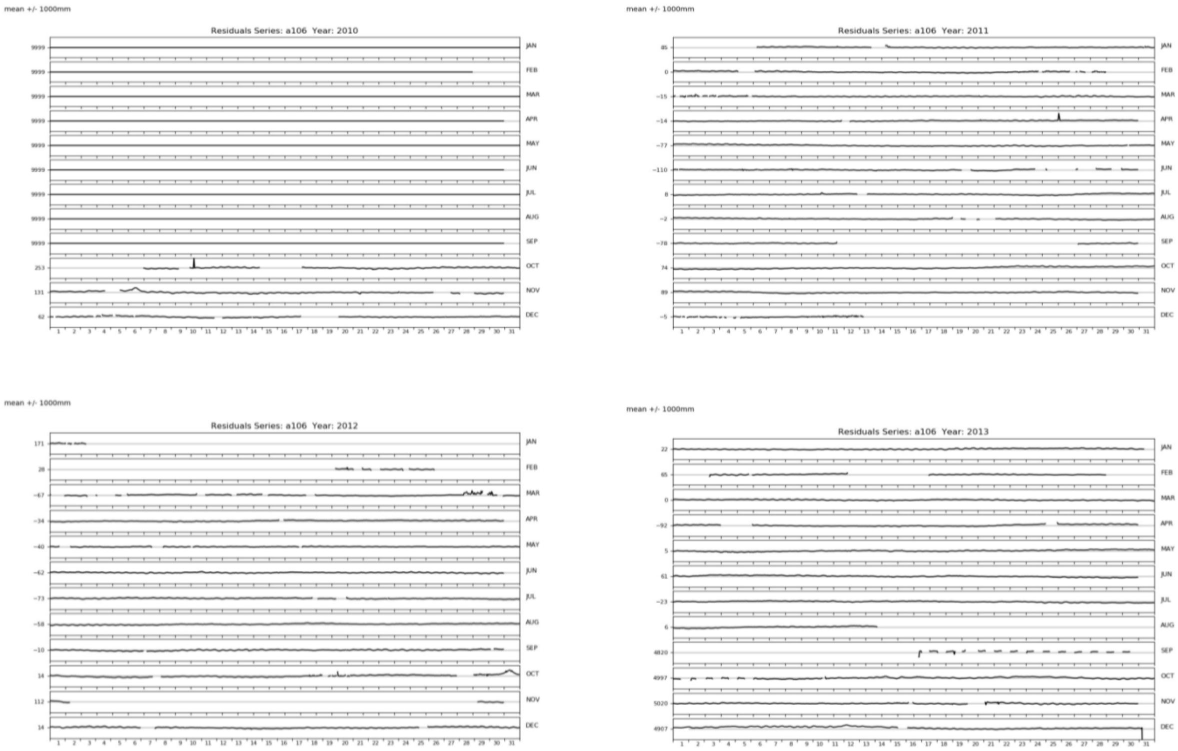
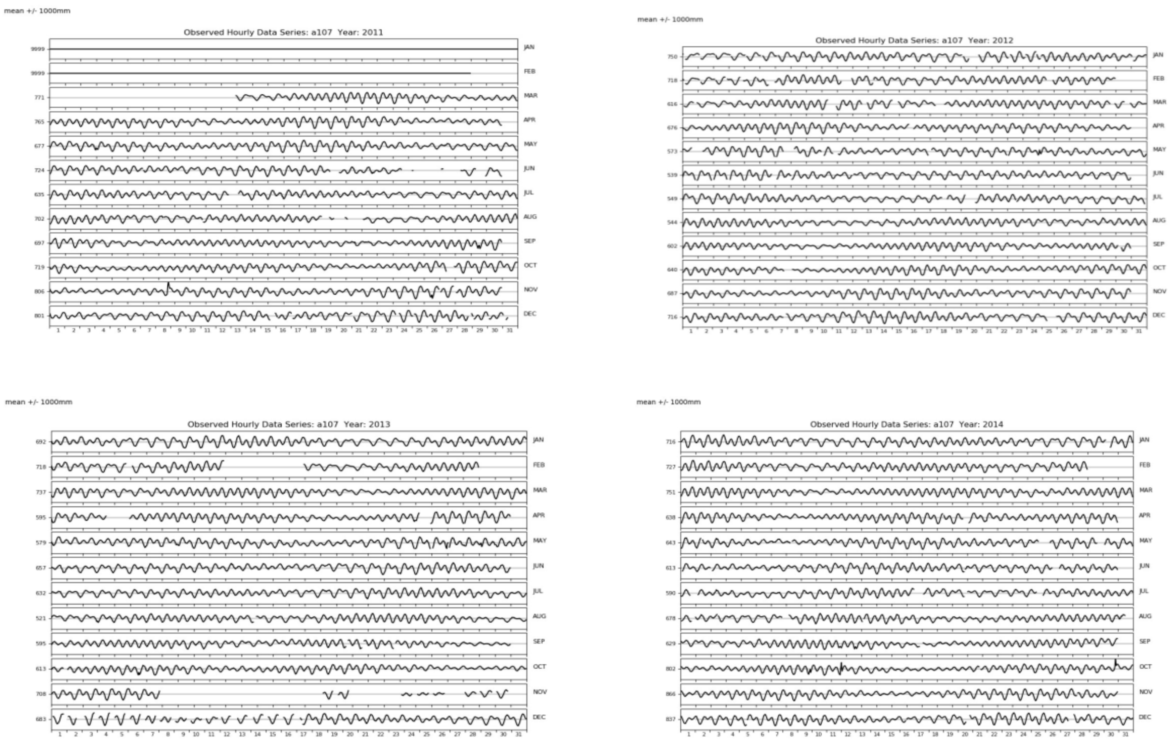


Figure 17: SLP64 results between observed data and residuals for the Chennai station from the period (2010-2013)

5.2 Cochin station:

The observed hourly data, the predicted tides, and the residuals (defined as observed data minus predicted tides) have a similar SLP64 format and file naming convention. Each file consists of a year of values at an hourly sampling interval. Figure 18 shows the observed and residuals outputs from SLP64 software.



