CENTER FREQUENCY ESTIMATION FOR NON-GAUSSIAN DATA USING A NON-LINEAR METHOD, THE SYMMIKTOS METHOD

Thomas L. Lagö¹, Sven Olsson²

¹ Jonkoping University, Embedded Systems Lab, Jonkoping, Sweden
² Blekinge Institute of Technology, Telecommunication and Signal Processing Soft Center, Ronneby, Sweden

ABSTRACT

When estimating the Doppler frequency, it is common to use a covariance approach. This is especially true for applications such as underwater current profiling. The early developers of this technology made the assumption that the backscattered signal obeyed Gaussian properties. Based on that, a covariance approach have been used for the estimation of the center frequency in the Doppler signal. Due to a complex Doppler signal, the backscattering process is not always Gaussian, and the covariance method will create bias and wrong estimates.

The Symmiktos Method have been developed to handle the estimation of the Doppler frequency in a more robust sense, and independent of whether or not the backscattering signal obeys Gaussian properties. The method is non-linear, and is performed in frequency domain. The Symmiktos Method also gives less bias; a problem that is inherent in the classical method, and one that is increased if the variance is decreased. Modeled data as well as real-life data is used to show typical behavior of the two methods.

The analysis of the data clearly shows that the covariance method is more sensitive and biased in the presence of signals that do not obey the Gaussian assumption. Also, the large database with backscattering signals from four different locations has been used in the comparison of the two methods. The comparison shows a large difference between the two estimators. All tests on simulated signals have shown that the Symmiktos Method is superior compared to the covariance method for all the data analyzed.

NOMENCLATURE

- $S_x(f)$: Spectrum of the received Doppler signal
- $f_s$: Sampling frequency
- $f_c$: Center frequency
- $f_0$: Transmitted Doppler frequency
- $f$: Frequency
- $x_r(t)$: Doppler signal
- $x_k$: Discrete function
- $\Delta f$: Doppler shift
- $S(f)$: Spectrum
- $Q$: Mean Square Distance to a Gaussian distribution
- $T_i$: Significance levels
- ADCM: Acoustic Doppler Current Meters

1. INTRODUCTION

The time signal response from a ping, generated by the ADCM transceiver, is collected in a block of data. The current profiling is then accomplished by analyzing the data from the bottom up to the surface in blocks of 200 samples each, which is the data collection principle for the ADCM-300 used in this research project, [1]. Each block of data compares to a specific depth, or distance from the transceiver placed at the bottom of the sea.

We assume that the received signal consist of a signal with only one sinusoid, broadened by the burst time. This is the case when there is one target with a steady velocity. The first moment will give a good estimate of the center frequency. The Doppler frequency estimate will be independent of the width of the spectrum. Given the first moment as the estimate of the center frequency, we obtain

$$f_c = \frac{\int f S_x(f) df}{\int S_x(f) df}$$

Figure 1 illustrates how the frequency content will shift upwards with the Doppler shift $\Delta f$.

Figure 1. Example of a transmitted spectrum that is Doppler shifted in frequency due to the target velocity. In this figure it is assumed that the center frequency $f_c$ is very large compared with the Doppler shift $\Delta f$. In this case, the stretching of the spectrum is negligible.
The assumption that there will be one sinusoid would be accurate if this would have been a police radar, that uses the same fundamental Doppler principle as this system. Since the reflections or backscattering is accomplished by many entrained air bubbles, there will be a large amount of individual reflexes, generating the classical assumption that these reflexes are generated by randomly distributed air bubbles, with a size that is small in comparison to the burst wavelength, thus generating a Gaussian signal, with a peak in frequency domain where the main current (Doppler frequency) is located.

