ERDC/CHL CHETN-III-81
November 2018

US Army Corps of Engineers®

Assessing Jetty Effectiveness via Statistical Analysis of AIS Data

by David L. Young and Brandan M. Scully

PURPOSE: This Coastal and Hydraulics Engineering Technical Note (CHETN) describes a pilot study exploring the use of statistical analysis of Automatic Identification System (AIS) data to assess the sheltering effectiveness of jetties.

BACKGROUND: Coastal navigation structures are frequently provided to improve the controllability of navigating vessels. Structures are typically designed to withstand exposure to wave loads that far exceed safe vessel operating conditions. Even damaged structures may continue to shelter vessels during normal vessel operating conditions. A structure functionality measure provides a mechanism to prioritize structures for maintenance and rehabilitation whereby damaged structures that continue to demonstrate higher functionality receive lower priority.

Globally, certain commercial vessels are required to broadcast specified information using AIS transceivers. The transmitted data are collected by monitoring stations across the United States and archived by the U.S. Coast Guard (USCG). Detailed discussion of AIS purpose, carriage requirements, data content, and availability is provided by Scully and Mitchell (2015). USACE staff may obtain archival AIS data from the USCG (USCG 2016) or from the developing AIS Analysis Package (AISAP) website (Scully and Mitchell 2015). AIS technology documents vessel behavior in high resolution and can be used to quantify vessel-structure interaction. A procedure using AIS data parameters listed in Table 1 was developed to enable statistical comparison of AIS-derived vessel performance metrics for vessels sheltered by jetties compared to those exposed to open ocean conditions.

Table 1. Information from AIS records utilized in this analysis (ITU 2014).

Attribute	Description
Maritime Mobile Service Identity (MMSI)	Unique Vessel ID#
TX_DTTM	Transmission Date-Time stamp
LAT	Latitude coordinate in decimal degrees
LON	Longitude coordinate in decimal degrees
Course Over Ground (COG)	Angle of vessel track in 10 × degrees true (0 – 3600)
Speed Over Ground (SOG)	Speed of the vessel (knots)
Heading (HDG)	Angle of vessel heading in 10 × degrees true (0 – 3600)
Ship and Cargo Type	Number indicating the type of vessel and cargo
DIM_BOW	Distance from the AIS unit to the vessel's bow (m)
DIM_STERN	Distance from the AIS unit to the vessel's stern (m)

STUDY AREA: The Columbia River Entrance at the border of Washington and Oregon was chosen as the pilot site for this investigation due to the large jetties north and south of the inlet, a climate of regularly occurring large waves, and the high volume of commercial vessel traffic through the inlet. Figure 1 displays the study site. The north and south jetties are outlined in red, and the sheltered/exposed demarcation boundary points are marked with a yellow circle. The location of the Columbia River channel, including the channel centerline, is highlighted in green (courtesy of

EarthNC Online). The USCG historical AIS data request system requires specification of the latitude and longitude of the upper-right and lower-left corners of the area of interest (AOI). The white-shaded area represents the area bounded by the upper right) (46° 15' 56" N, 124° 02' 50" W) and lower left (46° 14' 00" N, 124° 06' 00" W) coordinates. Full resolution AIS data in this AOI spanning 1 January 2011 to 31 December 2014 were obtained from the USCG (USCG 2016). Wind, wave, and water level data were acquired from National Oceanic and Atmospheric Administration (NOAA) buoys 46029 and 46243 and NOAA tide gage 9440581 (NOAA 2013)—marked with red triangles in Figure 1 (data products in Table 2).

