Clusterflux v1.0: Clustering solutions for flux estimates based on seabed gas seep observations in single beam echo-sounder data

Knut Ola Dølven

April 22, 2025

1 Introduction

Accurate and reliable quantification of underwater gas seepage is of great importance for environmental research as well as for monitoring various sub-sea human activities. Our knowledge about the sensitivity and extent of natural seabed seepage is still very limited and even though underwater oil and gas exploitation shall decrease in the near future, existing infrastructure as well as new industries considering e.g. seabed carbon capture projects need extensive monitoring methods for leak detection. The most effective method for detecting seabed seeps is by using ship-mounted single- and multibeam echosounders. While multibeam echosounders have the advantage of wide coverage, only Single Beam Echosounder (SBE) data can be currently used for gas flux estimates using inversion tools such as FlareHunter [Veloso et al., 2015] and VBA-lab (Virtual Bubble Analysis laboratory, plugin ESP3). These tools make it possible to extract seabed gas fluxes from seeps by relating the backscatter signal to gas volume in the water column and are crucial in both scientific and industrial applications. Single-beam echosounder systems are also relatively cost effective and of limited payload which might make these a viable option for autonomous vehicles.

In the case where flare observations with sufficiently overlapping footprints are obtained, i.e. as in the top-down view in Figure 1b, clustering is applied with the aim to counteract double counting and maximize the probability of a best estimate of the total flow rate. Without clustering, overlapping areas and their associated gas flow (assuming the gas flow is evenly distributed in the flare observation footprint) will be counted twice.

Until now, there has been no standard software used for flare clustering. Additionally, even though it prevents double counting, there are several challenges with the conventional way of doing clustering (mostly described in Veloso et al., 2015). To revisit and establish standard software and methodology on this topic is therefore of interest to improve the validity of seabed gas flux estimates and consistency between researchers when using FlareHunter/VBA-lab.

Here, I provide software that does flare clustering and calculates cluster flux with the option to use both the original solutions presented in Veloso et al., (2015) and newly developed solutions that I believe improves the validity of the end result. The scripts also allows for customization of the clustering parameters. The scripts also include a simple GUI option and can be used with minimal knowledge of python and no programming experience. The underlying theory of both the original clustering solutions and the newly developed solutions are presented herein, as well as justifications for why the new solutions gives a more realistic result. The original and new

methodology has also been tested, showing that the new methodology presented herein can have significant implications for both the total number of clustered seeps and total estimated seabed flux.

Everything outlined is available at `https://github.com/KnutOlaD/flare_clustering`.