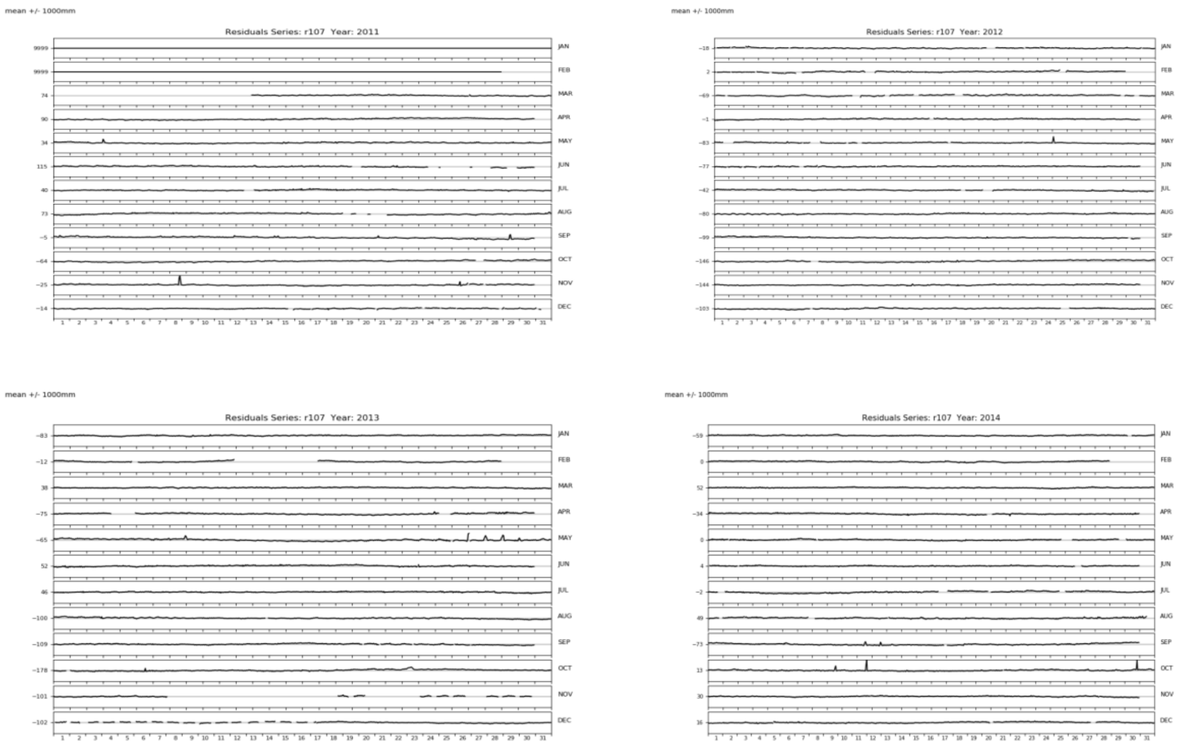
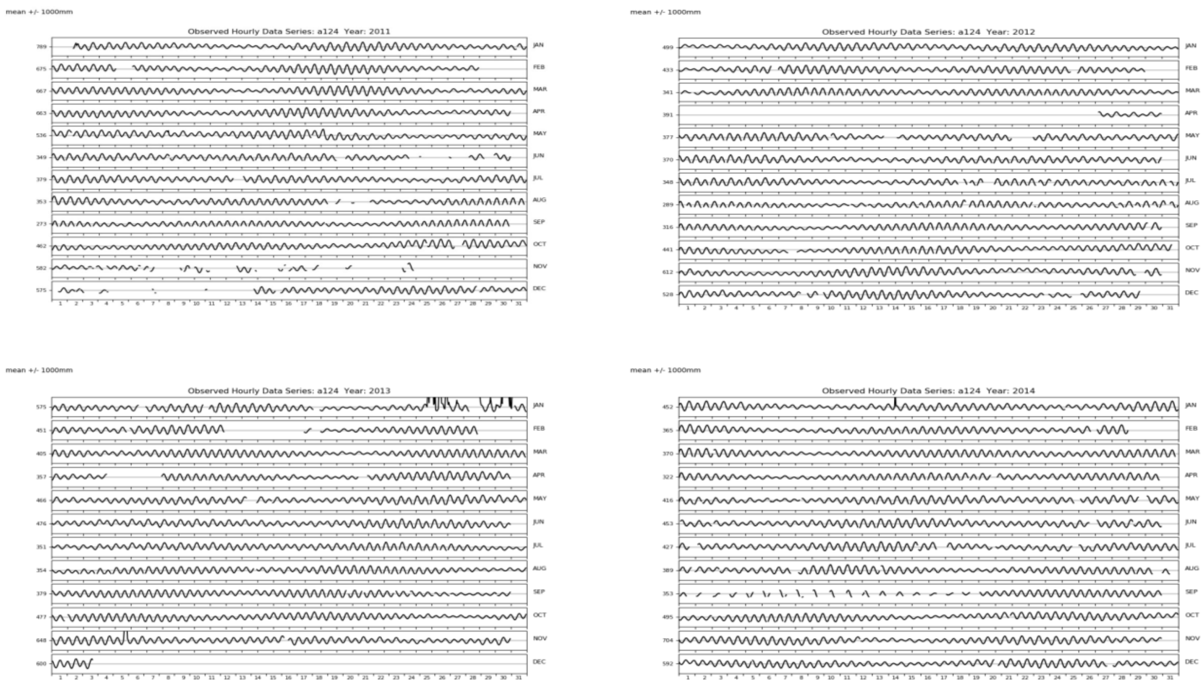


Figure 18: SLP64 results between observed data and residuals for the Cochin station from the period (2010-2013).