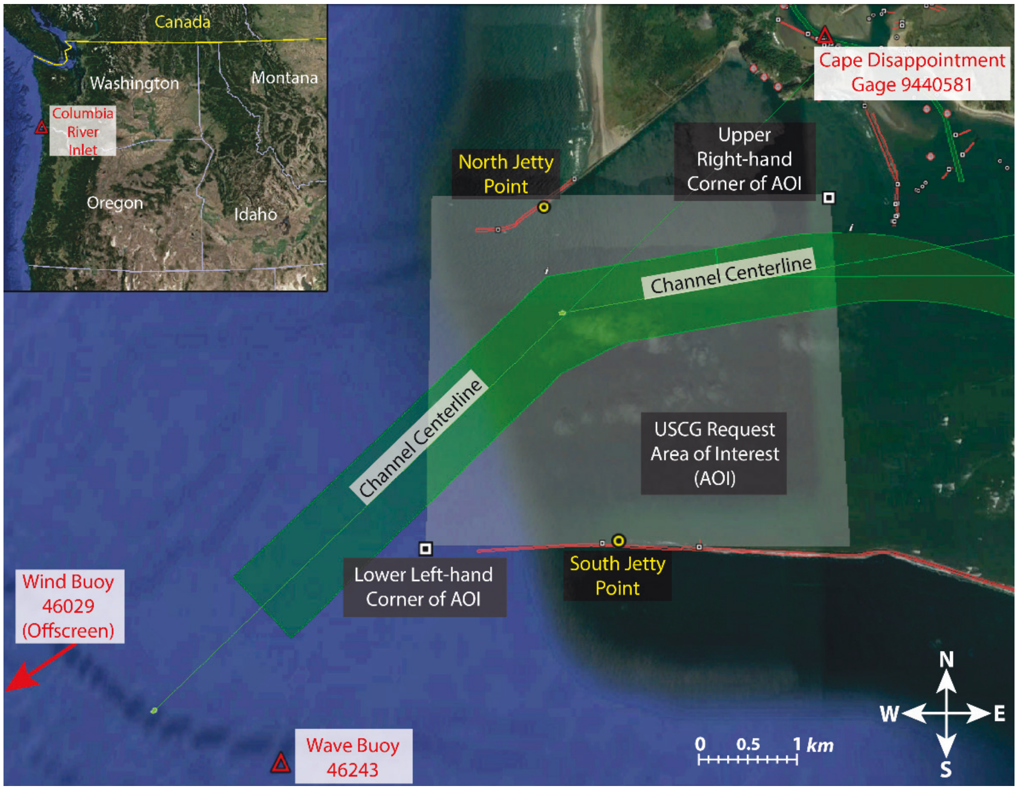

Figure 1. Map of the Columbia River Inlet and relevant features (images courtesy of Google Earth).

AIS Data Pre-processing and Transit Generation. Scully and Mitchell (2015) discuss the potential inaccuracies inherent in AIS data and recommend validation of AIS records with an authoritative source. The identities of vessels in the study area were validated by the USCG Authoritative Vessel Identification System (AVIS). AVIS compares AIS-transmitted information to a database of known vessel information to identify the best possible vessel match (Mitchell and Scully 2014). AIS data were received in monthly comma separated value (.csv) files. These monthly files were combined into a master database to simplify identification of vessel transits that cross months (e.g., a transit from 23:30 on 31 March 2017 to 00:30 on 1 April 2017). The USCG provided data from every monitoring station in the vicinity of the study area, resulting in duplicate vessel transmission records. Duplicate records beyond the first were removed.

This study is principally concerned with the jetty's effect on the performance of large commercial vessels transiting the entrance channel. The ship and cargo type and dimension elements were used to filter for those of interest by retaining only the entries with (1) ship and cargo type in the ranges 70–79 (cargo ships) or 80–89 (tanker ships) or (2) those vessels with unknown type but total vessel length (DIM_BOW + DIM_STERN) greater than 30 meters (m). For each vessel, individual transits are defined as a sequence of position reports where the time between consecutive reports does not exceed 3 minutes (min) (the longest allowable vessel reporting interval—for vessels anchored/moored or moving at less than 3 knots). Only transits with 50 or more vessel position reports are considered.

