101.02.1609.SSR
Oct 18, 2016

SeaSonde Wave Outlier Mitigation

Background

The SeaSonde wave extraction tool outputs local significant wave height, centroid wave period, and mean wave direction. These are stored in ASCII files prefixed with "WVLM_..." As with all wave sensors, real-time measurements contain obvious outliers that are not indicative of meaningful wave conditions. They arise from the random nature of ocean waves. In addition, for HF radars like SeaSonde, they can come from radio interference and natural radio noise. This is the motivation for developing and offering our outlier removal tool. It operates on the rawest WVLM wave file outputs, using a combination of filtering techniques, and averages over time and radar range, to produce a cleaner representation of wave outputs.

SeaSonde Outlier Removal Tool

A SeaSonde wave outlier removal tool is presented. We outline the filtering methodologies, and the choices the user has in configuring this. Then, by discussion and examples, we show the tradeoffs among them, starting from raw, unfiltered outputs.

Our premise for removing outliers is that there is a very large number of samples within a reasonable temporal and spatial (range) interval. Many of these are reasonable estimates. Some are clearly outliers that are not indicative of wave information. Our aim in outlier removal is to remove these spurious values and average the remainder to get a smoother estimate that connects with consecutive temporal neighbors in a smoother more natural manner.

Our present wave outlier process finds unnatural jumps and/or outliers in both wave height and wave period. When it finds an outlier in wave height, it eliminates it, as well as period and direction values at the same point. Then likewise for wave period, removing corresponding height and direction values. The premise is: if an outlier occurs in any of these parameters, all other values at that time are suspect. Since there are enough remaining non-outlier values, we use those to estimate the three wave parameters.

Details of Filtering Methods

There are three methods available in this tool to remove (or filter) outliers:

a. **Mean** (or simple average): This familiar method simply adds all samples and divides by the total. For outliers, it has the disadvantage that it can be significantly biased by large outliers (which for height and period, are by their nature positive).

b. **Median**: This is the middle value of an ordered array of the sample values (half on one side, half on the other). In itself, it excludes outliers if the field is large enough that outliers are a minority of the samples.

c. **Threshold** based on median: Here the mean value is taken of all samples that fall within ±XX% of the median of the group. We have found ±30% to be effective as a default, but this is user selectable.

Sequencing of Filtering Methods

Points for wave height and period are collected over a 2D space, comprised of time (e.g., every 10 minutes) and distance from the radar (e.g., range rings every 1 km). The normal default is to sequence the filtering so that time comes first, then range (this order can be reversed). Also, one can filter to remove outliers over time for each range cell, creating a file of time-filtered values vs. range. We recommend for operational use to do filtering over both time and range, so that a simple curve wave height and period are archived and displayable for the coverage area around each radar vs. time. After outliers are removed based on wave height and period filtering, these same time points are removed from wave direction; then the average of the sine and cosine of direction give a smoothed direction value. Folders containing ASCII files for these three outlier-filtered estimates (i.e., none, time filtered only, time and range filtered) are produced and archived in near-real time.

Settings in WaveModel.plist

We illustrate the setup preferences table to be used by the operator to set up outlier removal preferences for wave height and wave period . We explain each row entry of the table below.

WaveModel.plist

WaveModel.plist 〉 No Selection

Key	Type	Value
▼Root	Dictionary	(1 item)
▼WaveModelFiltering	Dictionary	(12 items)
MaximumWaveHeightMeters	Number	10
MaximumWaveHeightMetersChangePerHour	Number	1
MaximumWaveHeightMetersDropPerHour	Number	1
MaximumWavePeriodChangePerHour	Number	5
Method	String	Threshold ;; Header Average Median Threshold
MinimumVectors	Number	1
MinimumWaveDopplerPoints	Number	10
OutlierHeightPercentage	Number	80
OutlierPeriodPercentage	Number	120
RangedRateMultiplier	Number	1
ReduceTemporalThenSpatial	Boolean	NO
TwoPassFiltering	Boolean	YES

MaximumWaveHeightMeters: This is the absolute upper limit allowed for valid wave height data. In other words, eliminate any values that exceed this upper limit as an outlier (value shown here is 10 m). If set to 0, this filter is ignored.

MaximumWaveHeightMetersChangePerHour: This builds inertia to the outlier elimination process. Namely, significant wave height cannot realistically increase more than 1 meter per hour, based on the above entry. This is a good default. If set to 0, this filter is ignored.

