# Metadata for Gothenburg 2018 wake dataset

## Description of dataset

### Field measurements and instrument specifications

The data was collected during a field study in the large ship lane outside Gothenburg harbor (57.61178 N, 11.66102 E), using an Acoustic Current Doppler Profiler (ADCP). A bottom-mounted Nortek Signature 500 kHz Acoustic Current Profiler was moored on a buoy anchored to the sea floor at approximately 30 meters depth, for a duration of 4 weeks (28 August to 25 September 2018). The Nortek Signature 500 kHz Acoustic Current Profiler measured along beam current velocity and pressure, using four slanted beams (25° angle, frequency 1 Hz, cell size 1 m), and one vertical beam making high frequency measurements in echosounder mode (burst sampling rate 1 Hz). When recovering the instrument, we realized that beam 2 was malfunctioning, so we only have measurements from 4 beams: 1, 3, 4, and 5.

As the measured values are given as the average for each 1-m bin, the depth for the bind are in 0.5 m intervals. The area closest to the instrument and the area close to the surface do not have reliable data. The ADCP cannot measure all the way up to the surface because of **side lobe interference**. The area closest to the surface (5-10 % of the total distance) is not reliable, and therefore the dataset starts at 3.5 m depth. You also have a **blanking distance** closest to the instrument. This is because the instrument needs some time to get to “total rest” after the signal has been sent out, before it can receive unbiased data. Therefore, the signals that return from the particles very close to the instrument cannot be used, and the area closest to the instrument corresponds to the blanking distance. The blanking distance is the first measured cell + the size of the cells, and this is the distance from where the first measurement will be made (in our case it is 2 m). The max depth in the dataset is 30.5 m.

For the wake analysis I have used the 5th/vertical beam, as it has the strongest signal and the majority of the wakes are most clearly visible in that beam. However, when I identified the wakes, I did look at all available beams, and sometimes the wake is visible in some of the slanted beams and not the vertical beam (the files include plots of all beams so you can have a look by yourselves). Also, there is less noise in the slanted beams, so it might be valuable for you to use all 4 beams. In my Python program it is possible to chose which beam to use for the analysis, but the annotation is made for the 5th beam. But it will be possible for you to plot the data and see if the annotation works for the slanted beams as well (even though the manual adjustments have been made for the 5th beam).

### Periods with strong winds and noisy data

The following periods had strong winds which results in wavy conditions and noisy data. (I have noted if the data is noisy in the Excel sheet with all the wakes).

* 2018-09-08 07:00 – 2018-09-09 14:00
* 2018-09-11 07:00 - 2018-09-13 00:00
* 2018-09-16 12:00 - 2018-09-17 10:00
* 2018-09-18 08:00 - 2018-09-18 19:00
* 2018-09-19 12:00 - 2018-09-20 17:00
* 2018-09-21 00:00 - 2018-09-23 10:00 (last day is really bad!)

## Aim with analysis

The aim of the analysis is to identify the turbulent wakes in the dataset. The wakes are visible in the signal strength data/amplitude data, as the bubbles in the wake reflects the signal much more compared to the surrounding water. The turbulent wake can also be detected using the current velocity data from the 4 beams to calculate an estimate of the turbulence. I want to be able to find the wake area and retrieve the measured data for that area (i.e. the signal strength and turbulence for the wake area).

We have calculated two estimates of turbulence: Turbulent kinetic energy dissipation rate (epsilon ε, or eps for short) and velocity variance (uVar). As the calculations require averaging, the resolution of the data for the turbulence parameters is lower (averaged over 30s) compared to the signal strength (1 Hz, 1 measurement per second). So for each measured point (30s average) there is information about time, depth, signal strength, epsilon and velocity variance (the last two are calculated from the velocity measurements, which we do have in each point).

## Data types

### Matlab files

The “raw” matlab files retrieved from the instrument have been split o reduce the size of the files. I have split them into roughly 12-hour files (day and night), where I have manually checked where the noisy night data starts and ends (the migration of zooplankton). The files are named “day/night\_mm\_dd\_hh\_pythonfiles”. These files are in the folder **Matlab files** (they need to be in the same folder as the python files to run the program). (If you want to see the program used to save these files just let me know and I will send them)

The matlab files named “Passage\_date\_pythonfile\_bubble” are the turbulence files for the annotated wakes. The turbulence calculation of epsilon and uVar have only been made for the time just after ship passage (as there is a disturbance in the calculation caused by the pressure wave from the ship passage). Therefore, each annotated wake has its own turbulence file. The “bubble” part at the end indicates that the datafile starts (in time) when the bubble wake is first visible. The duration of each turbulence file is one hour. These files also need to be in the same folder as the Python program if you want to save the turbulence data as well as the signal strength/amplitude data.

### Python files

The file named “run\_file\_master\_sheet\_wake\_annotation\_only.py” is the run-file. If you run that (and all the matlab, python and csv-files I have sent you are in the same folder as the run-file), you should get a series of plots of the annotated wake areas and a wake list containing ShipWake-objects. Hopefully you can use this list to save the data you are interested in, in a format of your choice. The information saved for each ShipWake object is described in detail in the “ShipWake” python file, but here is a summary of the information I think you will be interested in:

self.bubble\_matrix = None # Matrix over the bubble wake that is used in the PCA analysis.

self.eps\_matrix = None # Matrix over the epsilon wake that is used in the PCA analysis.

self.uVar\_matrix = None # Matrix over the uVar wake that is used in the PCA analysis.