Transit Filtering. The last step before vessel performance criteria can be computed is to remove/correct points with a clearly erroneous GPS vessel position from the transits. The filtering script first searches each transit for point locations that indicate a large transit distance from the preceding location based on a maximum allowable vessel speed-over-ground (SOG) (35 knots) and the time elapsed between the point locations. Great circle distance is computed between sequential transmission coordinates and is compared to the maximum distance possible for the vessel to travel over the elapsed time based on the maximum allowable speed-over-ground (SOG). If the distance between the points exceeds the maximum possible distance (with a built-in tolerance of 20 m) the point is dropped from the transit. The remaining latitude/longitude (LAT/LON) coordinates of each point in the transit are smoothed individually using a forward/backward low-pass fifth-order Butterworth filter with a cutoff period of 30 seconds (sec) (Choi et al. 2004). To implement the filter, the LAT/LON coordinates of the transit point locations must be interpolated onto a uniform time-step (1 sec). The filter is run on the uniform time-step LAT/LON coordinates, and the filtered coordinates are re-interpolated back onto the time-steps of the original data. The effect of this filter is to remove the high-frequency GPS "hopping" while retaining the trajectory of the vessel.

Post-filtering, the vessel transits are redefined using the original criteria (i.e., blocks of the sorted combined AIS data with identical MMSI number [Table 1] and in which the time between consecutive points does not exceed 3 min), and all of the redefined transits that contain fewer than 50 points are deleted. The remaining redefined transits are again searched for points that are too far away from the preceding points (using the original criteria), and any transits that continue to contain such points after filtering are deleted. In the pilot dataset analysis, the number of transits dropped post-filtering was approximately 10% of the total number of transits prior to filtering.

Metocean Data and Demarcating Jetty Sheltering. Metocean data describing vessel operating conditions, including (1) wave direction, height, and dominant period, (2) wind direction and speed, and (3) water level, were added to the AIS dataset. Data were indexed onto the vessel timestamp from the nearby NOAA buoys and hindcast from the NOAA tide gage (see Figure 1 for locations). Metocean data is listed in Table 2.

Table 2. Description of hydrodynamic/meteorological data added to the vessel transits.

Parameter	Description	Units	Gage/Buoy
Significant Wave Height (H_s)	Significant wave height (mean of the highest 1/3) over the 20 min sampling period	meters	NOAA Wave Buoy 46243
Dominant Wave Period (DPD)	Period of maximum wave energy over the 20 min sampling period	seconds	NOAA Wave Buoy 46243
Mean Wave Direction (MWD)	Direction from which the dominant period waves are incoming (20 min sampling period)	Decimal degrees clockwise from true north	NOAA Wave Buoy 46243
Mean Wind Speed (WSPD)	Mean wind speed 5 m above sea level (8 min sampling period)	meters per second	NOAA Wind Buoy 46029
Mean Wind Direction (WDIR)	Mean wind direction 5 m above sea level (8 min sampling period)	Decimal degrees clockwise from true north	NOAA Wind Buoy 46029
Mean Water Level (MWL)	Mean water level in MSL datum	meters	Hindcast from NOAA Cape Disappointment Tide Gage 9440581

The newly indexed wave direction data are used to determine if individual points in the vessel transits are located within the shadow of the jetty (i.e., sheltered by the jetty). A line with a slope parallel to the wave propagation direction (averaged over the transit) is created to delineate the transition from exposed to sheltered by the jetty. If the waves are incoming from 0° to 90° or 270° to 360° (clockwise relative to true north), then the LAT/LON coordinates for the tip (yellow circle in Figure 1) of the north jetty are converted to Universal Transverse Mercator (UTM) coordinates and used as the point through which the line passes. For waves incoming from 90° to 270°, the coordinates for the visible tip of the south jetty are used. The vessel position coordinates are likewise converted to UTM, and the side of the demarcation line that the vessel position falls on determines whether it is sheltered by the jetty (sheltered) or exposed to ocean waves (exposed). If the point in the vessel transit is to the inland of the line (i.e., in UTM coordinates the x-value of the point is larger than the x-value of the line at the same y position), then the point is flagged as sheltered (see Figure 2). Note that the wind and wave period data were not incorporated into the analysis presented herein. Furthermore, current data were not available to be included in the final dataset, though current direction and velocity are known factors affecting navigability in entrance channels (Scully 2015).