5.3 Nagapattnam station:

The observed hourly data, the predicted tides, and the residuals (defined as observed data minus predicted tides) have a similar SLP64 format and file naming convention. Each file consists of a year of values at an hourly sampling interval. Figure 19 shows the observed and residuals outputs from SLP64 software.



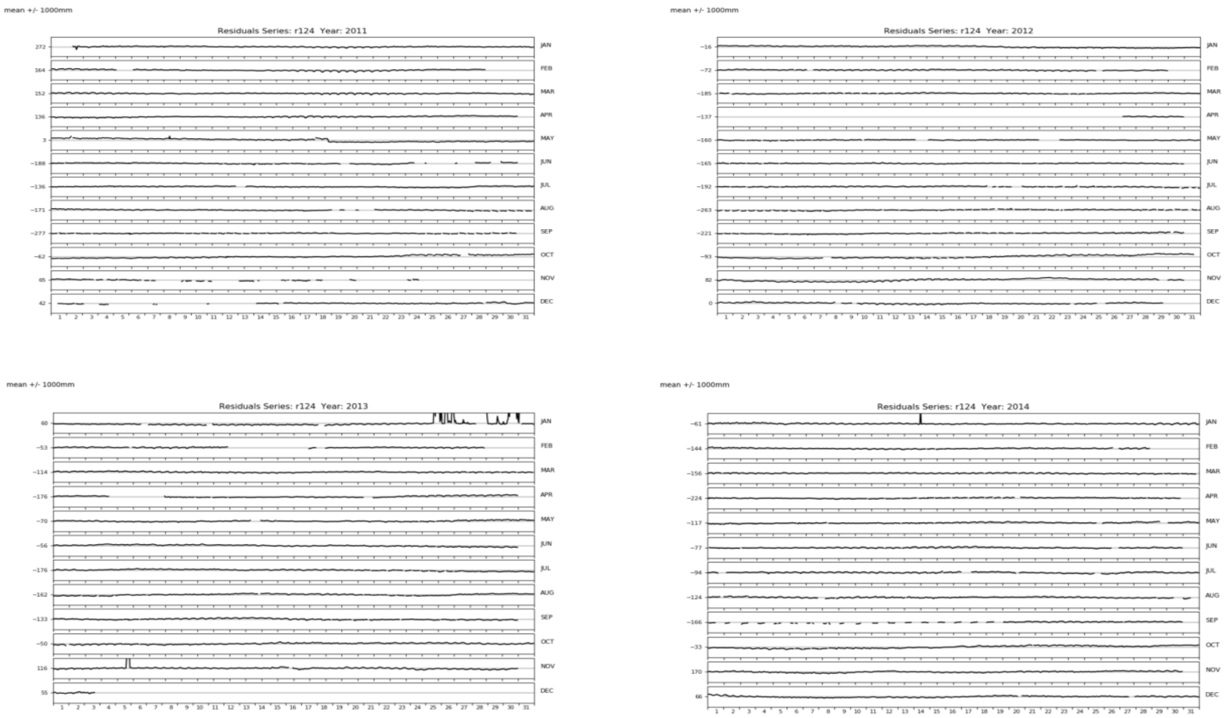
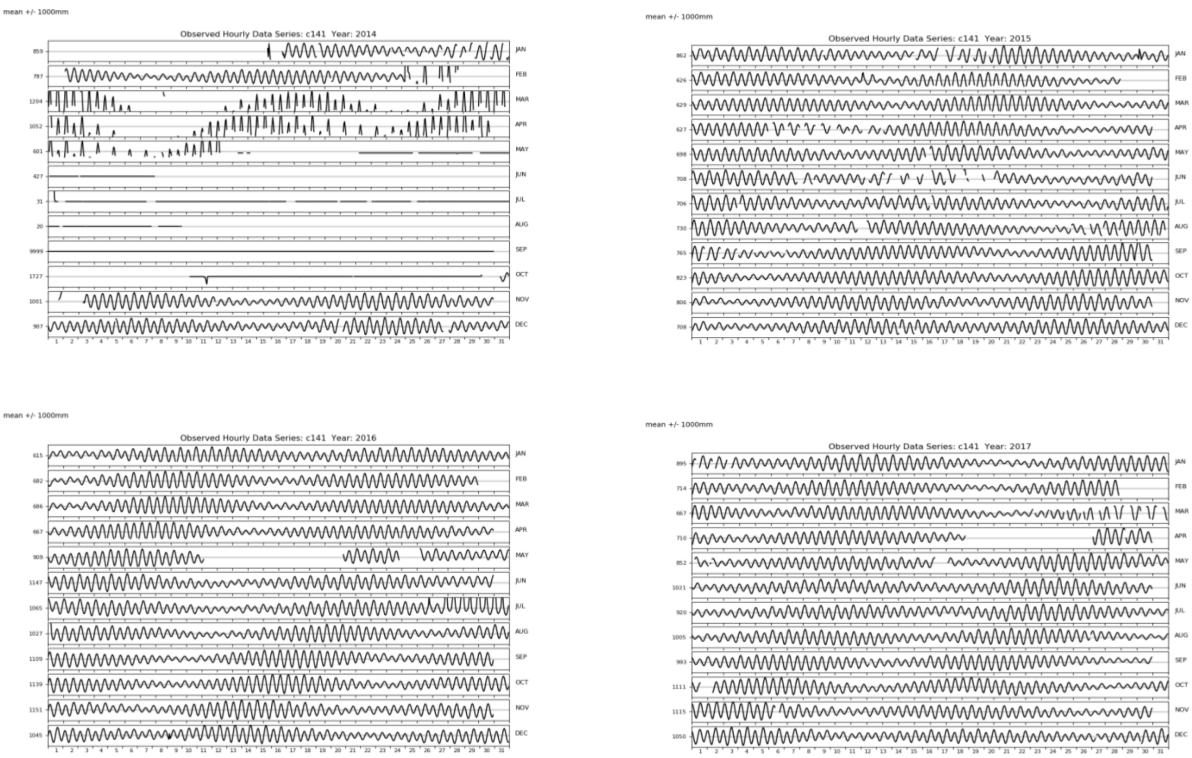


Figure 19: SLP64 results between observed data and residuals for the Nagapattnam station from the period (2010-2013)

5.4 Visakhapatnam station:

The observed hourly data, the predicted tides, and the residuals (defined as observed data minus predicted tides) have a similar SLP64 format and file naming convention. Each file consists of a year of values at an hourly sampling interval. Figure 20 shows the observed and residuals outputs from SLP64 software.