Either the time or frequency domain can be used to calculate the center frequency $f_c$, which directly compares to the water velocity, just like the police radar. Each domain has its advantages, but in principle, time and frequency representation contain the same information. It is more a question of how to extract the information desired, and what errors will affect the calculation in what way. It is thus possible to distinguish between two major categories of estimators, each with different hardware implementation possibilities and functionality, time and frequency. Issues such as price, power consumption, accuracy etc. should be considered when deciding. Commonly used estimators in commercially available ADCM systems are:

1. Covariance approach, analog or digital (time domain)
2. Peak-picking approach (frequency domain)
3. First moment (frequency domain)
4. Zero-crossing (time domain)
5. The Symmiktos Method (frequency domain)

The different methods will be very similar, given a "simple" signal, either in the form of a tone, or in the form of a well defined "peak." In reality, the background noise can be substantial, there could be multiple peaks, and there could be disturbance signals from radio equipment, with close frequencies. Such signals have been documented and are presented in [7][8][9]. In these cases it makes a rather substantial difference which approach one uses. In figure 2 below, an illustration is made, explaining what type of difference in terms or result is expected using different estimation approaches.

As indicated by figure 2, there can be substantial differences in between different Doppler responses, with the result that the different center frequency estimation methods finds very different results. The bias due to the background noise is also an issue. The smaller the bandwidth of the measurements system, the less bias due to the random background noise. This, however, will limit the current measurement range. It is possible to one estimation, finding the current (frequency value) with a rather broadband filter, and then narrow the filter, ones the peak has been found. This requires more than one measurement but also assumes that the input data is rather stationary. It is often difficult with the stationarity and it is more important to average correct measurements than to perform multiple measurements for one result. Theoretically, it is a possible approach.

2. NON-GAUSSIAN DATA

The backscattering signal from an acoustic Doppler current measurement system is assumed to obey Gaussian properties in the time domain and the energy from the Doppler signal should be concentrated in a narrow frequency area. This assumption would make the estimation of the Doppler shift, directly proportional to the ocean current, a classical signal processing problem. It is thus of great interest to analyze and verify that the data, does in fact obey these Gaussian properties. Otherwise, rather substantial estimation errors can become the result.

It is common practice in many commercial ADCM systems to use the covariance approach, [1][2]. It has been noted through the use of ADCM systems, that they are not always as accurate as they should have been, [1]. This could be because the Doppler signal was more complex than have been assumed, but there might also be other reasons. NOAA in the USA performed a measurement series and found substantial differences, especially at 2-10 meters depth, when comparing broadband ADCPs with narrowband ADCPs, [3]. The latter is not only assumed to be less accurate, but also more sensitive to the Gaussian assumption. This leads us to believe that the Gaussian assumption could have a larger impact than known before.

![Figure 3. A typical time signal from the Trubaduren light house. The 300 kHz signal is sampled with 48 kHz.](image-url)
This research project has been aimed at analyzing large data sets and prove if the assumption is correct. Different statistical measures have therefore been used, such as 3D spectrograms, normal probability tests, Chi-Square tests, Skewness analysis and Kurtosis analysis, to verify and investigate if the Doppler signal obeys Gaussian properties. In this paper, real data from four locations has been analyzed, using a variety of signal analysis and statistical tools.

In this work, data from four different oceanic locations have been used for the analysis, locations with rather different properties. The locations are:

1. Trubaduren, a light house off the west coast of Sweden. The depth is about 27 meters. The lighthouse stands on an underwater hill, 200-300 m in diameter. This hill can generate vortices and complex current situations. The current is often very high, typically 100 cm/s. Salt water.

2. Fladen, off the west coast of Sweden. The depth is about 14 meters and the bottom is flat and sandy. Currents are high, like at the Trubaduren, typically up to 100 cm/s. Salt water.

3. Almagrundet, off the east coast of Sweden. The depth is about 18 meters and the bottom is stony with algae growth. There are currents in several directions. Brackish water.

4. The Hongkong harbor, where the meter is placed in the Ma-Wan channel. The depth is about 40 meters and the bottom is stony with algae growth. Very high current, up to typically 500 cm/s. The water is quite dirty with a visibility of less than 10-15 cm.