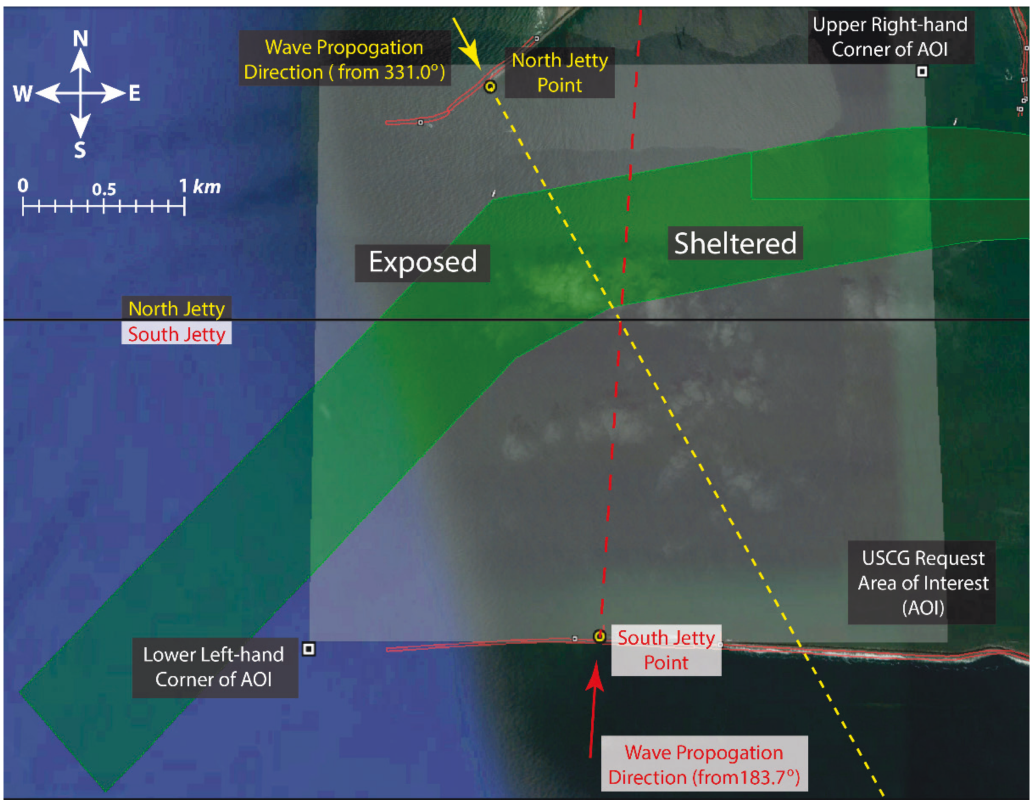

Figure 2. Delineation of exposed and sheltered positions within the channel for waves incoming from 183.7º (red) and 331.0º (yellow) clockwise from true north. (Image courtesy of Google Earth).

Once all points in the vessel transit are flagged as sheltered or exposed, the sheltered portion of the transit is defined as the largest block of consecutive sheltered points within the vessel transit (the exposed portion is similarly defined as the largest block of consecutive exposed points). If the wave direction is undefined during the vessel transit due to gaps in the wave buoy data, or if either the exposed *or* sheltered portions of the transit consist of fewer than 50 points, then the transit is deleted.

Vessel Performance Metrics. The metrics used to quantify vessel performance from the AIS data are principally concerned with two general parameters. The first is the deviation of the vessel's heading (HDG) from its COG—subsequently referred to as "heading deviation-from-course" (HD). In general, this is a measure of the helmsman's ability to maintain a desired course in the presence of external stimuli (Lewis 1989). Larger HD indicates difficulty maintaining the desired course, requiring increased magnitude and/or frequency of steering corrections (with changes to vessel HDG). The performance metrics are computed for individual vessel transits and subsequently used to compare exposed and sheltered vessel behaviors. Metrics that relate to HD are (1) the mean heading-deviation-from-course (M. HD in decimal degrees) and (2) the standard deviation of the heading-deviation-from-course (S. Dev. HD in decimal degrees).