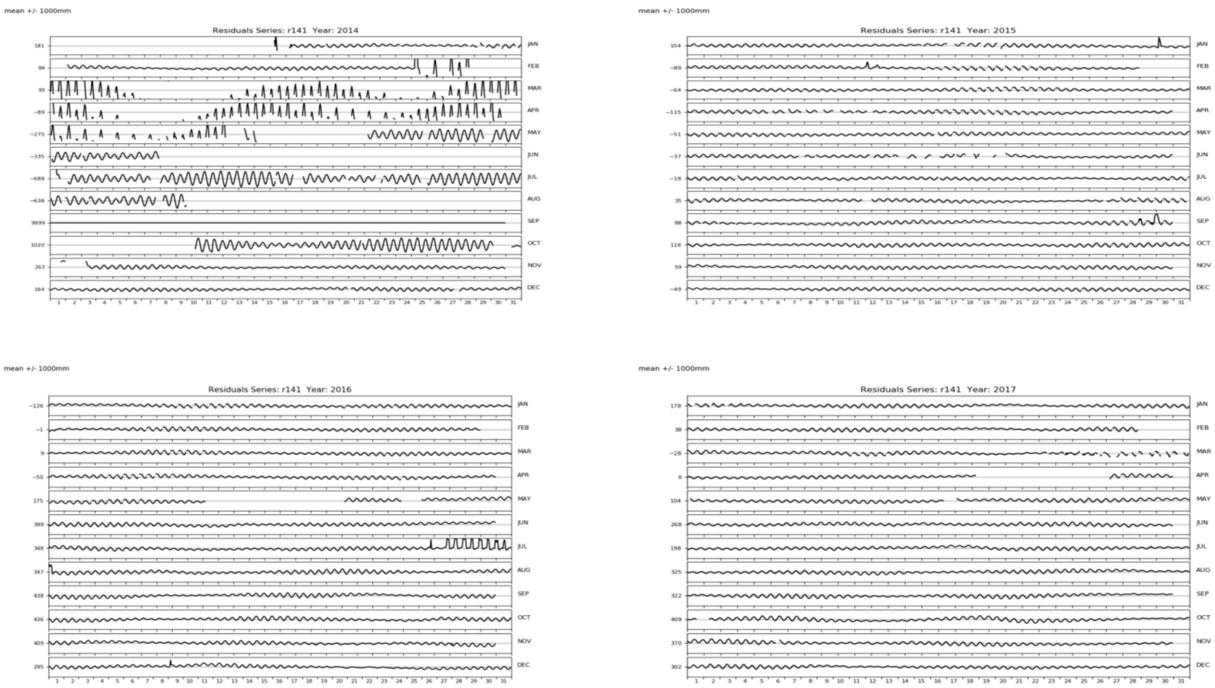
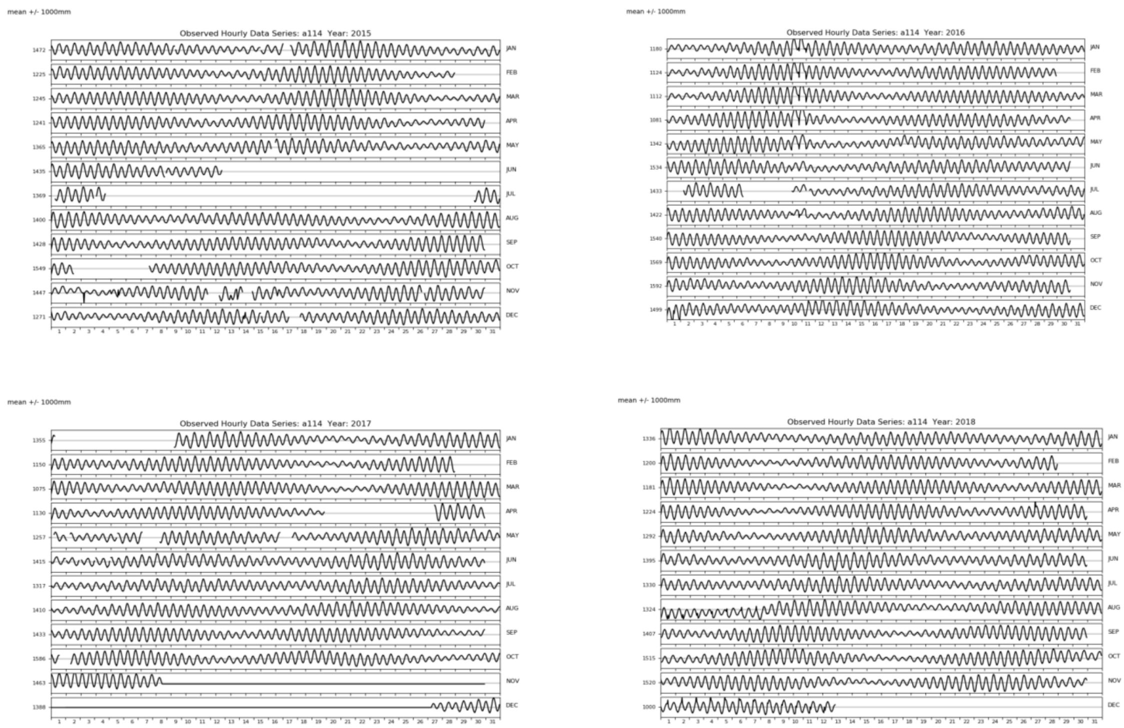


Figure 20: SLP64 results between observed data and residuals for the Visakhapatnam station from the period (2014-2017)

5.5 Kakinada station:

The observed hourly data, the predicted tides, and the residuals (defined as observed data minus predicted tides) have a similar SLP64 format and file naming convention. Each file consists of a year of values at an hourly sampling interval. Figure 21 shows the observed and residuals outputs from SLP64 software.



