The Marine and Hydrokinetic ToolKit for Data Quality Control and Analysis

Preprint

Sterling Olson,[1] Rebecca Fao,[2] Ryan Coe,[1] Kelley Ruehl,[3]
Frederick Driscoll,[2] Budi Gunawan,[1] Chitra Sivaraman,[3]
Carina Lansing,[3] and Hristo Ivanov[2]

1 Sandia National Laboratories
2 National Renewable Energy Laboratory
3 Pacific Northwest National Laboratory

Presented at the 14th European Wave and Tidal Energy Conference
Plymouth, United Kingdom
September 5–9, 2021

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

Conference Paper
NREL/CP-5700-79696
October 2021

Contract No. DE-AC36-08GO28308

The Marine and Hydrokinetic ToolKit for Data Quality Control and Analysis

Preprint

Sterling Olson,[1] Rebecca Fao,[2] Ryan Coe,[1] Kelley Ruehl,[3]
Frederick Driscoll,[2] Budi Gunawan,[1] Chitra Sivaraman,[3]
Carina Lansing,[3] and Hristo Ivanov[2]

1 Sandia National Laboratories
2 National Renewable Energy Laboratory
3 Pacific Northwest National Laboratory

Suggested Citation
Olson, Sterling, Rebecca Fao, Ryan Coe, Kelley Ruehl, Frederick Driscoll, Budi Gunawan, Chitra Sivaraman, Carina Lansing, and Hristo Ivanov. 2021. *The Marine and Hydrokinetic ToolKit for Data Quality Control and Analysis: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-5700-79696. https://www.nrel.gov/docs/fy22osti/79696.pdf.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5700-79696
October 2021

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Water Power Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

The Marine and Hydrokinetic ToolKit for Data Quality Control and Analysis

Sterling Olson, Rebecca Fao, Ryan Coe, Kelley Ruehl, Frederick Driscoll, Budi Gunawan, Chitra Sivaraman, Carina Lansing, and Hristo Ivanov

Abstract—**The ability to handle data is critical at all stages of marine energy development. The Marine and Hydrokinetic Toolkit (MHKiT) is an open-source marine energy software, which includes modules for ingesting, applying quality control, processing, visualizing, and managing data. MHKiT-Python and MHKiT-MATLAB provide robust and verified functions that are needed by the marine energy community to standardize data processing. Calculations and visualizations adhere to International Electrotechnical Commission technical specifications and other guidelines. A resource assessment of National Data Buoy Center buoy 46050 near PACWAVE is performed using MHKiT and we discuss comparisons to the resource assessment provided performed by Dunkle et al. (2020).**

Keywords—**Marine Energy, MHKiT, Software**

I. INTRODUCTION

High-quality resource and performance data collected through simulation, laboratory, and open-water measurements for validation of marine energy (ME) technologies provide critical information for design optimization, technology evaluation/demonstration, and certification. Once collected, these data require appropriate, robust, and verified data processing techniques such as those set by the International Electrotechnical Commission (IEC). Without a trusted ME data processing tool, continuous, repeated effort is needed to discover, adapt, and/or develop processing capabilities by multiple teams leading to longer development timelines and increased project costs.

The Marine and Hydrokinetic Toolkit (MHKiT) is an open-source functional toolbox for data ingestion, quality control (QC), analysis, visualization, and management that provides the ME community with high-quality validation and performance calculations critical for device development and design iteration. MHKiT provides unparalleled access to ME-specific standardized, open-source data processing, management, and QC software solutions verified by continuous integration tests and developed by subject matter experts from the National Renewable Energy Laboratory, Sandia National Laboratories, and Pacific Northwest National Laboratory. Experts in each of these labs have diverse expertise in areas including land and marine measurement; data acquisition and instrumentation system development; data management; software development; laboratory and field testing; open-source code development; ME device testing and simulation; resource characterization; and environmental impacts covering the gamut of ME-specific and support expertise.