The second parameter used to quantify vessel performance is the channel centerline distance (CCD) defined as the perpendicular distance from a point in the vessel transit to the centerline of the channel. This is a proxy measure of a vessel's ability to maintain a smooth trajectory within the channel bounds (Lewis 1989). The Columbia River channel centerline is shown as a green line in Figure 1. To determine the perpendicular distance (d) from the channel centerline, the LAT/LON coordinates of two points on the channel centerline are converted to UTM. The UTM coordinates are then used to determine the perpendicular distance from the line connecting them to each point in the vessel transit (also converted to UTM).

Figure 1 clearly shows a bend in the channel through the inlet. To address this, the above procedure is performed for the channel centerlines pre- and post-bend, and the perpendicular distance used is the smaller of the distances to the two lines. The performance metrics related to the CCD are (1) the mean channel centerline distance (M. CCD in meters), (2) the standard deviation of the channel centerline distance (S. Dev. CCD in meters), and (3) the root-mean-square error (RMSE) of the channel centerline distance (CCD RMSE in meters). The root-mean-square error is computed by assuming the distance to the channel centerline (CCD) is the *error* between the vessel trajectory and the *correct* trajectory that exactly follows the channel centerline.

Statistical Methods. The first statistical test used to compare performance variables for the exposed and sheltered portions of the vessel transit is the paired sample *t*-test (also referred to as a "repeated measures" *t*-test). This test is appropriate when data are collected in pairs, one pre- and one post-treatment (where the "treatment" is the sheltering of the vessel by the jetty). Some assumptions of the paired sample *t*-test are (1) the metric tested contains numeric values over a continuous range, (2) each observation is independent, (3) the dependent variable (i.e., the performance metric) is normally distributed, and (4) the dependent variable is free of outliers (Rice 2006). The vessel performance metrics likely violate assumption (2) to some extent, as it is probable that the data contain at least a few instances of vessels transiting the inlet simultaneously, as well as instances of the same vessel transiting the inlet multiple times. Additionally, some vessel performance metrics violate assumption (4) to some degree, particularly the standard deviation of the heading-deviation-from-course. In practice, it is often difficult to find a statistical test with assumptions that are all perfectly met by the data; thus, use of "robust" statistical methods are preferred—those which are resistant to errors in the results due to deviations from the assumptions (Huber 1981). The paired sample *t*-test is robust under all but the most egregious violations of the above assumptions (Bland 1995).

The second statistical test used is the two-sample Kolmogorov-Smirnov test, often abbreviated as the KS-test. This test determines if the empirical cumulative distribution functions (CDFs) of two sets of data come from the same parent distribution (Shorack and Wellner 1986). The most attractive feature of the KS-test is that the test's *only* assumption is that the distributions are continuous. Furthermore, the KS-test statistic directly quantifies the separation distance between the two empirical CDFs and consequently is sensitive to the general locations of the variability in the CDFs, as well as the CDFs shapes.

Finally, the Benjamini and Hochberg (1995) post-hoc *p*-value correction method was employed to remove the potential for false significance (Type 1 error) that could result from a large number of statistical tests. The *p*-values (probability that the null hypothesis is true) from each statistical test in the study are ranked from smallest to largest. The ranked *p*-values are adjusted by multiplying them by a factor of N/n, where N is the total number of variables tested, and n is the rank of the *p*-value of the statistical test within the ordered list. These adjusted *p*-values are compared to the statistically significant *p*-value (e.g., $p < 0.05$) to determine whether to reject the null hypothesis that vessel sheltering from the jetty had no effect on vessel HD and CCD metrics.

RESULTS AND DISCUSSION: Table 3 lists the results of the paired sample *t*-test and the two-sample KS-test comparing the performance metrics for exposed and sheltered vessels. The values reported in the Exposed and Sheltered columns of Table 3 are the average of the performance metric across all transits that contain both an exposed and sheltered value (5462 vessel transits (n)). Figure 3 compares empirical CDFs (plotting position $P = m/(n\text{-}1)$ for sample of rank m (Makkonen 2006)) for HD -based metrics, and Figure 4 compares CDFs for CCD-based metrics between exposed and sheltered vessels.