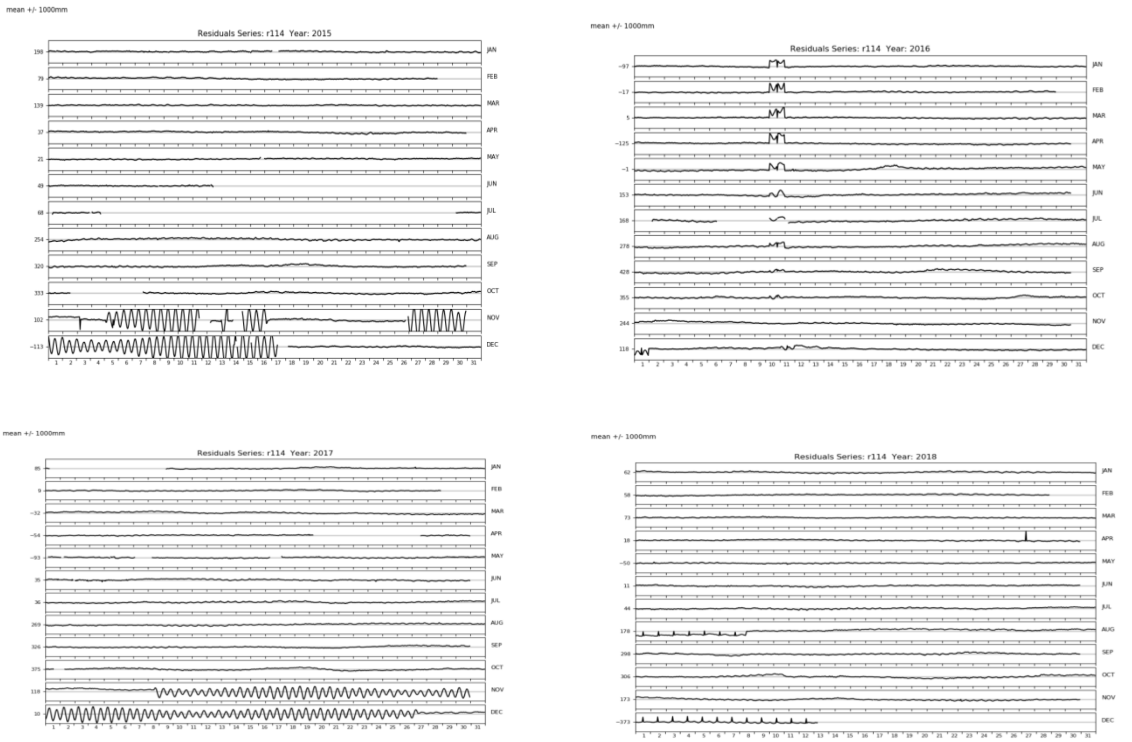


Figure 21: SLP64 results between observed data and residuals for the Kakinada station from the period (2015-2018).

5.6 Tidal analysis and prediction

This package is fixed to allow a maximum of 68 constituents which can be achieved through analysis of a record length of 366 days. The program will function for record lengths less than 366 days; however, the number of outputharmonic constituents will be fewer and subsequently the quality of the constituents will be less.For instance, a record length of 30 days will produce 30 constituents and a record length of 14 days will produce 11 constituents. From the hourly plots choose a time span of at least 366 consecutive days for a period with apparently good data (no obvious spikes or corruption) and with minimal data gaps. The time span can begin on any hour of any day and can end on any hour of any day in the following year, as long as onedoes not exceed 13 months. One restriction is that the time span CAN NOT cross the century boundary; that is, the input span cannot include the consecutive months of December 1999 and January 2000. Some of the tidal constituents have been shown below for the five tide gauge stations.

Chennai station tidal constituents

NO	NAME	FREQUENCY	STN	M-Y/ M-Y	Amplitude (A)	Phase (G)	AL	GL
1	Z0	0	106	111/1211	79.3259	0	79.3259	0
2	SSA	0.00023	106	111/1211	6.3481	136.11	6.3481	295.74
3	MSM	0.00131	106	111/1211	1.8594	345.81	1.8594	235.75
4	MM	0.00151	106	111/1211	0.6049	353.37	0.6049	65.44
5	MSF	0.00282	106	111/1211	0.6514	208.98	0.6514	170.99
6	MF	0.00305	106	111/1211	1.6528	332.31	1.6528	93.94
7	ALP1	0.0344	106	111/1211	0.0631	320.76	0.0628	343.36

8	2Q1	0.03571	106	111/1211	0.1425	15.04	0.1477	286.33
9	SIG1	0.03591	106	111/1211	0.2012	3.03	0.1997	99.69
10	Q1	0.03722	106	111/1211	0.3353	11.93	0.3397	357.17
11	RHO1	0.03742	106	111/1211	0.0635	342.3	0.0593	151.63
12	O1	0.03873	106	111/1211	3.0067	237.41	2.9867	296.53
13	TAU1	0.03896	106	111/1211	0.2094	252.31	0.2315	293.2
14	BET1	0.04004	106	111/1211	0.0439	270.17	0.0437	41.32
15	NO1	0.04027	106	111/1211	0.6509	243.67	0.8209	172.4