These three ShipWake attributes contains wake matrices for the signal strength/bubble wake, epsilon and uVar. The wake matrix is a square area covering the wake, where the annotated wake area has data (signal strength, epsilon or uVar) and the area outside the wake is padded with zeroes. time is normalised to the time when the ship passes closest to the instrument. The format of the attributes is a list containing the following information: [wake\_matrix, normalised\_time\_vector, depth\_vector].

There are also attributes for the start time and end time for the wake, which might be of interest.

The data saved in the wake-matrix is the raw signal strength/amplitude, epsilon or uVar, but the calculation is made using the normalized, less noisy data (se section Analyses made).

The plotting of the wake area is adjusted in the ShipWake file. There is a section with a TODO-comment (TODO: PLOTTING, line 157), where there are three possible plots that can be commented/uncommented for plotting or suppressing the plots. The file I send you has the second alternative uncommented to produce a plot. If the plot is produced the program should be working! Also, one plot is produced for each wake (89 of them), so there will be a lot of plots.

### Excel files

The ‘Summary metadata info’ Excel file contains a list of all the wakes in the dataset. For each wake I have marked if the wake is annotated or not and written a comment about the quality of the data (noise level) and the size of the wake.

|  |  |
| --- | --- |
| **Noise categories** | **Comment** |
| Excellent | Perfect to use |
| Good | Good to use |
| Ok | Probably ok to use |
| Bad | Noisy, big variation within this category |
|  |  |
| **Wake size categories** |  |
| Very big | Extreme examples |
| Big | Deep, clear and/or long |
| Small | The average wake |
| Tiny | Very small or very weak wakes |
| Ambiguous | Noisy data so difficult to tell if there is a wake |
| Internal | Internal wave, do not include |
| Fish | Wake like signal that probably is fish |
| Overlap | Several wakes overlap |
| Not visible | There is no visible or clear wake |
| Combinations of the above |  |

The reason to why there are wakes marked as non-visible is that there might be a visible wake in one of the other beams, or there is a ship passing close by that could/should create a wake, but there is none visible.

### CSV-files

I have attached a csv-file named ‘test\_master\_sheet\_7’, which is used to load and read the data in the Python program. It contains the filenames for the matlabfiles to use for the analysis, the manual adjustments for each wake (both the turbulent wake and the bubble wake), as well as the ship that has been manually matched with each wake. It can be used to plot the annotated wakes.

As the current Python program need this csv-file to run, I will need to make some adjustments in the program to be able to plot the entire dataset (including the non-annotated wakes). If you would like that, let me know and I can fix it.

### Figures

The figure folder contains the following folders:

**Figures beam a1, a3, a4, and a5**: These folders contain the raw amplitude data for the entire period for each beam. The data is plotted with three subplots in each plot, each containing data for a duration of one hour.

**Wake figures beam 5**: Contains plots of all the annotated wakes (signal strength), including the name of the matched ship. This data is the “filtered data”, no the raw data, so the data is much smoother compared to the figures in the folder above. The wake is centered in the plot area and the wake are is indicated by white arrows (horizontal) and a white line (vertical).

**figures beam 5 corr eps and corr uVar:** Contains plots of all the annotated wakes for the epsilon wave corrected data set and the uVar (velocity variance) dataset respectively. The wake is at the beginning of the plot area and the wake are is indicated by black arrows (horizontal) and a black line (vertical). Note, the axis is logarithmic. To get a good contrast in the figure the log scale does not include the extremes in the dataset, and there are sometimes white areas in the plot where the data is outside the scale (both too low and too high).

I have the same plots for non-wave corrected data if you are interested.

**Wake matrix figures beam 5:** Contains plots of the wake-matrix for each wake. The wake matrix is a square area covering the wake, where the annotated wake area has data (signal strength, epsilon or uVar) and the area outside the wake is set to 0. The wake matrix for the signal strength/bubbles are usually quite nice, but the turbulence matrices are not very good. This is an illustration of the data that is calculated when running the Python run-file (see Python files section).

## Analyses made

For each annotated wake, the wake area had been defined, both for the signal strength dataset (bubble wake), and the TKE dissipation rate (ε) and the velocity variation (uVar). To reduce noise, the data was normalized with respect to distance from the instrument, assuming exponential decay of the signal strength.

The wake area was defined by comparing the wake area to the daily/nightly mean, and all values ~15% higher than the mean was considered part of the wake. As this procedure identified noise as part of the wake, both the percentage limit and the start time, stop time and maximum depth to include in the calculation, were manually defined for each wake to exclude noise (that is what is in the csv-file).

## Reflections/comment

The files I am sending you right now is what I have and can give you without too much extra work. You will probably need different formats of the data, but It is probably easier if you see what I have first, and then I can help you find your way through my messy maze of files and programs.

As I am very new to programing, I have tried to make as many notes as possible in the scripts, but many of the comments might be difficult for anyone except me to understand. So please, ask questions and I will do anything I can to answer them.