The ME sector is best served by an open-source platform, such as MHKiT, because it allows users to easily apply standards and methods, collaborate on a common set of data processing functionality, and modify them to fit their specific needs. Providing the ME community with the ability to quickly and easily perform data analysis catalyzes the industry by allowing researchers and businesses to focus on cutting-edge research and producing high-impact results at a lower cost than would otherwise be needed. Individuals may perform data acquisition and calculations by following a variety of specifications, such as those set by the IEC [1]. This approach tends toward the repeated development of code across the industry. MHKiT provides access to comprehensive ME-specific standardized, open-source data processing, management, and QC software solutions verified by appropriate experts. Further, the open-source nature of the tool allows MHKiT to be customized by users to fit their specific needs and collaborate with others to improve and grow organically with the industry and community. A verified suite of ME data processing functions aids the ME sector and adds value by:

- Enabling rapid data processing
- Eliminating common code duplication
- Offering reproducible QC, analysis, and visualization capabilities
- Providing a standardized, referenceable, and readable codebase
- Using a common development platform where issues are discussed and features are expanded
- Assisting developers in device certification for insurance, regulator bodies, and investors.

2024: Wave / tidal device development and testing.

Ultimately, these capabilities will accelerate technology development and reduce costs by providing more information to feed modeling, design, and validation activities.

As can be seen from the ME code on the MRE CodeHub [2], many tools are available that solve specific problems in ME; however, no ME software has sought to provide general data processing from these tools. The release of MHKiT in January 2020 brought not only a platform for community-developed ME functions but provided the first codification of IEC 62600 technical specifications for marine energy. This open-source code base allows current and future ME industry members to quickly and easily apply IEC specifications to their data—a critical step in gaining regulatory compliance, insurance, and investors. Each of these bodies may read, review, and reference the code as free and open source software without the need to buy and run the code. Further, as an open-source code, ME users can easily modify and expand the code for their specific interests. The user may then choose to submit changes for publication in the open-source code base following a review by the MHKiT team. By submitting the code to MHKiT, the user will have access to that function in the future through the MHKiT software installation, which is maintained, tested, and hosted on GitHub.

II. MHKiT Functionality and Structure

The initial MHKiT release included modules for wave, tidal, and river energy converter power performance, resource assessments, data quality control, and functions for data ingestion with functionality based on the technical specifications developed by IEC TC114 62600 [1]. MHKiT leverages prior investments by the U.S. Department of Energy, the offshore engineering and measurement community, and the ME sector. With versions in Python [3] and Matlab [4], MHKiT is designed to reach a broad user base, and thereby meeting ME data processing needs for users with either programming language preference. Detailed documentation and examples are included for all modules, which enable the easy application of MHKiT to a variety of data sets. Since the initial release, MHKiT v0.3.1 has added capabilities for mechanical loads, power quality, and resource characterization (Fig. 1).

Each of the resource-based modules (wave, river, and tidal) have a respective I/O (input/output), resource, performance, and graphics module for handling data and calculations specific to that resource. The I/O module in each section contains functions for reading data specific to the resource. For example, the river I/O module can request data from U.S. Geological Survey servers [5], the tidal module can request data from National Oceanic and Atmospheric Administration servers [6], and the wave module can request data from National Data Buoy Center (NDBC) servers [7]. That ability to quickly request ME data from popular databases increases the utility of data processing in MHKiT and allows for rapid data analysis.