MaximumWaveHeightMetersDropPerHour: This builds inertia to the outlier elimination process. Namely, significant wave height cannot realistically drop more than 1 meter per hour, based on the above entry. This is a good default. If set to 0, this filter is ignored.

MaximumWavePeriodChangePerHour: Like the previous line, this gives inertia to the wave period (in seconds), which is the centroid of the fitted wave energy spectrum. In the table here, 4 seconds setting as a maximum change per hour is reasonable, based on wave physics. If set to 0, this filter is ignored.

Method: On the preceding page, three methods were given for filtering or smoothing outliers. The user selects here which method is to be applied.

MinimumVectors: This allows the user to select the minimum number of available wave samples allowed for use. For example, suppose over seven time points and three range cells, there were only two values calculated. The user might decide not to try to do an outlier removal, and just leave the output blank (a gap). In this case, any setting less than 2 would do this. Use of 1 ensures that anything greater than zero will be used to estimate the wave parameter. If set to 0, this filter is ignored.

MinimumWaveDopplerPoints: Wave extraction estimates are extracted from second-order Doppler spectral peaks. These are usually a spread continuum of several points. Some have found that by restricting these, outliers can be eliminated. For example, if only three points comprise this second-order continuum visible above the noise, most would feel this will undoubtedly give bad wave results. If set to 0, this filter is ignored.

OutlierHeightPercentage: For the threshold method, this is the threshold percentage above and below which points the median wave height that are to be considered outliers and are excluded from averaging.

OutlierPeriodPercentage: For the threshold method, this is the threshold percentage above and below which points the median wave period that are to be considered outliers and are excluded from averaging.

ReduceTemporalThenSpatial: When this true (check marked), it means temporal outlier filtering will be done first for each range cell. The default setting is false or off.

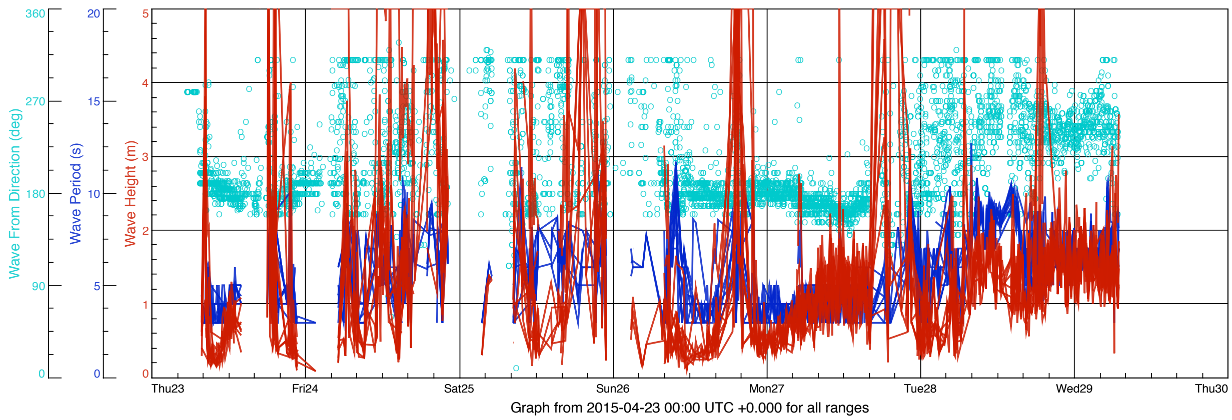
TwoPassFiltering: When true, raw data is run through a filtering first pass and a second pass is then used to filter outliers based on the first pass. This allows the raw data to have more variation averaged out instead of the single pass which will remove highly varied data as outliers and will likely bias the result lower. If the unfiltered data is not highly variable then one pass should be used.

RangeRateMultiplier: When two pass filtering is used, this multiplies the maximum wave height and period change rates in order to let more data through for the first pass.

