Hohonu

Quality Control Procedures for Real-Time Water Level Data

Date	Revision Description	Notes
July 2022	• Original Document	
November 2023	• Formal organization of Hohonu QC methodology • Added comparison of raw and clean data • Added comparison between clean Hohonu and NOAA data	• Identified future upgrades to QC methodology to increase precision and alignment with NOAA methods

Purpose and Scope

Hohonu provides precision water level monitoring and forecasting to help communities prepare for and respond to flooding. The organization maintains a network of over 100 water level sensors located in 14 states, and it uses the latest technologies in hardware, software, and data science in order to deliver reliable and accessible water level data to its customers.

Adherence to rigorous quality assurance (QA) and quality control (QC) protocols is essential for providing high-quality data to stakeholders. The U.S. Integrated Ocean Observing System (IOOS) provides authoritative procedures through the Quality Assurance/Quality Control of Real-Time Oceanographic Data (QARTOD) program[1]. Hohonu implements these procedures within its own QA/QC program.

This document provides an overview of the QC methodology used by Hohonu to 'clean' raw sensor data and ensure the provision of high-quality real-time water level measurements to its customers. It is intended as technical support for users of Hohonu's data. While procedures documenting data acquisition QA is outside the scope of this document, additional information on sensor specifications, calibration and maintenance, and leveling are referenced at the end[2,3].

Data Processing

After water level data are acquired from a Hohonu sensor, a set of internal and external QC steps are applied to convert 'raw' data into 'clean' data ready for storage, reporting, and forecasting (Figure 1). The below processes are implemented for all live Hohonu sensors.

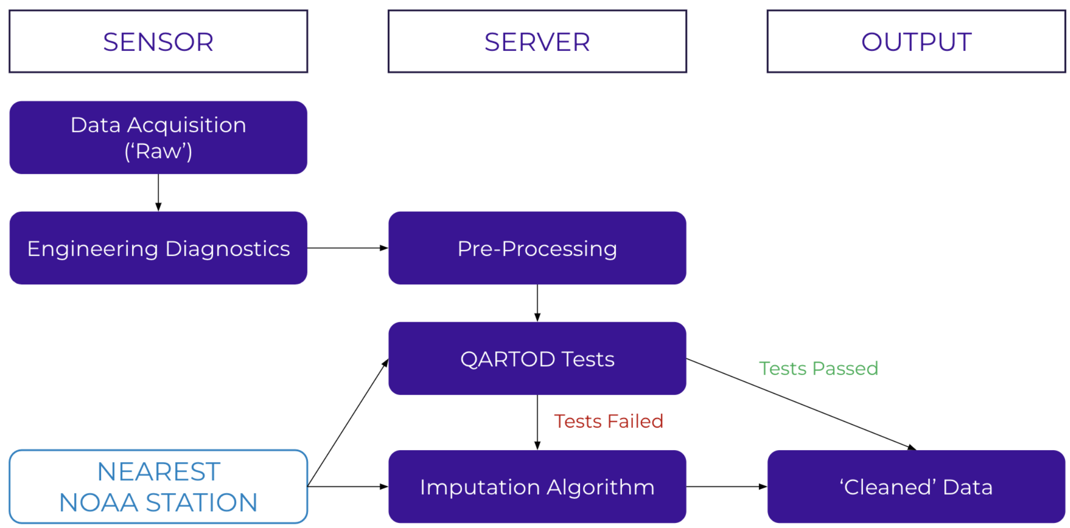

Figure 1. Overview of the Hohonu QC methodology

<u>1. Engineering Diagnostics</u>
Frequency: every sample
Hohonu sensors sample at a rate of 1 Hz and internally calculate a 6-minute mean from them. The first step of QC is performed onboard by removing all samples measured outside of 3 standard deviations from the mean. Data are then transmitted to the Hohonu server.

<u>2. Data Pre-Processing</u>
Frequency: every 6 hours
Prior to implementing QARTOD tests, Hohonu pre-processes water level datasets by first aligning samples to a 6-minute interval and identifying missing data points. Duplicate measurements and any samples measured at less than 0.5m D2W (distance from sensor to water) are removed. Flags are generated if sensor data indicate improper readings or when a sensor has become detached.

Finally, water level measurements are converted from the station datum (D2W) to a geodetic datum (i.e. NAVD88) or, if unavailable, a tidal datum (typically mean low low water, or MLLW) using a pre-calculated conversion factor. The purpose of this step is to make comparable Hohonu data to the nearest NOAA station for certain QARTOD tests and missing data imputation. The tidal data datum for each station is calculated per the NOAA-aligned methodology referenced below[1].