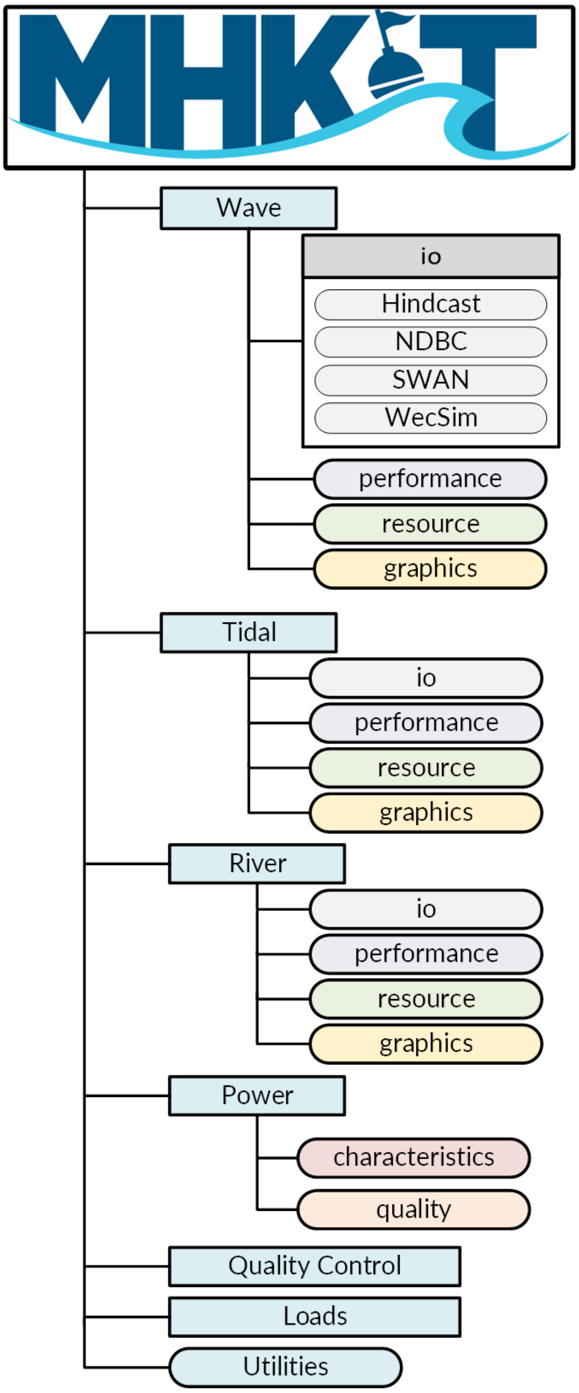

Fig. 1. The MHKiT v0.3.1 file structure. Folders are shown as squares and files are shown as pill boxes. MHKiT modules are shaded light blue. Submodules are shown in a consistent color across modules.

Further increasing the I/O functionality of MHKiT for popular data types and sources is a near-term goal with plans of adding hosted data, such as the Coastal Data Information Program [8], and instrumentation-specific data ingestion, such as acoustic Doppler current profiler (ADCP) readings.

The resource submodules contain functions utilized to characterize the available resource in a study or testing region. The functionality is largely focused on the IEC 62600 technical specifications for resource characterization for each respective resource (e.g., wave, tidal, river). In the river module, this includes functions to calculate the Froude number and exceedance probability. The tidal resource module includes a function to calculate the principal flow directions. The wave resource module contains functions for calculating idealized spectrums, and metrics such as significant wave height and energy period.

The performance submodules are also based on IEC 62600 technical specifications for device power performance characterization. For river and tidal this includes projected area and tip-speed-ratio calculations. For wave, the performance module includes functionality for calculating the capture length and power matrices.

The graphics submodules contain functions for visualizing outputs from other MHKiT functions. Graphics functionality includes generating capture length and power matrices (wave), tidal flow and joint probability distributions (tidal), and power curves (river). Where a function may be useful across different resources it is written in one module (e.g., river) but is made accessible by calling from the other module (e.g., tidal).

The other modules within MHKiT are designed to be resource-type agnostic. The QC module contains functionality for monitoring and flagging the quality of data. These functions include checks for time stamps, data spikes, data outliers, and stagnant data, among others. It is recommended that users apply the QC functions to their data first in order to identify and handle anomalies before proceeding with data analysis.

The loads module contains functionality for assessing mechanical loads following the IEC 62600-3. Functionality includes capabilities to calculate damage equivalent loads, loads statistics, and blade moments for rotor-style devices. The loads module also contains a graphics submodule for visualizing loads statistics.

The power module is broken down into two submodules. The first is the characteristics submodule, which contains general code for calculating quantities of interest from power measurements that are useful for analysis not specifically associated with the 62600-30 technical specification for power quality. The characteristics submodule contains code for calculating active power, among others. The second submodule in the power module is the quality module. This module contains code for calculating quantities specifically used for assessing power quality following the 62600-30 technical specification. These functions include capabilities to compute harmonics, interharmonics, and harmonic distortion.

The final module currently contained in MHKiT is the utils module, which contains functions that aid users to process their data more easily with MHKiT. For example, the utils module contains functions for converting time stamps to the MHKiT prescribed date-time format. The utils module also contains functionality that is useful across all the modules, such as calculating statistics.

To aid in the use and understanding around the functionality within MHKiT, detailed documentation (mhkit-software.github.io/MHKiT/) and examples for all the functions are available. Providing information about the general MHKiT development, sources where MHKiT functionality is pulled from, installation instructions, examples, and collaboration options, the MHKiT documentation is a go-to resource for learning more. The examples walk the users through accessing data through the I/O functions, applying MHKiT processing to data, and displaying the results. Additionally, the documentation provides in-depth information about each MHKiT function including the required and optional input parameters, what data types and units are expected, and what the function outputs will be.