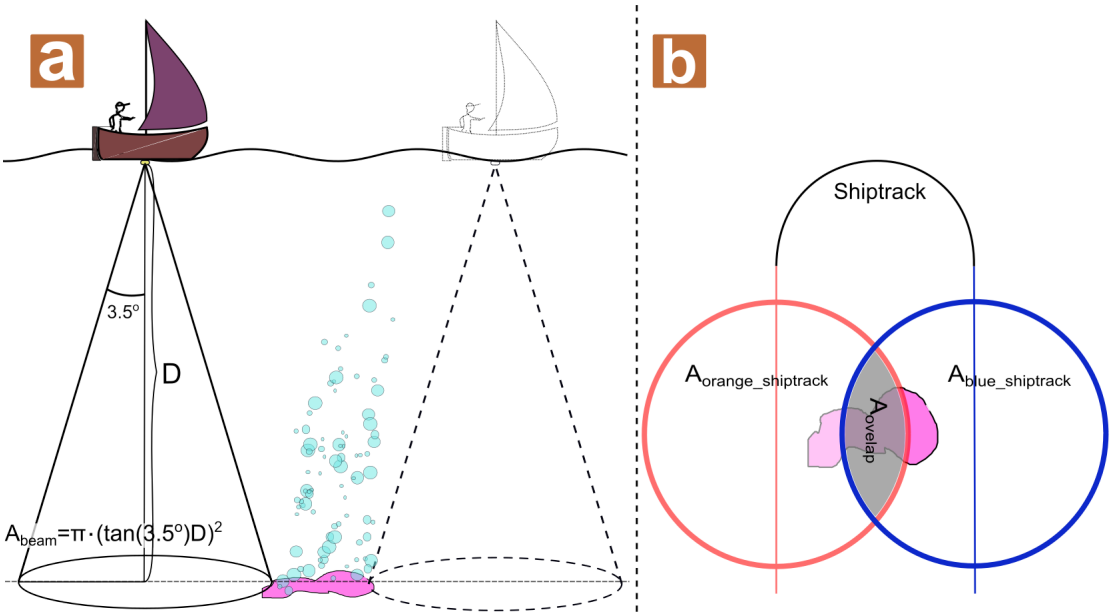

Figure 1: Conceptualized figure showing a) Insonified volume during an echosounder survey and observation of seabed seepage here as a constrained area of seepage and b) top down view of the acoustic footprint illustrating double counting of the same seepage area and an overestimate of the total flow rate during an echosounder survey.

2 Flow rate estimates using single-beam echosounder data

The information used for flow rate estimates in Veloso et al. (2015) is the acoustic "target strength" obtained as the logarithm of the summed backscatter cross-section of scatterers in the insonified volume. This data is obtained for each ping and at all depths as given by the vertical resolution Δd of the echosounder. Thus, the insonofied volume here refers to a truncated cone with thickness Δd (see illustration in Veloso et al., 2015). The method assumes that the total backscattering cross-section (TS) can be directly linked with the amount of gas present in the water column. By inferring a gas rising speed, the seabed flux is derived.

A typical singlebeam echosounder insonofies a cone-shaped volume of water where the horizontal acoustic footprint A_{beam} at a given depth D is given by

$$A_{beam} = \pi \left(D \tan(\frac{\theta}{2}) \right)^2 \tag{1}$$

where θ is the acoustic beam opening angle of the echosounder. For the single-beam echosounder, which is often used for seep detection and flow rate calculations, the opening angle of the cone is 7^o. The resulting horizontal acoustic footprint of the echosounder can therefore be quite

large at typical depths of interest, e.g. at 220 meter the acoustic footprint area is $A_{beam} = \pi \left(220m \tan(\frac{7}{2})\right)^2 = 569m^2$ (see Figure 1a)

The location of the single scatterers within the acoustic footprint of the echosounder is considered to be arbitrary, thus implying that the TS can be produced by several bubble streams with unknown locations within the footprint. In other words, it is assumed that the seepage is evenly distributed within the footprint when flow rates are estimated.

2.1 Glossary for observing seeps with single beam echo-sounders

There are a lot of terms that are easily mixed up within this topic. I therefore provide a short glossary here to avoid any confusion in the present text. This glossary also includes the abbreviations for the different solutions for calculating cluster flux and seep area - the details of these will be described in the text.

Flare	Acoustic signature in echo-sounder data that originates at the seabed resembling a gas flare
Flare observation	An isolated observation of a flare in an echogram
Gas seep	Point or area where gas comes out of the seabed or pipeline etc.
Gas seep observation	Observation of a (true) gas seep, i.e. here a flare observation that is actually a seep
Average Everything flux (AEflux)	The original method of calculating cluster flux as described in Veloso et al., (2015)
Gridded Average flux (GAflux)	The new proposed method to calculate cluster fluxes
Acoustic Footprint Estimate (AFE)	The original method of estimating seep area
Seep Area Estimate (SAE)	The new proposed method to estimate seep area

The main point here is that a "flare observation" is referring to the flare shape in the echogram, which has different characteristics than the gas seep that we believe are resulting in the echogram flare shape. For instance is a bubble stream V-shaped due to the horizontal dispersion of bubbles in the water column, while the acoustic footprint has more of an upward pointing triangular. It should also be noted that a flare observation might also not be a gas seep observation.

3 Flare observation clustering

Clustering is done to avoid double counting of the same seep in post-processing of SBE data. In practice, flares that are clustered are treated and counted as seep clusters instead of individual flares and assigned a single flow rate value.

3.1 Clustering threshold methods

There are two options for choosing clustering thresholds in clusterflux (with associated threshold parameter choices) - center difference (half-radius) and fractional overlapping area with associated parameter choices.