Examples to Demonstrate Outlier Tool

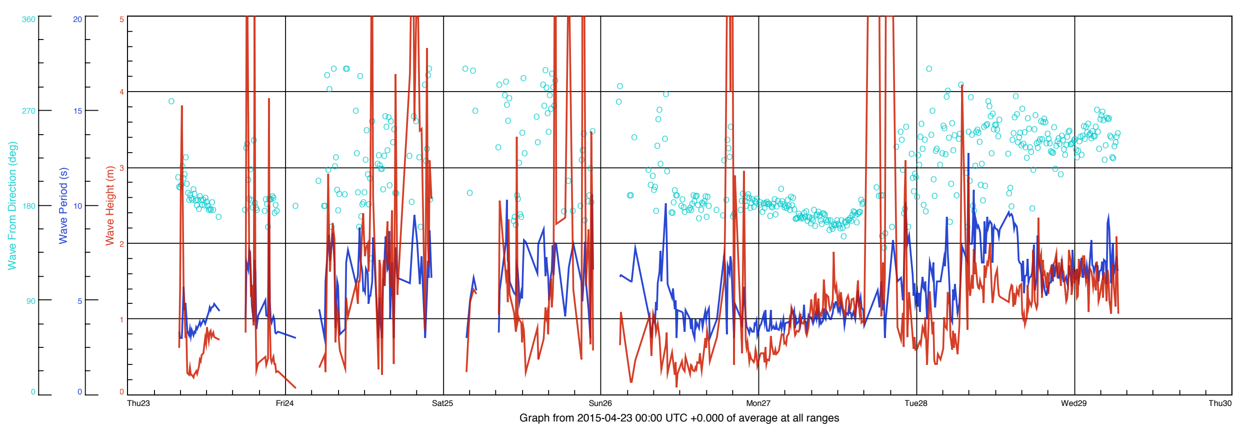
We applied the tool to Livorno, Italy — a 13 MHz site in the Tuscany region of Italy that looks Westward. A single week of data is examined (April 23-30, 2015). The latest wave tool were used to create the WVLM files in the rawest form. We first show below our standard outputs with no outlier removal, and then applied our new outlier removal tool with various preference settings discussed earlier.

A. Raw wave outputs every 10 minutes and showing all range cells



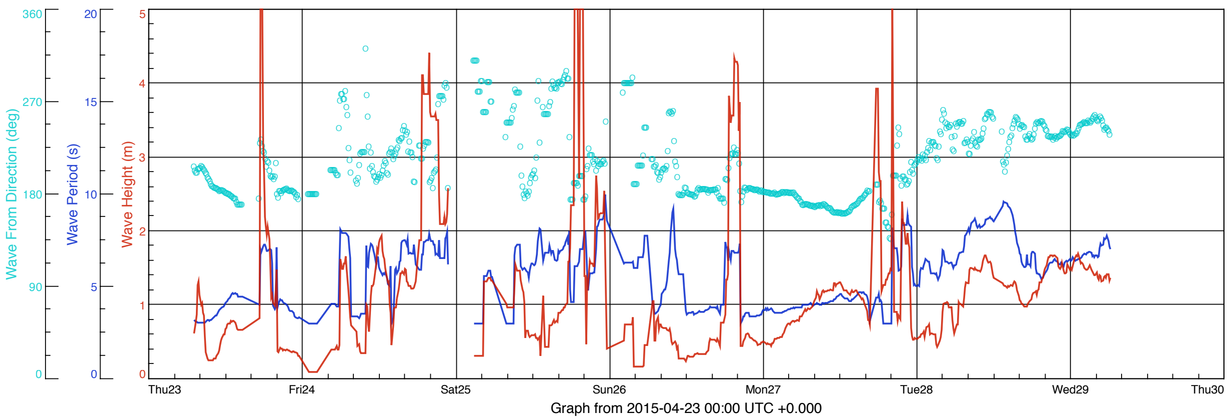
There was a lot of radio interference at this site during this time period. It is responsible for most of the outliers seen, especially in the wave height (red) and wave period (blue). In the use of subsequent outlier removal, this is the raw data (WVLM file contents) that should be used as the input.