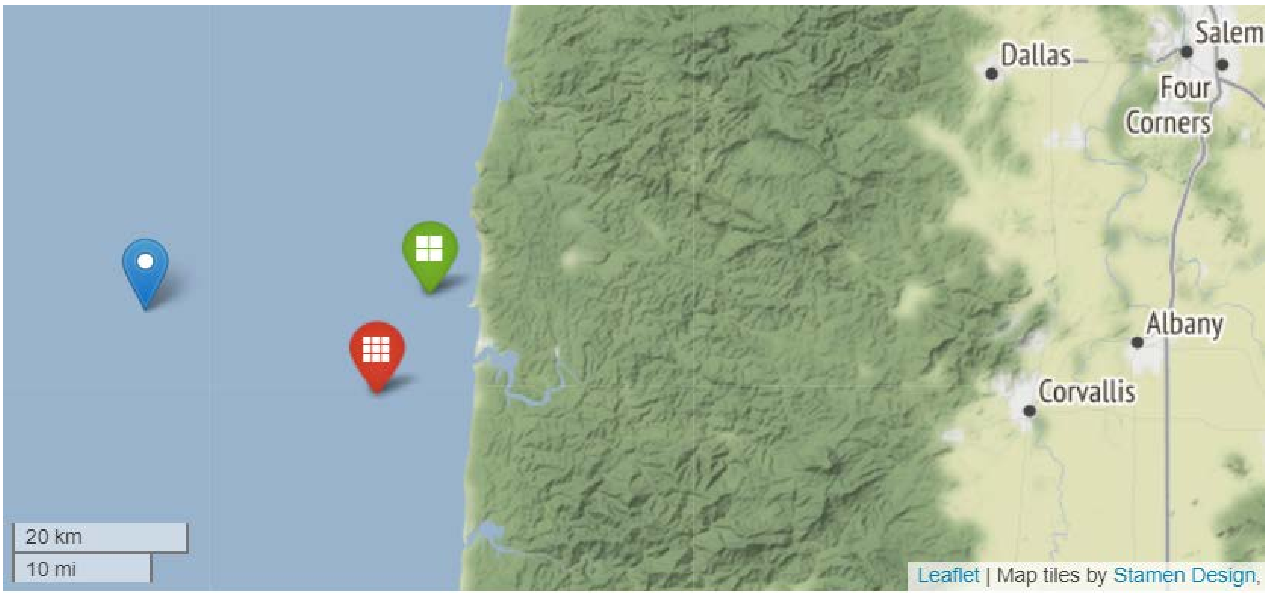

Fig. 2. NDBC 46050 shown in blue with PACWAVE north as a green four-square grid and PACWAVE south as a red nine-square grid.

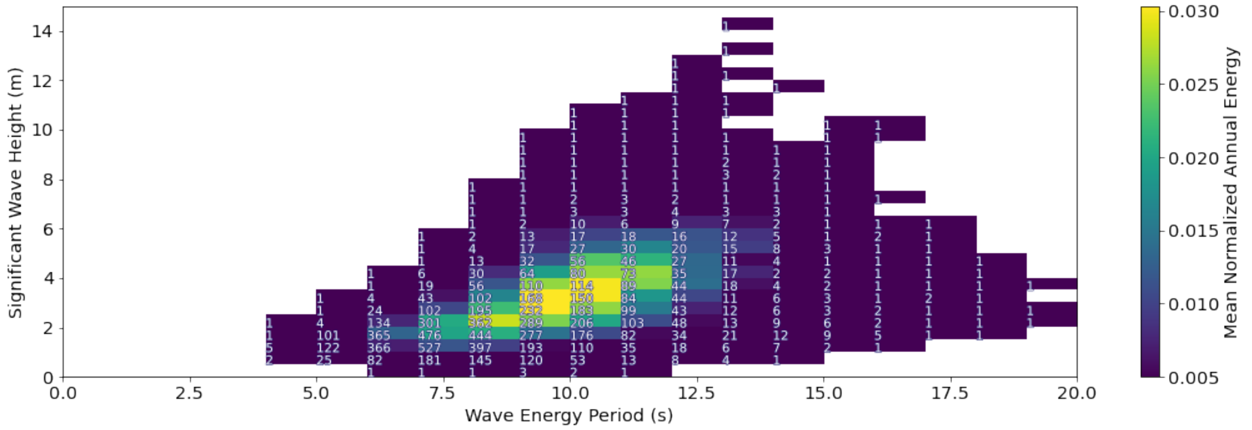

Fig. 3. NDBC 46050 energy flux (J) mean annual histogram for years 1996 to 2019 for each month. Bins are numbered by number of hours each year a sea state occurs.