Table 3. Statistical comparison of the exposed and sheltered performance metrics.

	Exposed	Sheltered	Paired *t*-test *p*-value (Adjusted)	KS *p*-value (Adjusted)
M. HD (°)	5.92	4.55	5.2×10^{-59} (1.3×10^{-58})*	1.0×10^{-274} (2.5×10^{-274})*
S. Dev. HD (°)	3.86	4.13	1.7×10^{-2} (1.7×10^{-2})*	1.0×10^{-170} (1.7×10^{-170})*
M. CCD (m)	125.28	128.98	4.4×10^{-4} (5.5×10^{-4})*	6.6×10^{-35} (6.6×10^{-35})*
S. Dev. CCD (m)	65.35	39.27	0.0 (0.0)*	0.0 (0.0)*
CCD RMSE (m)	143.39	137.52	3.6×10^{-8} (6.0×10^{-8})*	1.6×10^{-44} (2.0×10^{-44})*

*significant at $p < 0.05$

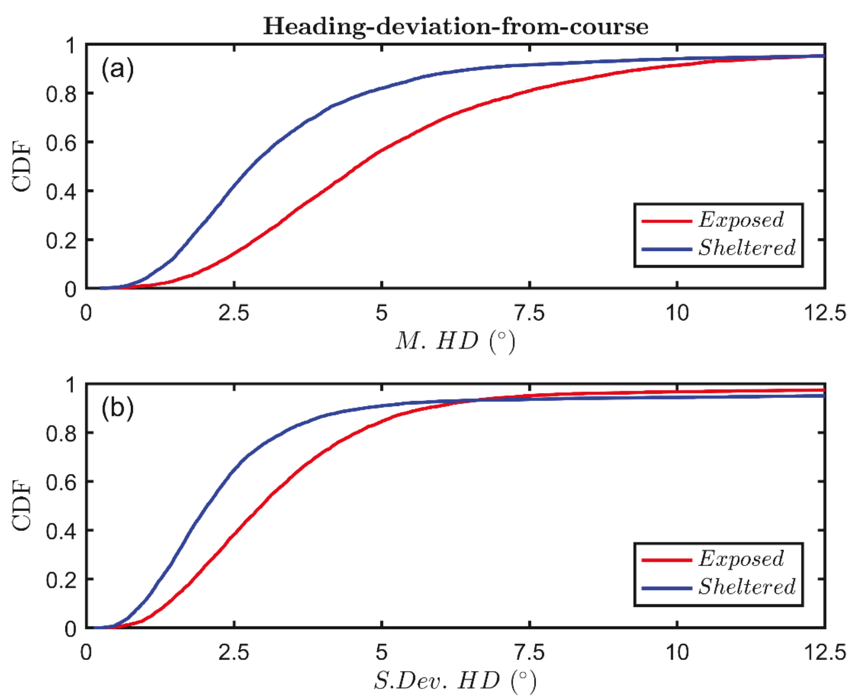

Figure 3. Comparison of the CDFs of sheltered and exposed vessels: (a) M.HD (b) S.Dev.HD.