B. Raw wave outputs every 10 minutes from standard WaveDisplay with all ranges averaged.

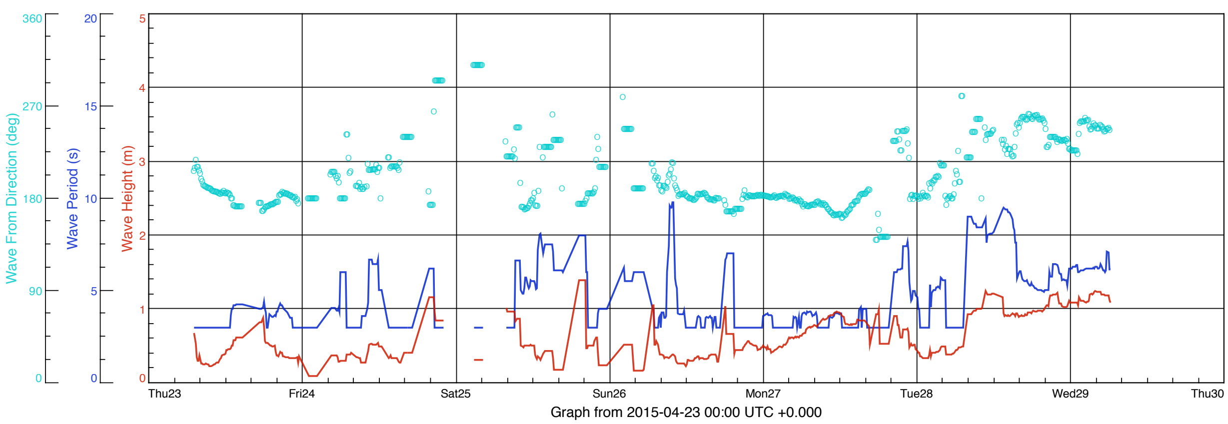


The plot above shows the output of our present WaveDisplay, for data every 10 minutes but with the option set to average all range cells. With the large volume of data decimated, believable features can be seen in the wave parameters, but wave height and period are still contaminated by outliers.

C. Outlier removal tool applied with threshold averaging method only.
The plot above shows the removal of some outliers, but cannot remove instances where most of the points are outliers.



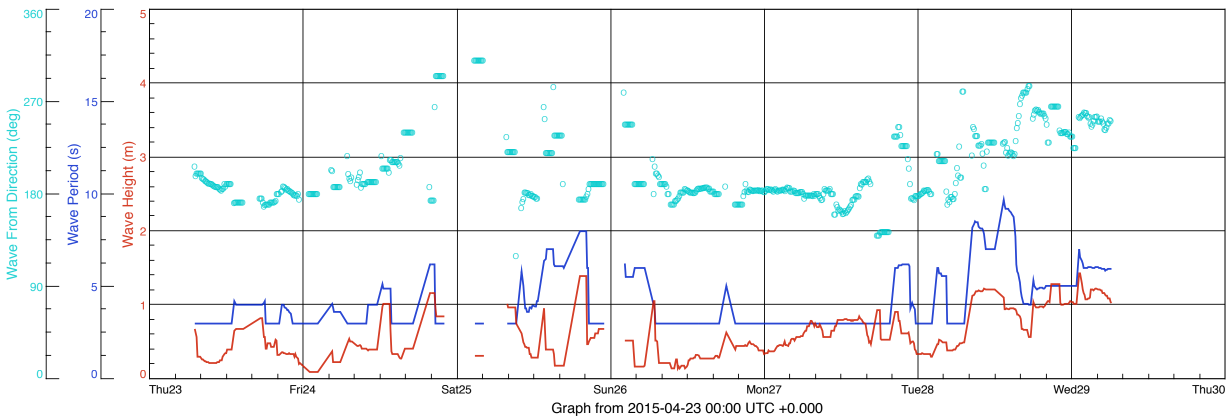
D. Outlier removal tool applied to wave height rate change with threshold averaging.



In the above plot, we have applied outlier removal, where first the time span used for filtering was 95 minutes (meaning the outputs of nine CSS files were selected, i.e., about an hour and a half). A maximum upper limit for wave height of 10 m was set into the preferences table. We allowed a maximum wave height change per hour of 1 meter. All range cells were averaged also. Finally, we used the "Filtering method of threshold based on median", with ±30% as upper/lower threshold limits.

This has greatly improved the quality of wave data — the filtered outputs are observing features that are believable. Outliers have been removed, as well as short-time noisiness that is not endemic to actual average wave conditions. However, note that — although outliers have been removed nicely from wave height — there are still a few noticeable ones in wave period. That will be addressed next.

E. Outlier removal tool applied to wave height and wave period
After experimenting with parameters for wave period, we found it best to use 4 seconds for the change parameter. This means that if period jumps more than 4 s during one hour, the point is considered an outlier and removed. Shorter



times for this cause unnaturally smoothed periods. All other parameters for outlier filtering are the same as the previous example.

Revision History

June 20, 2015 – First Draft.
July 8, 2015 – Final Draft.
Sep 24, 2016 – Minor formatting changes for Release 8
Oct 18, 2016 – Update plist parameter to reflect latest filtering.

Copyright and Disclaimer