III. APPLICATION AND DEMONSTRATION

An application of MHKiT is shown herein for considering wave resource analysis using omnidirectional wave data and following guidance codified into MHKiT from the IEC TS 62600-101 [1]. Analysis is compared against [9], who provided analysis of the PACWAVE sites using SWAN hindcast data. Here, comparisons will be made between the NDBC buoy 46050 and the PACWAVE south site, which is approximately 40 km closer to shore with a depth of 60 m, as shown in Fig. 2. As discussed in [10], the primary differences seen in a resource assessment as waves approach the shore for depths greater than 8 m is in the directionality; therefore, only omnidirectional wave metrics are compared in this study.

Access to spectral wave density data from the NDBC database is provided directly within MHKiT, allowing for easy and rapid data analysis. Using the MHKiT NDBC request data function, spectral wave density data from 1 January 1996 to 31 December 2020 were pulled from NDBC Buoy 46050 Stonewall Bank 20 nautical miles west of Newport, Oregon. The buoy is approximately 40 km from the PACWAVE south site and has a water depth of 160 m.

With the data in the MHKiT environment, it can easily be processed using the MHKiT wave resource module to calculate significant wave height, energy period, peak period, mean zero-crossing period, and energy flux. From significant wave height and mean zero crossing period, additional metrics of interest can also be calculated, such as wave steepness, which was calculated using (1) [10].-

$$S_m = \frac{H_{m0}}{\frac{g}{2\pi}T_z^2} \qquad (1)$$

Where H_{m0} is significant wave height, g is gravity, and T_z is the mean zero-crossing period.

With the quantities of interest, calculated analysis of the site was initiated by recreating Fig. 2 in [9], which displays the average number of hours a particular sea state occurs over the 24-year data set and the average normalized annual energy in each sea state defined by significant wave

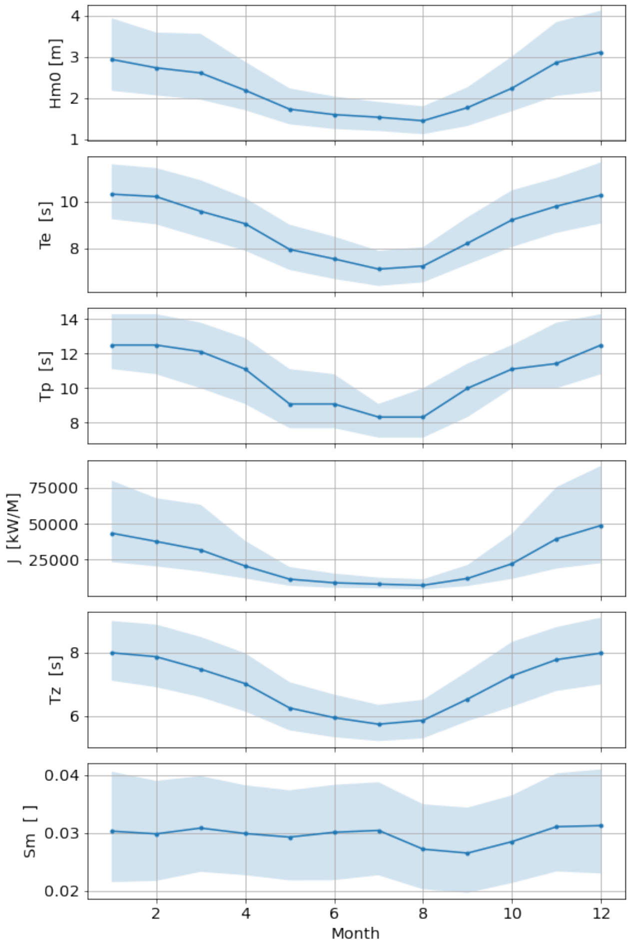

Fig. 4. NDBC 46050 mean for significant wave height (Hm0), energy period (Te), energy flux (J), peak period (Tp), mean zero-crossing period (T_z), and wave steepness (Sm) for years 1996 to 2020. The shaded area represents the 25th and 75th percentile of the data.