In reference [8], a full analysis of several data sets from locations with different properties, are presented in full. In order to determine if the background noise fulfills the Gaussian assumptions, a number of tests were performed.

The backscattering data consist of the Doppler signal, but also background noise. It is well-known that the low frequency background noise in the ocean is far from Gaussian. However, the noise considered in this case, is centered around 300 kHz, and with a very small bandwidth, about 6 kHz. The first data to be analyzed is the background noise. Figure 4 below represents a typical plot of the background noise in a normal plot diagram.

As can be seen from the diagram, the background noise seems to be Gaussian when viewed with the naked eye. It is, however, important to perform a more explicit evaluation like the Chi square test. The table presents data from the Trubaduren light house. The parameter \( Q \) represents the mean square distance to a Gaussian distribution.

\[ T_1, T_2 \text{ and } T_3 \text{ in table 1 below are the 95\%, 99\% and 99.9\% significance levels. If } T_1 < Q < T_2 \text{, there is a * probability that the data does not belong to a Gaussian distribution, [7].} \]

**Table 1:** Background noise from the Trubaduren lighthouse. \( T_1, T_2 \text{ and } T_3 \text{ in the table below are the 95\%, 99\% and 99.9\% significance levels. } Q \text{ is the mean square distance to a Gaussian distribution.} \)

<table>
<thead>
<tr>
<th>( Q )</th>
<th>( T_1 )</th>
<th>( T_2 )</th>
<th>( T_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.71</td>
<td>22.36</td>
<td>27.69</td>
<td>34.53</td>
</tr>
<tr>
<td>15.67</td>
<td>21.03</td>
<td>26.22</td>
<td>32.91</td>
</tr>
<tr>
<td>14.83</td>
<td>21.03</td>
<td>26.22</td>
<td>32.91</td>
</tr>
<tr>
<td>11.36</td>
<td>19.68</td>
<td>24.72</td>
<td>31.26</td>
</tr>
<tr>
<td>9.08</td>
<td>19.68</td>
<td>24.72</td>
<td>31.26</td>
</tr>
<tr>
<td>14.67</td>
<td>22.36</td>
<td>27.69</td>
<td>34.53</td>
</tr>
<tr>
<td>29.50*</td>
<td>22.36</td>
<td>27.69</td>
<td>34.53</td>
</tr>
<tr>
<td>13.60</td>
<td>22.36</td>
<td>27.69</td>
<td>34.53</td>
</tr>
<tr>
<td>8.25</td>
<td>22.36</td>
<td>27.69</td>
<td>34.53</td>
</tr>
<tr>
<td>4.48</td>
<td>21.03</td>
<td>26.22</td>
<td>32.91</td>
</tr>
</tbody>
</table>

After evaluating the table, it is evident that a Gaussian assumption for the background noise model is realistic since \( Q < T_1 < T_2 < T_3 \).

It is interesting to determine if also the received signal fulfills the requirements for a Gaussian distribution. A backscattered signal is, however, somewhat more complex. The received signal is caused by many small entrained air bubbles in the water. If the bubbles have the same diameter, and there are many bubbles, they will have the same movement in an average sense, and the response will obey Gaussian properties. The real total response received is, however, likely to be a mixture of several distributions with different means and variances.

![Figure 4](image.png) Figure 4. Normal cumulative probability plot of background noise data collected at the Trubaduren lighthouse. The Gaussian assumption seems to be valid.

![Figure 5](image.png) Figure 5. A typical time signal from the Trubaduren light house. The 300 kHz signal is sampled with 48 kHz.
The next set of data is collected at 22 meters depth at the Trubaduren lighthouse off the Swedish west coast. Figure 6 presents a normal cumulative probability plot of the Doppler shift data at this depth, suggesting the presence of several processes acting simultaneously.

A Chi-square test on the data was performed. The tests quantify the distance to a Gaussian distribution, as previously mentioned. A calculation based on the same data is presented in Table 2 below.