3. QARTOD Tests

Frequency: every 6 hours

The IOOS QARTOD procedures provide a hierarchy of 11 QC tests for real-time water quality data. Hohonu implements 6 of these tests on its server using automated scripts modified from those provided on the IOOS Github repository[4]. These tests include:

- *Time/Gap Test*: Determines that the most recent data point has been measured and received within the expected time window
- *Gross Range Test*: Checks that values are within reasonable range bounds. This range is established using 3 time the standard deviation
- *Spike Test*: Determines if there is a spike in the data given the previous point. Spikes are assessed using a cutoff of 3 times the standard deviations
- *Rate of Change Test:* Checks the first order difference of a series of values to see if there are any values exceeding a threshold. This threshold is established using 3 times the standard deviation.
- *Flat Line Test*: Checks for consecutively repeated values within a tolerance. The tolerance is set as 0.00001m.
- *Neighbor Test:* Compares Hohonu data with nearest NOAA station to detect anomalies. Flags occur when the stations' difference in subsequent samples is greater than 1m.

All test thresholds are determined automatically using the data. If any of these tests fail, the data point is flagged and its value is removed. Hohonu only uses the GOOD, FAILED, and MISSING flag types provided in the QARTOD procedures.

Of the remaining QARTOD tests, two are implemented onboard the sensor (Syntax Test and Location Test), one is performed manually upon daily station review (Climatology Test), and one is not applicable given the sensor type (Multivariate Test). Finally, Hohonu is presently assessing incorporating the Attenuated Signal Test into its automated procedures.

4. Imputation Algorithm

Frequency: every 6 hours

Any missing data or data that failed QARTOD tests are filled using an imputation algorithm. The algorithm is a multivariate imputer that estimates the missing values by utilizing available data in similar features[5]. In this case, the similar feature refers to data collected from the nearest NOAA station. Data are considered 'clean' upon passing all QARTOD tests or after imputation (Figure 2). Finally, clean data are converted back into the original station datum.

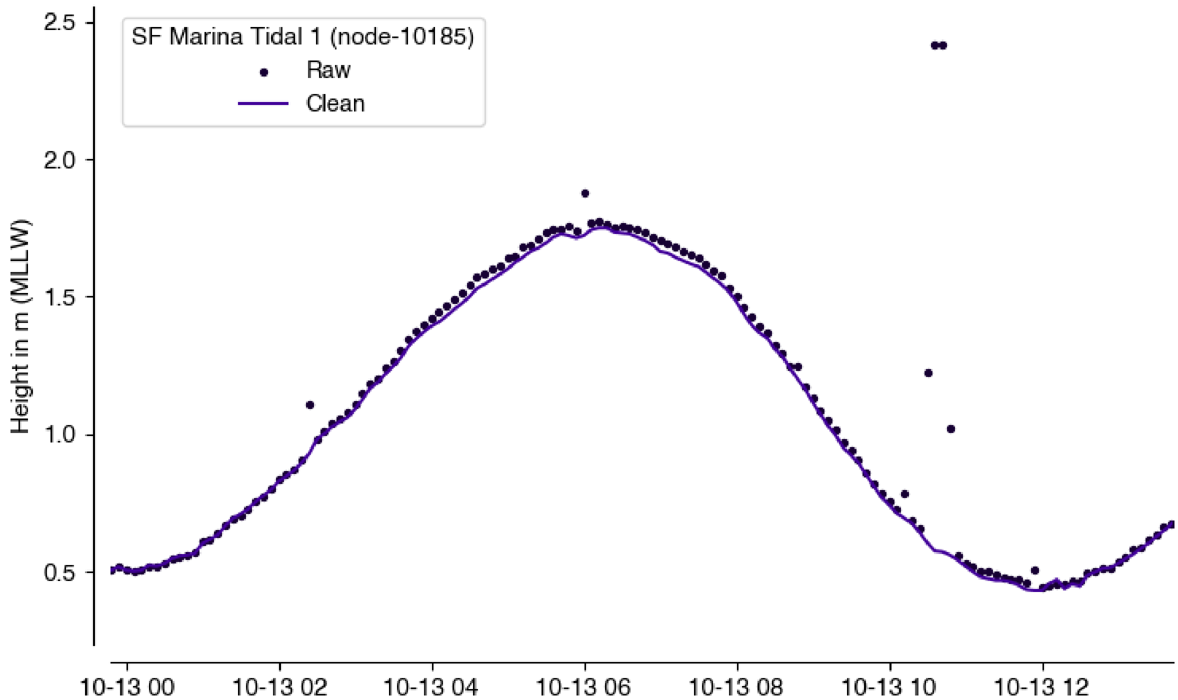

Figure 2. Comparison between raw and cleaned data collected in San Francisco, CA. In this example, the raw data failed the spike, rate of change, and neighbor tests in multiple points, and were accordingly flagged, removed, and imputed.