$$S_m = \frac{H_{m0}}{\frac{g}{2\pi}T_z^2} \qquad (2)$$

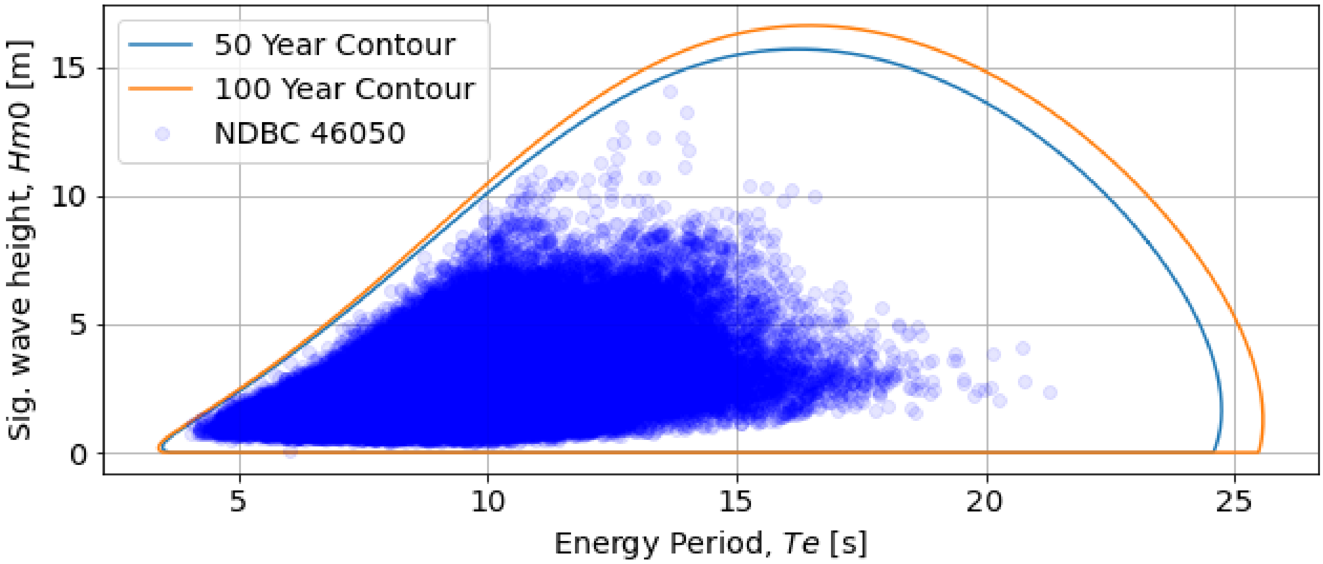

Fig. 5. Contours of 50-year and 100-year extreme sea-state contours for NDBC 46050 for 24 years of hourly spectral data from 1996 to 2020.

height and energy period. Missing data and outliers were dropped from the data set. Fig. 3 shows that the most frequent sea state occurs on average 527 hours per year at an energy period of 7.5 s and a significant wave height of 1.25 m. Dunkle et al. [9] reported a similar most frequent sea state with a slightly longer energy period at 8.5 s and having a 1.75-m wave height for 528 hours per year. The highest average annual energy sea state at buoy 46050 occurs at an energy period of 9.5 s and a significant wave height of 2.75 m and occurs on average for 168 hours per year. Further, [9]. reported the most energetic sea state on average to occur at 2.75 m and 10.5 s for 231 hours per year. The comparison is similar, as expected, for the omnidirectional wave resource. It is also worth noting that the analysis that the bin at 9.5 s and 2.5 m occurs for 232 hours and is 99.7% of the maximum annual energy bin reported above and the number of hours seen agrees with the maximum reported by [9].

Next, statistics were found for each of the quantities of interest grouped by month, as shown in Fig. 4. The shaded region shows the values between the 25% and 75% quantiles, whereas the line is showing the average value for a given month. The top subplot in Fig. 4 shows significant wave height to have a maximum mean value in December at 3.11 m, which corresponds well with Fig. 5 in [9]. The higher significant wave height also brings higher variability in the winter months than in the summer months, which shows a minimum value around 1.44 m in August.