3.1.1 Center distance as cluster threshold

An integral part of clustering is determining the metric by which two flare observations should be clustered. In Veloso et al. (2015), all overlapping flares are clustered. While this can be achieved by gridding and searching for overlapping grid cells, it has typically been done using the distance between the centers of flare observations (personal communication Stetzler et al., 2023). In practice, flare observations A and B are clustered if their separation distance Δ_{AB} is $\leq \Gamma_{dist} \frac{R_A+R_B}{2}$, where Γ_{dist} is a threshold limit parameter (for instance 1.8). However, determining thresholds this way can have severe limitations in certain situations. For instance, when individual flare observation areas have very different sizes (e.g., when $R_1 \ll R_2$), it is possible that flares with no overlap at all are clustered. This can, of course, be circumvented by gridding the field and using a simple "if" statement that demands the flares overlap. However, this approach restricts the boundaries for how thresholds are determined (e.g., if the user wants all clustered flares to have a certain amount of overlap). In Clusterflux, an additional cluster threshold option is therefore included, which addresses these issues.

3.1.2 Overlapping area as cluster threshold

It is possible to determine a threshold based on overlapping flare observation areas instead of center distances. To do this, (still) assuming circular seep areas, following Weisstein (2023), we can calculate the overlapping area between two flare observations as

$$\Omega_{A\cap B} = R_A^2 \cos^{-1}\left(\frac{\Delta_{AB}^2 + R_A^2 - R_B^2}{2\Delta_{AB}R_A}\right) + R_B^2 \cos^{-1}\left(\frac{\Delta_{AB}^2 + R_B^2 - R_A^2}{2\Delta_{AB}R_B}\right) - 0.5\sqrt{(-\Delta_{AB}+R_A+R_B)(\Delta_{AB}-R_A+R_B)(\Delta_{AB}-R_A+R_B)(\Delta_{AB}+R_A+R_B)} \quad (2)$$

where $\Omega_{A\cap B}$ represents the overlapping area between the flare observation areas. The fractional overlapping area is then found simply by dividing $\Omega_{A\cap B}$ by the individual areas Ω_A, Ω_B of circle A and B. Clustering threshold can then be given as a fractional value ($\Gamma_{area} \in \{0, 1\}$) where the flare observations are clustered if either or both of the statements $\Omega_{A\cap B}\Omega_A \leq \Gamma_{area}$ and $\frac{\Omega_{A\cap B}}{\Omega_B} \leq \Gamma_{area}$ are true. This implies that, depending on the relative sizes of A and B, one flare observation might be determined to be clustered into the other but not vice versa. However, this scenario is not problematic. As long as one flare observation is deemed "cluster-worthy," the flare observations are clustered (with the smaller flare observation likely representing just a fraction of the larger seep).

3.2 Estimating flux from clusters

Flare clusters are then treated as single flares and the flux is calculated based on the individual flow rates of the individual flare observations that comprise the cluster. This has traditionally been done using the Average Everything (AEFlux) solution. Here, we have developed a new solution for estimating cluster flux, the Clustered Average (GAFlux) solution, which overcomes some of the drawbacks with this method. Clusterflux includes the option to run both these methodologies.

3.2.1 Original solution: Average Everything flux

This method is included as the *cluster_flowrate_average_everything* function in the clustering.py script in the github repository (`https://github.com/KnutOlaD/flare_clustering`)

The clustering technique suggested by Veloso et al. (2015) assigns an average flow rate to the total area of flares that are close enough to each other to be considered a seep cluster. In Veloso et al. (2015), any flare observation with overlapping acoustic footprint at the seabed, defined by a threshold value, is clustered. The threshold is typically defined as a center distance of 1.8 times the radius of the averaged acoustic footprint (of two adjacent flare observations). This applies to all flare observations, but it is sufficient that only one flare observation has enough overlap in a cluster, implying that a flare cluster can stretch over large areas and contain hundreds of flare observations that have no overlapping area (see Figure 13a in Veloso et al., 2015 and Figure 2 in this document). Once clustered, Veloso et al. (2015) estimate the flow rate of the cluster by first calculating the average gas flux per unit area F_{avg} in the cluster as

$$F_{avg} = \frac{1}{K} \sum_{i=1}^{K} \frac{F_i}{A_i}, \tag{3}$$

where F_i and A_i are the flow rate and seabed footprint of flare observation i in the cluster, respectively. The total cluster flow rate F_{total} is then calculated by multiplying the average flow rate per unit area by the total area of the estimated cluster using a gridded numerical solution

$$F_{total} = F_{avg} A_{cluster} = F_{avg} N \Delta x \Delta y, \tag{4}$$

where N is the number of grid cells with cell size $\Delta x \cdot \Delta y$ in the cluster area $A_{cluster}$.