<table>
<thead>
<tr>
<th>Q</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
</tr>
</thead>
<tbody>
<tr>
<td>187.05</td>
<td>21.06</td>
<td>29.14</td>
<td>36.12</td>
</tr>
<tr>
<td>83.41</td>
<td>22.31</td>
<td>30.58</td>
<td>37.70</td>
</tr>
<tr>
<td>127.67</td>
<td>22.31</td>
<td>30.58</td>
<td>37.70</td>
</tr>
<tr>
<td>109.77</td>
<td>22.31</td>
<td>30.58</td>
<td>37.70</td>
</tr>
<tr>
<td>228.44</td>
<td>21.06</td>
<td>29.14</td>
<td>36.12</td>
</tr>
<tr>
<td>267.59</td>
<td>21.06</td>
<td>29.14</td>
<td>36.12</td>
</tr>
<tr>
<td>128.84</td>
<td>22.31</td>
<td>30.58</td>
<td>37.70</td>
</tr>
<tr>
<td>89.19</td>
<td>22.31</td>
<td>30.58</td>
<td>37.70</td>
</tr>
<tr>
<td>294.37</td>
<td>21.06</td>
<td>29.14</td>
<td>36.12</td>
</tr>
<tr>
<td>239.82</td>
<td>21.06</td>
<td>29.14</td>
<td>36.12</td>
</tr>
</tbody>
</table>

It is a surprisingly large distance to a Gaussian distribution as reported by the Chi-square test. The data constitute a strong indication that the Gaussian assumption is not valid. When studying the data there also seems to be a difference with depth. This compares well to the findings by NOAA in [3].

Consequently, the next set of data was collected closer to the surface. For larger amplitudes, there is a larger mismatch. Consequently, the tails in the distribution plot are curved. It is quite obvious from the analysis of these data that there is a large variation in the individual pings. These tests have been performed using a large amount of data, and no evidence for a Gaussian distribution has been found.

After analyzing a large set of data, it is clear that given the data used so far, the Gaussian assumption is not valid for the backscattered Doppler signal. For the background noise it is, however, a good model.

3. THE SYMMIKTOS METHOD®

When designing the Symmiktos estimator, a thorough investigation of different estimators was made. Advantages and disadvantages of the estimators were noted, and work was initiated on a new estimator which did not have the disadvantages of the covariance and peak-picking estimators. In order to handle the complex Doppler signal and the background noise, a nonlinear approach was taken. The Symmiktos Method is an estimator that implements the first moment calculation after linear and nonlinear signal processing has been applied, suitable for the Doppler signals generated by ADCMs for current measurements.

The Symmiktos Method consists of the following six steps:

1. Estimate the noise power spectrum.
2. Estimate the signal power spectrum (see figure 7).
3. Make a new estimate of the noise power spectrum, and compare this noise level with the noise level received before the measurement (step 1).
4. If this noise level is close to the first noise level, subtract the estimated noise power spectrum level from the signal spectrum estimate (see figure 7).
5. Insert a clip level, calculated from the peak in the Doppler spectrum.
6. Estimate the Doppler frequency using an accurate first moment calculation algorithm on the peak above the clip level (see figure 7).
4. RESULTS

The result is an estimate with a very low bias, and low variance. The non-linear operation, subtracting the noise level, helps to suppress interfering signals, such as disturbing transmitter signals and background noise. The non-linear operation, clipping, helps avoid a bias due to background noise and also prevents calculation of the first moment, including several low level signals, which could bias the estimate, [5]. The covariance estimator and the Symmiktos Method have been compared for several data sets. In order to illustrate the differences, however, simple simulated signals have been used. These signals have then, successively, been made more complex. The sinusoidal signals used for this evaluation represented a 0.9 m/s current, 0.9 m/s current plus white noise with a signal-to-noise-ratio of 30 dB, a sinusoid representing a 0.9 m/s current, white noise (as before) plus some sinusoids with random amplitude and random frequency. This signal is closer to the "real world" than the sinusoid with white noise added and a "real world signal" from the Trubaduren lighthouse off the Swedish coast.