Cochin station tidal constituents

NO	NAME	FREQUENCY	STN	M-Y/	M-Y	A	G	AL	GL
1	Z0	0	107	114/	115	70.6548	0	70.6548	0
2	SA	0.00011	107	114/	115	11.0044	340.3	11.0044	162.54
3	SSA	0.00023	107	114/	115	3.4313	90.85	3.4313	248.96
4	MSM	0.00131	107	114/	115	0.5649	54.78	0.5649	136.17
5	MM	0.00151	107	114/	115	2.0872	308.74	2.0872	95.04
6	MSF	0.00282	107	114/	115	1.2756	244.49	1.2756	112.18
7	MF	0.00305	107	114/	115	1.1131	335.29	1.1131	1.09
8	ALP1	0.0344	107	114/	115	0.0999	41.12	0.0881	358.42
9	2Q1	0.03571	107	114/	115	0.4829	343.12	0.4249	23.42
10	SIG1	0.03591	107	114/	115	0.3167	297.7	0.2666	39.11
11	Q1	0.03722	107	114/	115	1.9453	347.79	1.6606	172.39
12	RHO1	0.03742	107	114/	115	0.4521	347.43	0.3707	231.09
13	O1	0.03873	107	114/	115	8.7878	335.57	7.2377	304.49
14	TAU1	0.03896	107	114/	115	0.1895	83.44	0.2192	23.36
15	BET1	0.04004	107	114/	115	0.2343	29.44	0.1873	260.85

Nagapattnam station tidal constituents

NO	NAME	FREQUENCY	STN	M-Y/M-Y	A	G	AL	GL
1	Z0	0	124	111/1211	52.0443	0	52.0443	0
2	SSA	0.00023	124	111/1211	1.1457	114.14	1.1457	273.77
3	MSM	0.00131	124	111/1211	2.3624	345.61	2.3624	235.55
4	MM	0.00151	124	111/1211	0.9389	81.54	0.9389	153.6
5	MSF	0.00282	124	111/1211	0.5961	339.45	0.5961	301.46
6	MF	0.00305	124	111/1211	1.9245	3.14	1.9245	124.78
7	ALP1	0.0344	124	111/1211	0.0248	322.88	0.0247	344.63
8	2Q1	0.03571	124	111/1211	0.3247	2.7	0.3404	272.72
9	SIG1	0.03591	124	111/1211	0.2756	57.3	0.2736	153.73
10	Q1	0.03722	124	111/1211	0.4979	65.07	0.5067	49.65
11	RHO1	0.03742	124	111/1211	0.1425	224.68	0.1328	33.7
12	O1	0.03873	124	111/1211	2.2622	224.7	2.246	283.76
13	TAU1	0.03896	124	111/1211	0.0337	214.68	0.0373	256.48
14	BET1	0.04004	124	111/1211	0.1449	253.68	0.1445	24.83
15	NO1	0.04027	124	111/1211	0.4134	240.84	0.5234	173.89

Visakhapatnam station tidal constituents

NO	NAME	FREQUENCY	STN	M-Y/ M-Y	A	G	AL	GL
1	Z0	0	141	115/1215	70.974	0	70.974	0
2	SSA	0.00023	141	115/1215	4.4947	151.96	4.4947	311.53
3	MSM	0.00131	141	115/1215	1.5917	198.63	1.5917	115.16
4	MM	0.00151	141	115/1215	2.4197	131.57	2.4197	195.68
5	MSF	0.00282	141	115/1215	0.747	37.28	0.747	17.92
6	MF	0.00305	141	115/1215	0.7722	255.35	0.7722	35.56
7	ALP1	0.0344	141	115/1215	0.263	136.13	0.214	123.81
8	2Q1	0.03571	141	115/1215	0.2536	7.09	0.2036	271.62
9	SIG1	0.03591	141	115/1215	0.4572	9.26	0.3653	59.83
10	Q1	0.03722	141	115/1215	0.1109	170.44	0.089	138.26
11	RHO1	0.03742	141	115/1215	0.3527	21.34	0.2585	135.14
12	O1	0.03873	141	115/1215	3.8289	242.13	3.0679	273.1
13	TAU1	0.03896	141	115/1215	0.0687	263.37	0.0873	271.54
14	BET1	0.04004	141	115/1215	0.2523	219.85	0.1954	347.49
15	NO1	0.04027	141	115/1215	0.2728	226.93	0.3049	139.3

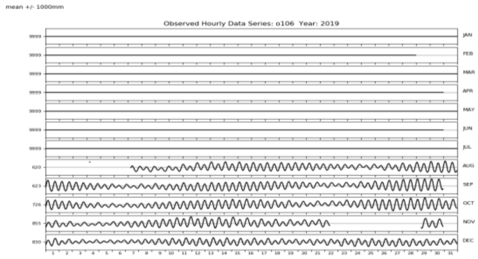
Kakinada station tidal constituents

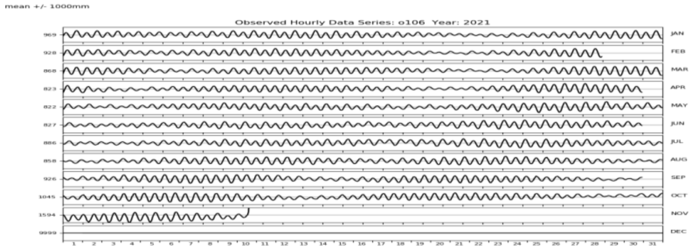
NO	NAME	FREQUENCY	STN	M-Y/M-Y	A	G	AL	GL
1	Z0	0	114	114/1214124.4720	0	124.472	0	
2	SSA	0.00023	114	114/1214	14.8157	154.41	14.8157	313.5
3	MSM	0.00131	114	114/1214	2.3438	35.04	2.3438	122.09
4	MM	0.00151	114	114/1214	0.4492	133.29	0.4492	286.12
5	MSF	0.00282	114	114/1214	1.0339	43.42	1.0339	283.3
6	MF	0.00305	114	114/1214	1.0398	327.88	1.0398	6.86
7	ALP1	0.0344	114	114/1214	0.2047	51.74	0.1742	156.14
8	2Q1	0.03571	114	114/1214	0.3475	51.01	0.2962	243.43
9	SIG1	0.03591	114	114/1214	0.3658	296.8	0.3051	192.92
10	Q1	0.03722	114	114/1214	0.1387	155.73	0.1166	139.96
11	RHO1	0.03742	114	114/1214	0.2889	174.81	0.2397	219.45
12	O1	0.03873	114	114/1214	3.48	243.66	2.876	19.71
13	TAU1	0.03896	114	114/1214	0.2133	197.62	0.2536	307.39
14	BET1	0.04004	114	114/1214	0.2477	115.45	0.198	159.84
15	NO1	0.04027	114	114/1214	0.3064	68.93	0.2717	228.7