Even though this solution provides good averaging in many cases, e.g. when only two flare observations are clustered, there are situations where considering averaging errors can arise using this methodology. Let's for instance consider the four seep observations presented in Figure 2 with estimated flowrates of $F_1 = F_2 = F_3 = 5$ and $F_4 = 205$ and the same seabed footprint of $A_1 = A_2 = A_3 = A_4 = 400$. The seep observations were obtained in a straight line with 80% overlap between A_1 and A_2, 80% between A_2 and A_3, and 20% overlap between A_3 and A_4 and were therefore considered a single seep cluster instead of individual seeps (we use areal overlap here for simplicity in the calculations and not radial as is the current norm for defining the threshold for clustering). This means that their individual flow rates were not considered and a common cluster flow rate was calculated (see Figure 2). Following the methodology of Veloso et al., (2015) as iterated above (the AEFlux solution), the total flow rate of the cluster is then calculated as

$$F_{total} = \frac{1}{4} \left(\frac{F_1}{A_1} + \frac{F_2}{A_2} + \frac{F_3}{A_3} + \frac{F_4}{A_4} \right) \cdot A_{cluster} = \frac{1}{4} \left(\frac{5}{400} + \frac{5}{400} + \frac{5}{400} + \frac{205}{400} \right) \cdot 880 = 121, \tag{5}$$

where the cluster area $A_{cluster}$ is calculated as $A_{cluster} = A_1 + 0.8A_1 + 0.2A_1 + 0.2A_1 = 880$ since $A_1 = A_2 = A_3 = A_4$ and we already knew the amount of overlapping area (see Figure 2 a).

In this case, AEFlux clustering estimates an average of the observations and multiplies it with the total shared area. In this example, it is quite obvious that the total flow rate of the cluster is underestimated. After all, considering that it is only 20% of the area of flare observation 4 that overlaps, the remaining 80% area of this observation alone should amount to a flow rate of $0.8 \cdot 205 = 164$. I believe that the issue here is caused by violating prior assumptions that are required to make the AEFlux solution valid. These are

1. All flare observations in the cluster must have the same amount of overlap (otherwise there needs to be some weighing added to account for this in the averaging operation).

2. Deviations from the cluster mean flow rate are due to random processes and/or that the central limit theorem applies.

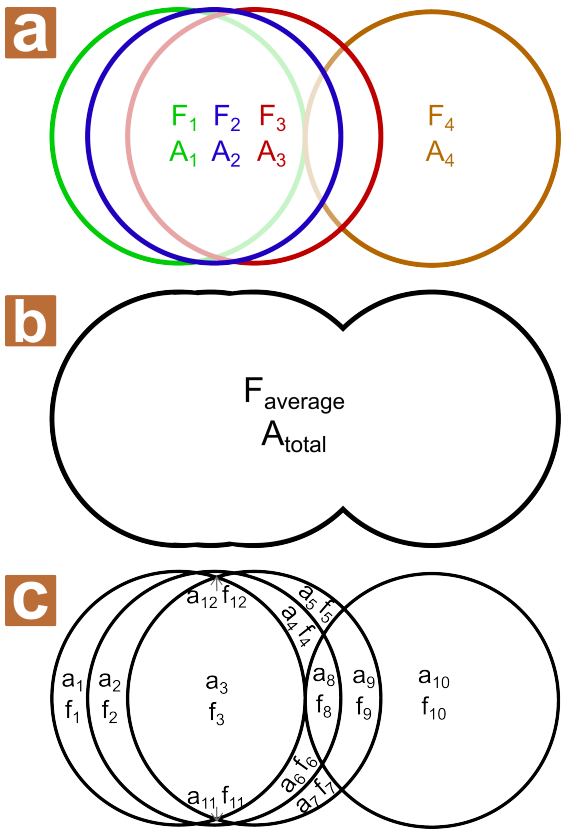

Figure 2: Idealized example of four clustered overlapping flare observations on a line with a) the four different observations and accociated areas and flowrates, b) how the AEFlux solution implements averaging in the cluater and c) how the Gridded averaging method implements averaging in the cluster. See text and equations for how the total flux is calculated.