The second and third plot from the top in Fig. 4 shows energy period and peak period each having a maximum value in January at 10.3 s and 12.5 s, respectively. The minimums also correspond to the same month of July with values of 7.12 s, and 8.33 s for energy period and peak period, respectively. Dunkle et al. [9] report a minimum energy period value of 8.5 s in July and maximum energy period of 11.3 s in February and do not report peak period monthly statistics.

The maximum energy flux occurs in December at 48889 kW/m while the minimum occurs in August at 7212 kW/m. These values come in lower than the results from [9]., which report values ranging between 70 and 80 kW/m in the winter months and a mean around 20 kW/m in the summer months.

The average monthly steepness, as described by (1), stays relatively constant throughout the year, ranging between 0.0265 and 0.0313. A discussion of monthly wave steepness was not held in [9] but would be interesting to compare for the PACWAVE south site.

A cumulative distribution of the energy flux, as described in the IEC TS 62600-101 [1], is shown in Fig. 6. The summer months have a lower maximum energy flux and are found left of the black data line representing the cumulative distribution of all collected data. April and October most closely follow the overall energy flux

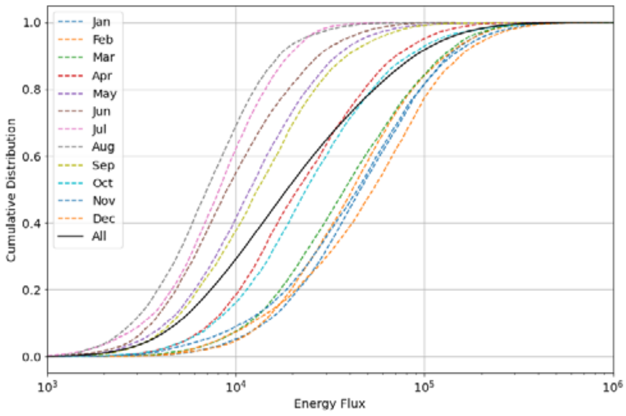

Fig. 6. NDBC 46050 energy flux (J) cumulative distribution for years 1996 to 2020 for each month. All months are shown as a solid black line.

distribution while the winter months show less variation than the summer months in their distribution.

Lastly, 50-year and 100-year extreme sea-state contours were calculated using MHKiT, as shown in Fig. 5 The environmental contours function in MHKiT was adapted from the Wave Energy Converter Design Response Toolbox (WDRT) [11]. The extreme sea state is calculated using a modified inverse first-order reliability method (I-FORM), as described in [12]. The methodologies for calculating environmental contours are an active area research and differences in methodology can be seen when comparing to the results discussed in [9].

Dunkle et al. [9] present peak 16.68 s and 12.49 m as the peak energy period and significant wave height for the 50-year contour, whereas the methodology applied in MHKiT returns a 50-year peak at 15.71 m and 16.24 s. Dunkle et al. [9] present a peak for the 100-year contour at 13.19 m and 16.85 s, whereas the MHKiT functionality returns 16.62 m and 16.43 s. While the exact methodology applied in Dunkle et al. [9] is not discussed in detail, the modified I-FORM methodology applied in MHKiT considers that the principal component analysis does not remove all the joint probability between significant wave height and energy period. The modified I-FORM method applied creates a relationship to describe this dependency in the principal component analysis space.

IV. FUTURE WORK

The team has created a public road map for additional modules and functionality to be developed in MHKiT. This plan is publicly available on our GitHub repository so the marine energy sector can provide feedback (https://github.com/MHKiT-Software/MHKiT-Python/projects/1). In the near term, the MHKiT team will extend existing functionality and create new functions for cloud-based processing of large data sets.

Additional capabilities in the coming years includes functions for mooring assessments, met-ocean calculations, uncertainty, noise characterization, an ocean thermal energy conversion (OTEC) module, acoustic characterization, an environmental impacts module, and functionality focused on Powering the Blue Economy™ applications.

Table 1. Future modules and their capabilities planned for inclusion in MHKiT.