5.7 Data short gaps and spikes

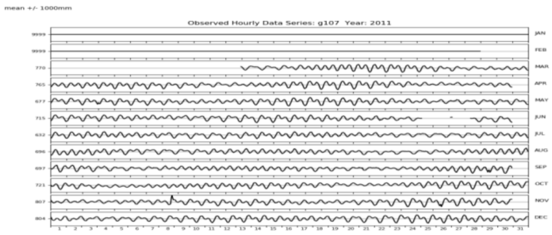
Short gaps and data spikes are a common problem in most sea level time series records. This section describes a technique for using interpolation to correct such problems. The best procedure for filling gaps is to replace the missing data flags (i.e. 9999) with quality controlled data from an auxiliary sea level gauge that is linked to the same datum. The predicted tide method for filling gaps requires yearly files of observed and corresponding predicted data. The predicted tides are shifted in time to match the timing characteristics of the observed series. The residuals between the predicted tides and the observed data are calculated. Then, a linear interpolation between the end points of the gap in the residual series is performed and each interpolation constant is added to the shifted predicted tides over thespan of the gap. For the time being, we are focusing on five stations: Chennai, Cochin, Nagapattnam, Visakhapatnam, and Kakinada (see Figures for examples).

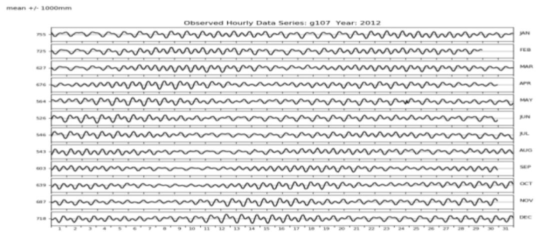
Chennai Station:

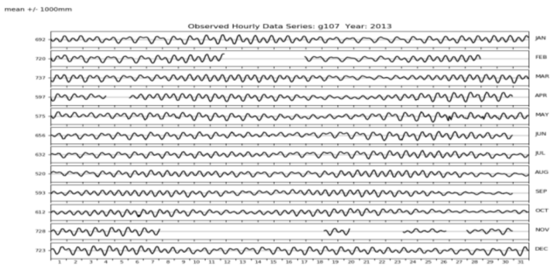


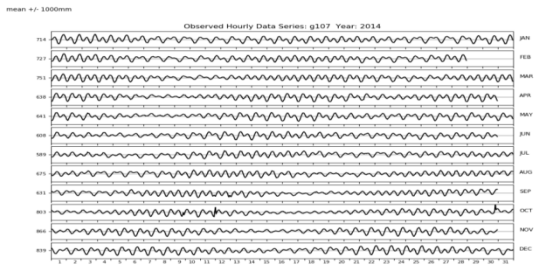


Cochin Station:

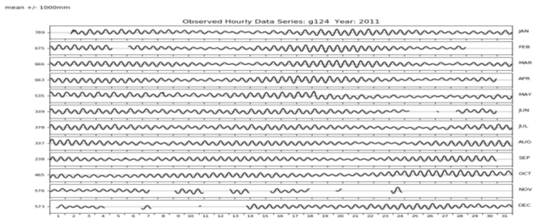


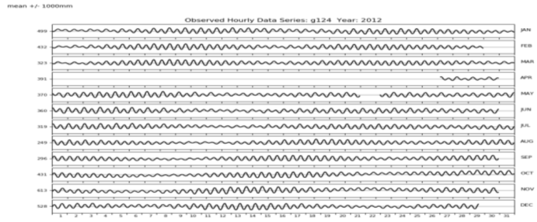


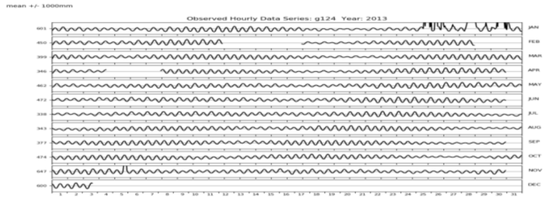


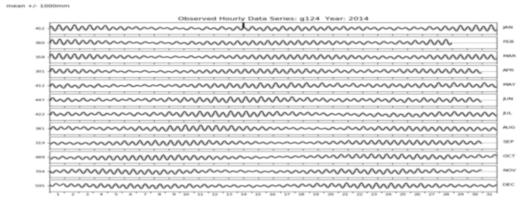


Nagapattnam Station:

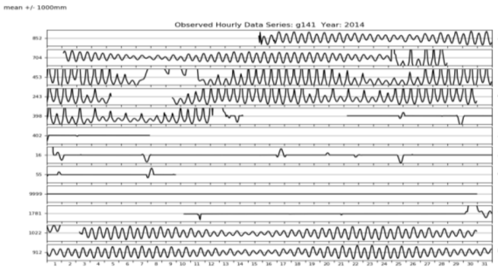


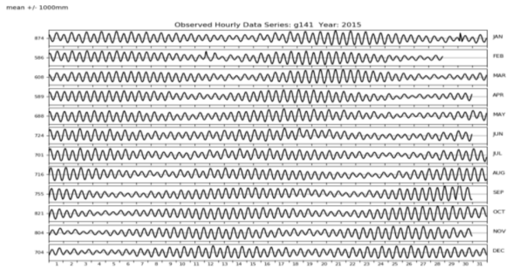


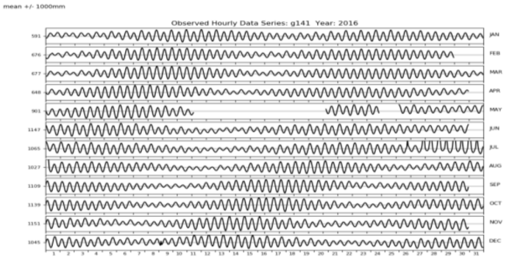


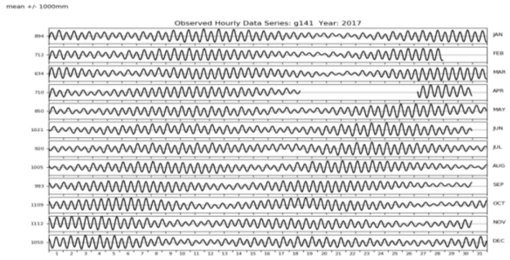


Visakhapatnam station:

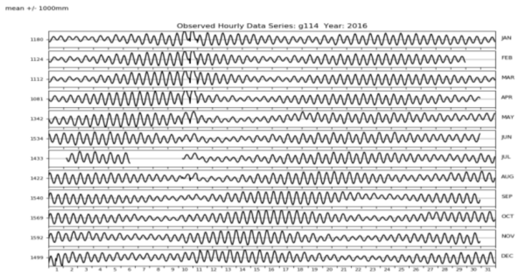


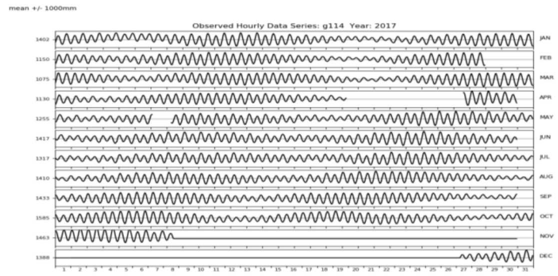


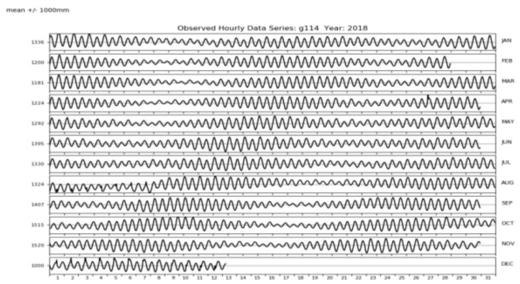


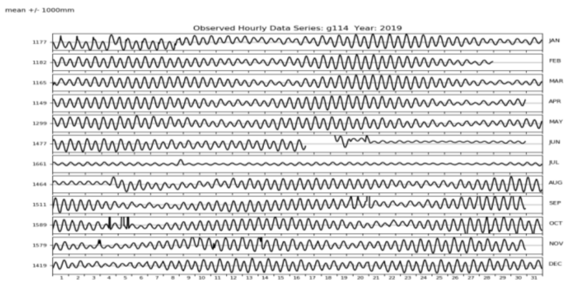


Kakinada station:









6. Conclusions and Future Scope:

This study presents data processing and quality control procedures demonstrated on five tide gauge stations maintained by INCOIS. The automated QC methods and validation techniques outlined here are designed to be applied consistently across the entire network of 36 tide gauge stations along the Indian coastline. Ensuring precise and continuous monitoring of water levels across all stations will enhance the scientific accuracy and operational utility of the data. Integration of tsunami and seiche detection algorithms, combined with robust data analysis tools, enables timely identification of faulty data and supports real-time operational oceanography. Table 5 summarizes the flagged data points identified through the QC process at several stations: Chennai (2,568 points, 3.3%), Cochin (3,349 points, 3.8%), Nagapattinam (451 points, 0.4%), Visakhapatnam (148 points, 0.1%) and Kakinada (42 points, 0.04%).

Station Name	Total data	Out of control Flag (4)	Outliers Flag (2)	Quality data Flag (1)	Total Flagged Data	Missing data Flag (9)	Percentage of flagged data
Chennai	75662	2568	0	73418	2568	47,050	3.3 %
Cochin	87421	417	2932	83,968	3,349	36,789	3.8 %
Nagapattnam	99317	17	434	98852	451	14,635	0.4 %
Visakhapatnam	100100	134	14	99,952	148	15,644	0.1 %
Kakinada	87279	10	32	87,226	42	26,673	0.04 %

Table 5. Details of the number of quality and flagged based on the QC procedure.

Validation of observed data against predicted tide levels using demeaning, detiding, residual analysis, RMSE, and standard deviation demonstrated strong agreement between observations and model outputs. To improve the overall tide gauge network's accuracy and scientific utility, it is vital to maintain continuous, high-quality monitoring. These improvements will support more precise tidal representations, facilitate real-time tsunami and seiche detection, and advance the use of tide gauge data in operational oceanography and high-frequency sea-level variation studies.

References:

1. Harikumar, R., et al. "Ship-mounted real-time surface observational system on board Indian vessels for validation and refinement of model forcing fields." Journal of Atmospheric and Oceanic Technology 30.3 (2013): 626-637.

2. UNESCO/IOC. 2020. *Quality Control of in situ Sea Level Observations: A Review and Progress towards Automated Quality Control*, Vol. 1. Paris, UNESCO. IOC Manuals and Guides No.83. (IOC/2020/MG/83Vol.1)

3. bservations: A Review and Progress towards Automated Quality Control, Vol. 1. Paris, UNESCO. IOC Manuals andGuides No.83. (IOC/2020/MG/83Vol.1).

4. Caldwell, P., and B. Kilonsky, 1992. *Data processing and quality control at the TOGA Sea Level Center.* Joint IAPSO-IOC Workshop on Sea Level Measurements and Quality Control, Paris, 12-13October, 1992. IOC Workshop Report No. 81, UNESCO. pp. 122-135.

5. SLP64 Software Manual

6. Intergovernmental Oceanographic Commission, 1992. *Joint IAPSO-IOC Workshop on Sea Level Measurments and Quality Control*. Workshop Report No. 81. Paris, 12-13 October, 1992. page 16.