Both these assumptions are violated in our example and will also be violated in practically all field observations unless there are only two flares in the cluster. The second assumption is also often violated in practice. A typical sufficiently large sample size for a random process for the central limit theorem to apply is N$\sim$30 whereas a typical flare cluster is considerably smaller in size.

3.2.2 New clustering solution: Gridded averaging

This method is included as the *cluster_flowrate_gridded_averaging* function in the clustering.py script in the github repository (`https://github.com/KnutOlaD/flare_clustering`).

Considering a cluster containing clustered flare observations $k \in [1..K]$, we first define a grid covering the total cluster area with grid cells $[i, j] \in [1..N, 1..M]$, resolution $\Delta x = \Delta y = \zeta$ and grid cell center locations (relative to the geographic zero reference) $[i\zeta, j\zeta]$. Given that each flare observation footprint area has outer bounds defined by center location $[k_x, k_y]$ and radius R_k, we can obtain the vector $\mathbf{I}_k$ containing the $[i, j]$ index pairs of all grid cells located within the footprint area by including all $[i, j]$ pairs where both $k_x - R_k < i\zeta < k_x + R_k$ and

$k_y - R_k < j\zeta < k_y + R_k$.

Furthermore, we can approximate the total area A_k of flare observation k by

$$\hat{A}_k = \zeta^2 \sum_{i=1}^{N} \sum_{j=1}^{M} \delta_{i,j}, \quad \text{where} \quad \delta_{i,j} = \begin{cases} 1 & \text{when} \quad [i,j] \in \mathbf{I}_k \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

and, assuming uniform seepage for each individual flare observation, we can calculate the flow rate from each grid cell, i.e. the gridded observed gas flux per unit area,

$$\phi_k = \frac{F_k}{\hat{A}_k} \tag{7}$$

where F_k is the estimated flow rate from flare observation using FlareHunter/VBA-lab.

We can now proceed to extend this formulation for a whole flare cluster. Let's assume a flare cluster consisting of K flare observations with indexed $[1, 2, ..., K]$. Furthermore, any flare observation in the cluster is given a flux per grid cell area ϕ_k and an associated delta function $\delta^k_{i,j}$ as defined above. The averaged flux from a single cell in the cluster can then be expressed by

$$\Phi_{i,j} = \frac{\sum_{k=1}^{K} \phi_k \delta^k_{i,j}}{\sum_{k=1}^{K} \delta^k_{i,j}} \tag{8}$$

and by simply summing over all grid cells in the cluster area we get

$$\Phi_{tot} = \sum_{i=1}^{N} \sum_{j=1}^{M} \Phi_{i,j} = \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{\sum_{k=1}^{K} \phi_k \delta^k_{i,j}}{\sum_{k=1}^{K} \delta^k_{i,j}}, \tag{9}$$

which gives the total flux of cluster K.

Essentially, we use a numerical grid to calculate the individually averaged flowrates of all the different configurations of existing areal overlaps. This way we can cluster flares without letting flowrates observed at one location impact our estimates at completely different locations. In other words: Where we have overlapping samples, i.e. *only where the footprints are overlapping*, we calculate the average flow rate of the overlapping samples. Otherwise we let the individual estimates remain unchanged.

We can illustrate this in a conceptual manner, disregarding errors associated by the nummerical gridding, by the areas $a_1....a_{12}$ and associated flowrates $f_1....f_{12}$ in Figure 2c.

For instance, area a_1 has no overlap with other observations and the flowrate f_1 is calculated by

$$f_1 = \frac{F_1 a_1}{A_1}, \tag{10}$$

thereby only taking the flowrate observed in flare observation 1 into account. Flowrate of area a_8, on the other hand, is calculated as

$$f_8 = \frac{1}{3}\left(\frac{F_2}{A_2} + \frac{F_3}{A_3} + \frac{F_4}{A_4}\right) a_8, \tag{11}$$

here the expression $\frac{1}{3}\left(\frac{F_2}{A_2} + \frac{F_3}{A_3} + \frac{F_4}{A_4}\right)$ gives the average flowrate per unit area of area a_8 specifically - taking all, but only the observations overlapping in area a_8 into account. The total flowrate is, conceptually, disregarding the numerical errors, in this example calculated as $\sum_{k=1}^{12} f_k \sim 192$.