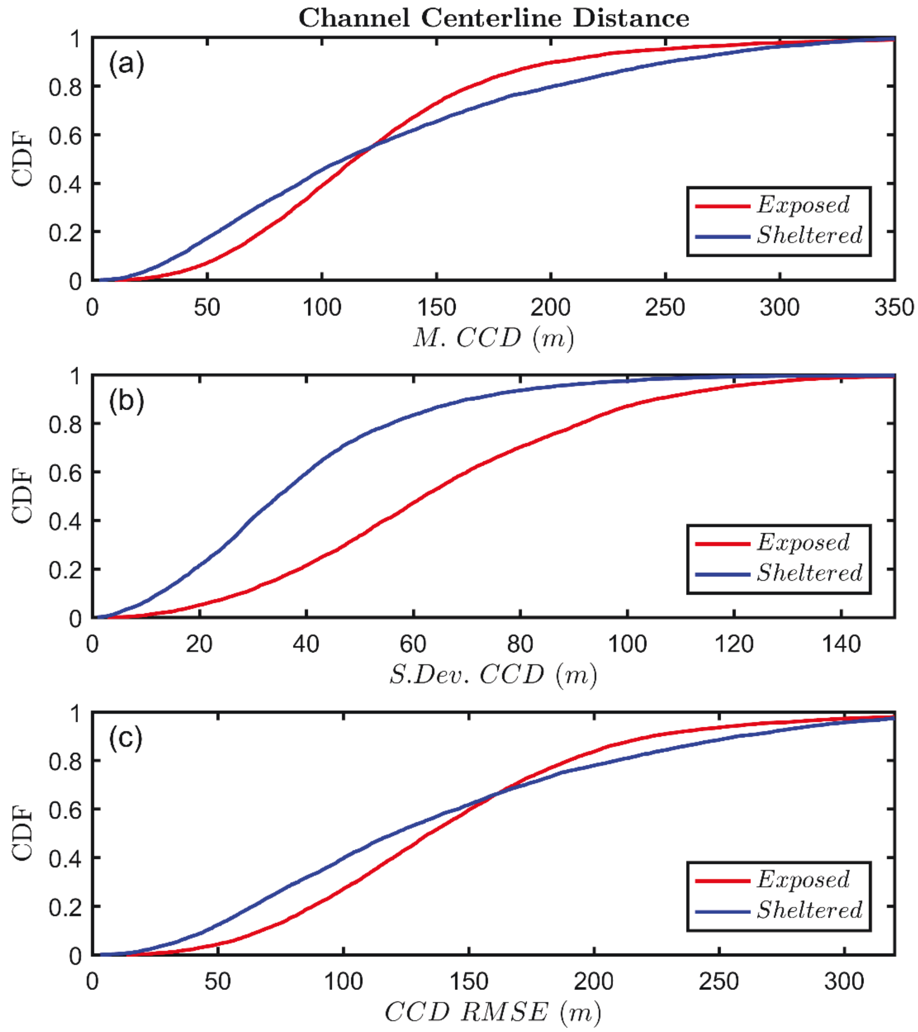

Figure 4. Comparison of the CDFs of sheltered and exposed vessels: (a) M.CCD, (b) S.Dev.CCD, (c) CCD RMSE.

The overwhelming majority of the paired sample *t*-tests and KS-tests were sufficiently significant so as to reject the null hypothesis that sheltering by the Columbia River jetties has no impact on the HD and CCD metrics. A reasonable interpretation of the rejected null hypotheses is that jetty sheltering has a clear impact on the vessel performance as measured by the performance metrics. Inside the jetty, the vessels maintain a smaller deviation between their heading and course (significantly smaller M. HD), indicating that the pilots are making smaller magnitude steering corrections in this region. Furthermore, although the S. Dev. HD is significantly higher in the region sheltered by the jetty, the CDFs in Figure 3(b) suggest that the majority of the vessel transits actually experience a smaller standard deviation inside the jetty than outside, and the mean of the sheltered standard deviation is biased by the comparatively few but large outliers in the sheltered standard deviation data (violating one of the assumptions of the paired sample *t*-test).

The M. CCD for exposed vessels is only slightly smaller than for sheltered vessels (3.7 m less), but the difference is significant. However, the S. Dev. CCD for exposed vessels is significantly and substantially higher than for sheltered vessels. Additionally, the CCD RMSE for exposed vessels is significantly higher than for sheltered vessels. This suggests that vessels outside the jetty are on average closer to the channel centerline, but the vessel broadcast position relative to the channel centerline is substantially more variable. Vessels are better able to maintain a trajectory parallel to the channel centerline inside the jetty, so much so that the sheltered CCD RMSE is significantly smaller than outside the jetty, despite the smaller M. CCD value outside the jetty.