In the first test the Covariance method and the Symmiktos method both show a negligible variance. The Covariance method has a small bias error. For a known current of 0.9 m/s, the Symmiktos Method estimated the value correctly, while the Covariance estimator estimated 0.808 m/s, a 9% error. Using the second signal with white noise added to the sinusoid, the Symmiktos Method is still accurate, and with negligible bias. The variance in the Covariance estimator is also negligible, but the bias changes with the amount of noise added to the signal, as already expected. In the third test, white noise and several sinusoids with random frequency and amplitude around the fundamental frequency were added. The Symmiktos method starts to change its estimate. The Covariance estimator has a large variance, and the accumulated current converges slowly, as illustrated in figure 8.

Another simulation showing the sensitivity of the covariance method is illustrated in figure 9 below. Here, a simulation with a 0.9 m/s current has been performed. However, a number of large disturbance signals have been introduced. The covariance method now performs catastrophically. It is a very mean test, but still, clearly illustrates the sensitivity to this type of signal with multiple tonal components.

A more realistic current signal was simulated, and the estimators were tested again. Figure 10 illustrates the result of this test. The curve with the larger variance represents the individual estimates. The other curve represents the successively averaged estimates.

![Figure 8](image1.png)

**Figure 8.** Comparison of the estimators' performance given a sinusoid, a sinusoid plus white noise, and added "random" sinusoids. The true current value is 0.9 m/s in all tests.

![Figure 9](image2.png)

**Figure 9.** The Symmiktos Method, and the Covariance estimator tested on synthesized signals. A main current of 0.9 m/s has been used and a number of disturbance components representing current values close to 0.9 m/s have been introduced. The covariance method is highly sensitive to this type of signal.

![Figure 10](image3.png)

**Figure 10.** Comparison of the estimators' performance given a sinusoid, plus white noise. The true current value is 0.9 m/s.

This simulated signal with a number of "random" sinusoids is not a particularly representative signal of the current, since the sinusoids represent a large current spread. The signal is, however, interesting from the robustness point of view. The covariance method is highly sensitive to "close" in frequency components.
When estimating the Doppler frequency, directly related to the water current, for Acoustic Doppler Current Meters (ADCMs), the covariance method is a common tool. The estimator approach is based on the assumption that the backscattered data obeys Gaussian properties. It has been found that ADCMs based on the covariance method, sometimes generates large errors [1][3], and in this paper suggestions for these errors are presented. An analysis of a large data set, investigating if the data is Gaussian or not, has been presented. The conclusion is that the background noise can be considered as Gaussian, but the received Doppler signal from the backscattering process cannot be considered as Gaussian, in general.

The analysis of large amounts of Doppler data has led to a new estimation model for the backscattered Doppler shifted signals. The frequency and time domain estimators of the first moment are similar if we look at the simplified mathematical model. However, we need to approximate the derivative in the moment are similar if we look at the simplified mathematical model. However, we need to approximate the derivative in the moment calculation, a robust and accurate estimation method is achieved.

Several examples have been presented, illustrating the differences in between the covariance method and the Symmiktos Method. For less complex signal, the main difference is the bias term in the covariance estimator. For more complex data, the covariance method start to generate more and more errors, apart from the bias term. For signals containing multiple frequency components, it is very difficult for the covariance method to converge at all, whilst the Symmiktos Method still finds the Doppler frequency.

There are applications for this method, or approach, outside of the ADCM and oceanographic society. It is not uncommon in the mechanical engineering society to be interested in a frequency value, and thus there could be other applications for the Symmiktos Method approach than the ones described in this paper.

ACKNOWLEDGMENTS

We wish to thank SMHI in Norrköping (The Swedish Meteorological and Hydrological Institute) for all their support in this research project, as well as the Blekinge Research foundation for financial support.

REFERENCES