Referring to Figure 1b, we average only in the grey area, instead of the combined area of the two footprints as in the AEFlux solution.

The only underlying assumption of this approach (which is also an assumption of the AFlux approach) is that *seepage is uniform within the area of each individual flare observation* unless we include information from overlapping observations. In other words GAFlux will always give a similar or more realistic result compared to AFlux and it adviced to use this GAFlux whenever possible.

In principle, one could argue that all overlapping areas should have an average flux estimate. If so, this formulation could be applied to the whole seep area and the whole clustering step (i.e. finding clusters) could be omitted. In essence, one would treat the whole surveyed seepage region as a single cluster. This would probably give the best total flow rate estimate, but would not give any information on the total number of seeps. However, the latter is of course something that can be obtained by other means. This is also easily achieved in the software by setting the clustering distance to a very large number which would effectively include all flare observation in a single cluster.

4 Seep area estimation using flare observation area

As discussed in the previous section, in AEFlux clustering it is assumed that the amount of gas coming from a seep is evenly distributed in the whole flare observation area, A_{fo} (see Figure 3). By "flare observation area", I here refer to the area reported in FlareHunter/VBA lab output data. A_{fo} is derived from the footprint of the acoustic beam of the SBE at the seabed and horizontal width (in the echograms) of the base of observed flares. Essentially, it is the accumulated coverage of the acoustic footprint of the echosounder that returns gas indicative data in the echo-profile for a particular flare observation. This is the area used when creating seep clusters, but does not correspond to the actual footprint area of the monitored seep. In fact, in the output data, only the radius is given, assuming a circular seep area. This is also slightly misleading, since the accumulated footprint is never a circle but typically an elongated shape (see Figure 3). Here, we aim to use a geometrical and statistical perspective to estimate the *gas seep area* A_{seep} as opposed to the *flare observation area* A_{fo} and describe the relation between them.

An example of the relation between the flare observation area A_{fo}, seepage area A_{seep}, and echosounder acoustic footprint area A_{beam} is shown in Figure 3a, where the echosounder is moving in a straight line along the x-axis of our coordinate system over a completely flat seabed. The flare observation area obtained when observing an arbitrary gas seep area in this situation, assuming the seep is located *at* the shiptrack, i.e. at $y = 0$, can be calculated as

$$A_{fo} = \pi R^2 + 2LR, \tag{12}$$

where $R = \tan(\theta)D$ is the radius of the acoustic footprint at the seabed given by the opening angle of the echosounder θ (here $\theta = 7^o$) and depth, D. L is the distance between the origin of the seep when it was first detected and where it leaves the acoustic footprint. This is the area reported in FlareHunter and ESP-3 output data and it will be referred to as the Acoustic Footprint Estimate (AFE). In order to proceed with finding an estimate of the seep area A_{seep}, two assumptions are made:

1. At least half of the seep area is captured by the echosounder beam
2. The seep area is circular

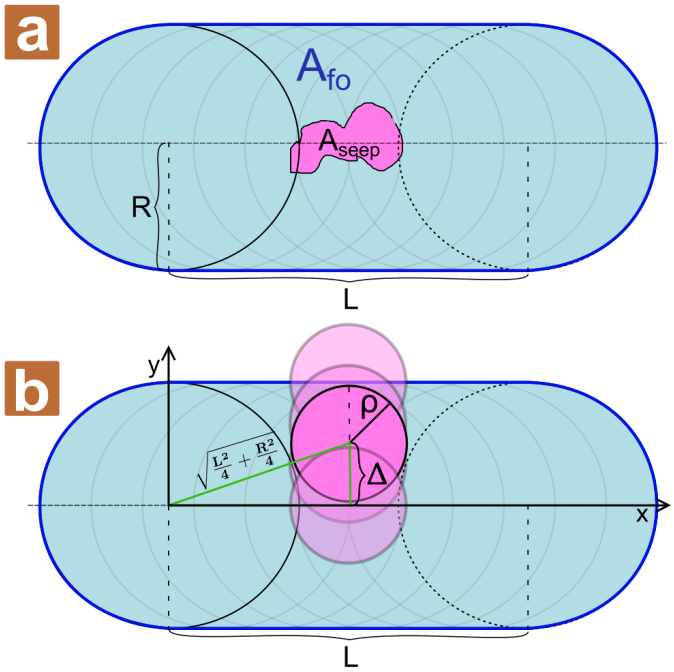

Figure 3: Concept drawing of how a) the flare observation area A_{fo} in light blue, of an arbitrary seep with area A_{seep} in pink is obtained using the accumulated coverage of the SBE acoustic footprint A_{beam} (the black circle with radius R) and used in clustering following flow rate estimates using FlareHunter and/or ESP-3 software and b) how an arbitrary seep area $\hat{A}_{seep}$ can be estimated assuming it is placed at an arbitrary orthogonal position to the ship track for the same flare observation area as presented in figure a)

While both these assumptions might be untrue, we argue that they hold relatively well on average, given that a seep located mostly outside of the acoustic footprint of the echosounder will likely result in correspondingly lower flowrate estimates and might not be interpreted as a gas seep at all given the relatively sharp opening angle of the echosounder. Assuming a circular shape follows directly the fact that the average shape of a large set of random shapes is a circle. The estimated seepage area $\hat{A}_{seep}$, assuming it is placed with a center-distance $y = \Delta$ away from the ship track line (see Figure 3) is then given by

$$\hat{A}_{seep} = \pi \left(\sqrt{\frac{L^2}{4} + \Delta^2} - R \right)^2, \tag{13}$$

where, using Equation 12, we can express L in terms of A_{fo}

$$L = \frac{A_{fo}}{2R} - \frac{\pi R}{2}. \tag{14}$$

We now only need to make an assumption about the y-location of the seep area center, Δ, to arrive at an area estimate.

Using the current output from VBA-lab, there is no way of determining the location of the seep along the y-axis (although it could be possible to derive this information for split beam

SBE's in the future). We therefore seek the expected value for Δ (Δ being the distance between the $x = 0$ line and the center of the seep). First, we assume that the origin for any detected seep is located within the interval $y = \pm R$ (even though it's theoretically possible to detect seeps outside of this range as well). This implies that when $L \geq 2R$ (when the gap between the big circles in Figure 3 is ≥ 0), the center of a detected seep can be placed anywhere between $y = R$ and $y = -R$. By trivial inference, the expected value for Δ then becomes $\frac{R}{2}$ (i.e. $y = \pm\frac{R}{2}$. When the circles start to overlap ($L < 2R$), the potential location of the seep becomes limited to the distance between the intersection point of the circles and $y = R$ (see Figure 3). The intersection point between the two circles can be found using the standard equation for the circles at $x = \frac{L}{2}$, i.e.,

$$\begin{aligned} \left(\tfrac{L}{2} + 0\right)^2 + y_c^2 &= R^2 \\ \left(\tfrac{L}{2} - L\right)^2 + y_c^2 &= R^2, \end{aligned} \tag{15}$$

where $y_c = \pm\sqrt{R^2 - \left(\frac{L}{2}\right)^2}$ are the intersection points. In other words, the seep must in this case be located between y_c and $y = R$ and the expected value for Δ becomes $\frac{1}{2}(y_c + R)$. We can omit L (since this is not explicit output data) using Eq. 14 and some algebra to get $L \geq 2R \implies (4+\pi)R^2 \geq A_{fo}$. Combining this with Eq. 14 and 13 we obtain the following expression for the estimated seep area (with only known variables)

$$\hat{A}_{seep} = \pi\left(\sqrt{\tfrac{1}{16}\left(\tfrac{A_{fo}}{R} - \pi R\right)^2 + \Delta} - R\right)^2, \quad \Delta = \begin{cases} \frac{1}{2}(y_c + R) & \text{when} \quad (4+\pi)R^2 < A_{fo} \\ \frac{R}{2} & \text{when} \quad (4+\pi)R^2 \geq A_{fo} \end{cases}. \tag{16}$$

It should be noted that it is also possible to calculate the theoretical maximum Δ using the recorded flare height in the FlareHunter/VBA-lab output data by assuming no horizontal bubble displacement in the water, but this is potential future work.