SUMMARY AND FUTURE DEVELOPMENT: The results of this pilot study suggest that statistical analysis of archival AIS data is a promising means of assessing functional jetty performance with respect to vessels in transit. Jetty sheltering was shown to have statistically significant effects on vessel controllability as measured by HD and CCD-based performance metrics at the Columbia River Entrance. This indicates that the sheltering of the jetty decreases the steering input required to maintain vessel course and improves the pilots' ability to maintain a course parallel to the channel centerline. This analysis methodology must be repeated for additional jetty systems and expanded to other coastal structures to fully assess the ability of AIS data to monitor structure effects on vessel performance, particularly given the wide variation in hydrodynamic conditions and structure-channel configurations across the USACE project portfolio. Once verified, quantifiable differences in vessel performance could be used as a prioritization metric for structure management. For example, structures demonstrating higher HD or CCD metrics could be assigned a higher priority for maintenance than those that demonstrate lower metrics.

ADDITIONAL INFORMATION: This Coastal and Hydraulics Engineering Technical Note (CHETN) was prepared as part of the USACE Coastal Inlets Research Program (CIRP) by Dr. David L. Young and Dr. Brandan M. Scully, U.S. Army Engineer Research and Development Center (ERDC), Coastal and Hydraulics Laboratory (CHL), Vicksburg, MS. Questions pertaining to this CHETN may be directed to Dr. Brandan M. Scully (Brandan.M.Scully@usace.army.mil) or to the USACE CIRP Program Manager, Mary A. Cialone (Mary.A.Cialone@usace.army.mil). Additional information regarding CIRP may be obtained from the CIRP web site http://cirp.usace.army.mil/. This technical note should be cited as follows:

> Young, D. L., and B. M. Scully. 2018. *Assessing Jetty Effectiveness via Statistical Analysis of AIS Data.* ERDC/CHL CHETN-III-81. Vicksburg, MS: U.S. Army Engineer Research and Development Center. http://dx.doi.org/10.21079/11681/30030

REFERENCES

Benjamini, Y., and Y. Hochberg. 1995. "Controlling the Fasle Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B* 57(1):289–300.

Bland, M. 1995. *An Introduction to Medical Statistics.* Oxford, UK: Oxford University Press.

Choi, K., A. Bilich, K. M. Larson, and P. Axelrad. 2004. "Modified Sidereal Filtering: Implications for High-Rate GPS Positioning." *Geophysical Research Letters* 31(22): L22608, doi: 10.1029/2004GL021621.

Huber, P. J. 1981. *Robust Statistics.* New York: J. Wiley & Sons.

International Telecommunication Union (ITU). 2014. *Technical Characteristics for an Automatic Identification System Using Time-Division Multiple Access in the VHF Maritime Mobile Band*. ITU-R M.1371.5. Geneva, Switzerland. http://www.e-navigation.nl/sites/default/files/R-REC-M.1371-5-201402-I!!PDF-E_1.pdf.

Lewis, E. V. 1989. *Principles of Naval Architecture Volume III: Motions in Waves and Controllability.* The Society of Naval Architects and Marine Engineers. Jersey City: New Jersey.

Makkonen, L. 2006. "Plotting Position in Extreme Value Analysis." *Journal of Applied Meteorology and Climatology: Notes and Correspondance* 45:334–340.

Mitchell, K. N., and B. M. Scully. 2014. "Waterway Performance Monitoring via Automatic Identification System (AIS) Data." *93rd Annual Meeting of the Transportation Research Board.* January 2014. Washington, DC, United States.

National Oceanic and Atmospheric Administration (NOAA). 2013. NOAA Tides & Currents – Water Levels – Station Selection. http://tidesandcurrents.noaa.gov/stations.html?type=Water+Levels.

Rice, J. A. 2006. *Mathematical Statistics and Data Analysis* (3rd ed). Pacific Grove, CA: Duxbury Press.

Scully, B. M., and K. N. Mitchell. 2015. *Archival Automatic Identification System (AIS) Data for Navigation Project Performance Evaluation.* ERDC/CHL CHETN-IX-40. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Shorack, G. R., and J. A. Wellner. 1986. *Empirical Processes with Applications to Statistics.* Hoboken, NJ: J. Wiley & Sons.

U.S. Coast Guard (USCG). 2016. NAIS Data Sharing Categories and Requirements. http://www.navcen.uscg.gov/?pageName=NAISdisclaimer.