Data Storage and Reporting

All raw and cleaned water level data are stored on the Hohonu server and are available for further QC, reporting, manipulation, and forecasting. Data may also be compared against NOAA control stations (Figure 3). Users can access these data via the user interface on the Hohonu Dashboard, via a mobile iOS app, or through the Hohonu Public API[6]. Present functionality also allows for water level data to be converted between different datums. References for data access are provided below.

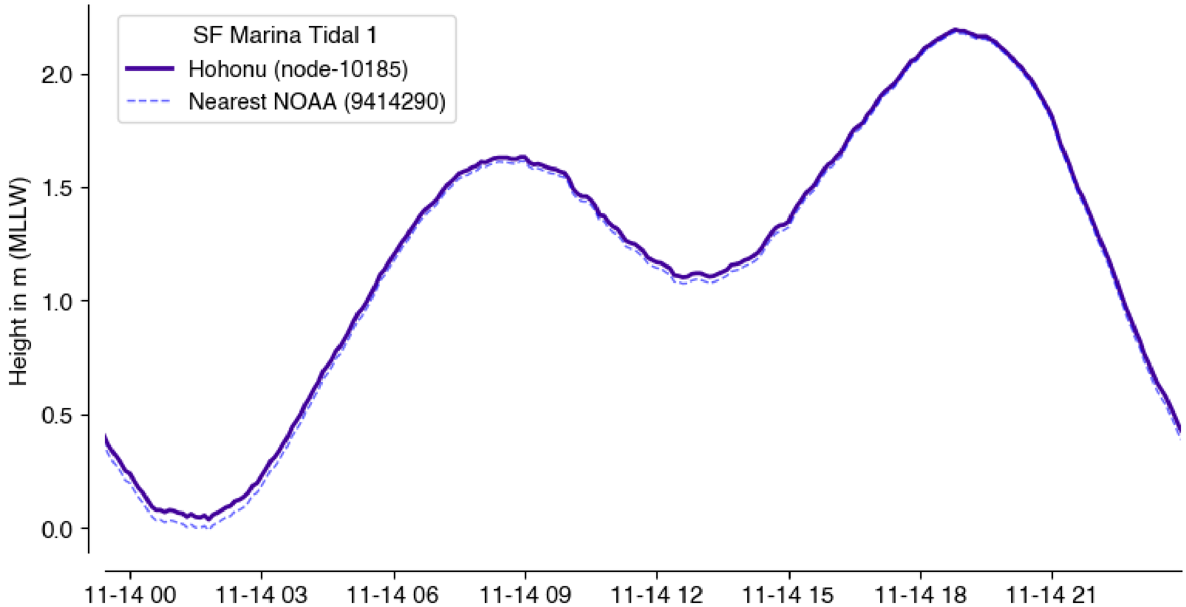

Figure 3. Comparison between cleaned Hohonu data and data from the nearest NOAA station (located approximately 1.4 miles away). Data at these stations are strongly correlated (Pearson correlation of 0.999 for the plotted time series and 0.996 over the preceding 5 months).

Any flags generated during the QC process are also stored on the Hohonu server and are examined daily by manually inspecting an aggregated station report. This inspection allows for response to issues with data stations. Hohonu is constantly working to update its protocols to align with the best available methodologies. Future updates will include more robust training modules for users and automatic thresholding to differentiate between bad data and acute events (e.g. storm surge or tsunamis).

References

1. National Ocean Service, Manual for real-time quality control of water level data: a guide to quality control and quality assurance for water level observations, 2021. https://repository.library.noaa.gov/view/noaa/29274
2. Sensor specs: https://maxbotix.com/pages/hrxl-maxsonar-wr-datasheet
3. Hohonu FAQ: https://www.hohonu.io/faqs
4. IOOS QC: QARTOD and other Quality Control tests implemented in Python. https://ioos.github.io/ioos_qc/
5. Scikit-learn IterativeImputer: https://scikit-learn.org/stable/modules/generated/sklearn.impute.IterativeImputer.html

6. Hohonu Public API. https://hohonu.readme.io/reference/getting-started-with-your-api