Module	
I/O	• Swan data ingestion • Water Power Technologies Office hindcast data ingestion • ADCP and acoustic Doppler profiler data ingestion ported from DOLfYN • Standardized data formats for upload to the Marine and Hydrokinetic Data Repository (MHKDR) • Data ingestion from the standardized MHKDR data formats
Mooring	• Assessing mooring design and loads • Functions for ingesting and processing data from mooring models as well as assessing strains on mooring lines
Met-Ocean	• Focus on identifying existing packages that could be leveraged by the ME community • Quantities of interest will include wind, water temperature, and water salinity
Uncertainty	• Calculating the uncertainty in measurements and derived quantities following the IEC technical specifications where applicable • Uncertainty in sensor calibrations, instruments, and data acquisition systems
Noise	• Filtering noise out of signals • Functionality for excluding data following the requirements prescribed in each technical specification will be developed and included within each relevant module
OTEC	• Resource characterization • Graphics
Acoustic	• The acoustic characterization module will include: o I/O functions for ingesting acoustic measurement data o Sound pressure calculations o Geo-referencing • Statistics • Graphics
Environmental Impacts	• New module with functionality for assessing the environmental impacts around deploying Marine Energy Converters
Powering the Blue Economy	• New functionality specifically geared toward Powering the Blue Economy applications. Current planned area of focus will be: o Water desalination o Aquaculture o Battery charging o Distributed/integrated applications
General/ Misc.	• Flicker characterization • Data transformation • QC graphics • Additional tidal and wave graphics

CONCLUSION

MHKiT is a functional toolkit for ME-specific data ingestion, processing, and visualization, which is accessible from Python and MATLAB. The tool is open source and adds value to the ME community by allowing for easy importation of popular hosted ME data, functions written based on IEC TS and adapted from ME tools such as the WDRT and provides access to useful graphics functions where appropriate. These tools combined allow for powerful results to be obtained in shorter time frames, which catalyzes and boosts the productivity of the ME community.

MHKiT is still in a nascent stage at slightly over 1-year since its release. At this time, the team is focused on adding more ingestion functions from popular data sources, simulation tools, and instrumentation. The team is looking to add additional IEC TS functionality, and include the functionality of other ME codes where appropriate. In 2021, the MHKiT team will extend existing functionality and create new functions for cloud-based processing of large data sets.

ACKNOWLEDGEMENT

This work was authored [in part] by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Water Power Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

REFERENCES

[1] IEC TC114: Marine energy- wave, tidal and other water current converters, IEC Geneva, Switzerland

[2] "MRE CodeHub" https://github.com/MRE-Code-Hub

[3] Driscoll, F., Fao, R., Klise, K., Ruehl, K., Olson, S., Shippert, T., & Sivaraman, C. (2020). MHKiT-Python (Marine Hydrokinetic Toolkit) (No. MHKiT). National Renewable Energy Laboratory (NREL), Golden, CO (United States); Pacific Northwest National Laboratory (PNNL), Richland, WA (United States); Sandia National Laboratories (Sandia), Albuquerque, NM (United States).

[4] Driscoll, F., Fao, R., Klise, K., Ruehl, K., Olson, S., Shippert, T., ... & Gunawan, B. (2020). MHKit-MATLAB. National Renewable Energy Laboratory (NREL), Golden, CO (United States); Pacific Northwest National Laboratory (PNNL), Richland, WA (United States); Sandia National Laboratories (Sandia), Albuquerque, NM (United States).

[5] "USGS Surface-Water Daily Data for the Nation." U.S. Geological Survey. https://waterdata.usgs.gov/nwis/dv/?referred_module=sw.

[6] "CO-OPS API For Data Retrieval." National Oceanic and Atmospheric Administration. https://api.tidesandcurrents.noaa.gov/api/prod/.

[7] "National Data Buoy Center" National Oceanic and Atmospheric Administration. https://www.ndbc.noaa.gov/.

[8] "Waves and Sea Surface Temperatures." Coastal Data Information Program. https://cdip.ucsd.edu/m/stn_table/.

[9] Dunkle, Gabrielle, et al. "PacWave Wave Resource Assessment." (2020).

[10] Myrhaug, D. (2018). Some probabilistic properties of deep water wave steepness. Oceanologia, 60(2), 187-192.

[11] Coe, Ryan, Michelen, Carlos, Eckert-Gallup, Aubrey, Yu, Yi-Hsiang, and Van Rij, Jennifer. (2016, March 30). WEC Design Response Toolbox v. 1.0 (Version 00) [Computer software]. https://www.osti.gov//servlets/purl/1312743.

[12] Eckert-Gallup, A. C., Sallaberry, C. J., Dallman, A. R., and Neary, V. S. (2016). Application of principal component analysis (PCA) and improved joint probability distributions to the inverse first-order reliability method (I-FORM) for predicting extreme sea states. Ocean Engineering, 112, 307-319.