This way of estimating seepage area, hereby referred to as the Seep Area Estimate (SAE) method, is not aimed at giving the exact area of a particular seep, but to provide an estimate of the expected value. It is still an improved estimate for any seep over using the AFE method, since AFE would consistently overestimate the area. Another advantage of using SAE is that it effectively avoids the flawed assumption of a circular accumulated acoustic footprint, and replaces it with a more reasonable assumption of a circular seep area. The SAE method should therefore be the preferred method to use when the seep area is of interest, such as when clustering seeps.

The area_estimator.py script automatically estimates the seep area using the SAE method. The input is standard output VBA-lab/FlareHunter.xlsx sheets (see test_data.xlsx) and the output is a new.xlsx sheet where the area radius column ('Average_Radius_Foot') is replaced by a SAE estimate and given a new column name ('Seep_Radius_Est').

5 Testing

We tested the different solutions on a data-set collected offshore Vesterålen, Norway (see Ferré et al., 2024). Flow rates were estimated using the ESP-3 / VBA-Lab software and flow rate output and accumulated acoustic footprint area output data (the AFE solution) were used for testing. We clustered the flare observations and their associated flowrate estimates using a distance threshold of 1.99 radii (see [Veloso et al., 2015]) and four different configurations: 1) IFA/AFlux, 2) IFA/GAFlux 3) SAE/AFlux 4) SAE/GAFlux.

Method configuration	Number of clusters	Clustered observations	Total flowrate [ml/min]
AFE and AEFlux	16	83	795
AFE and GAFlux	16	83	778
SAE and AEFlux	10	34	895
SAE and GAFlux	10	34	904

Table 1: Results of using different configurations, i.e. Instrument Footprint Area (AFE), Estimated Seep Area (SAE), the Average Flux (AFlux) solution and Gridded Average Flux (GAFlux) solution.

The largest impact on total flowrate estimates appears when changing the seep area (see Table 1). When SAE, the number of clustered observations decrease from 83 to 34 and the total flowrate estimate for the whole area increase with around 15%, from 778 to 895 (AEFlux) and 778 to 904 (GAFlux). This increase in flowrate is a result of fewer clustered flares, meaning that more of the measurements are fully taken into account, rather than being averaged with nearby flare observations. There are only minor differences between the averaging methods, however, this might be because many flare observations are not clustered at all (over 50% of the flares are not clustered in this particular dataset) or clustered with only one other observation. In both these cases the two methods are effectively identical. When looking at differences in flowrates for individual flare clusters there are clear differences, particularly for larger clusters (see Figure).

Even though the two methods presented here can provide more realistic flow rate estimates, there are several elements that have not been taken into account. One regards the assumption that the seep is placed at an arbitrary location between the ship track and the acoustic footprint border (that $y = 0 \leq \Delta \leq y = R$). The outer bound for Δ is also actually limited by the observed flare height, since this is limited by the vertical boundary of the cone shaped acoustic footprint of the echosounder. Additionally, when estimating the true seep area, we assume that the seep is located directly at the center of the accumulated acoustic footprint area. This is also not a good assumption, since the seep area in reality can be located anywhere between the boundaries set for the Δ-value. This can actually increase the number of flare observations that should be clustered and reduce the difference between the results. By using a 2-dimensional probability distribution for the placement of the seep area or a Monte Carlo simulation, this effect can probably relatively easily be simulated to give even more valid and realistic clustering and hence flow rate estimates for the seep site. Additionally, split beam echosounders are becoming industry standard which unveils an even more promising potential future solution. In this case, seep location and/or seep area can be derived from the angle information in split-beam echosounder data.

References

[Stetzler et al., 2023] Stetzler, M., Veloso, M., and Moser, M. (2023). Personal communication. September 1 - December 31, 2023.

[Veloso et al., 2015] Veloso, M., Greinert, J., Mienert, J., and De Batist, M. (2015). A new methodology for quantifying bubble flow rates in deep water using splitbeam echosounders: Examples from the Arctic offshore NW-Svalbard. *Limnology and Oceanography: Methods*, 13(6):267–287.

[Weisstein, 2023] Weisstein, E. (2023). Circle-circle intersection. `https://mathworld.wolfram.com/Circle-CircleIntersection.html`. Accessed: 2023-09-10.