Uncertainties associated with the use of erosional cave scallop lengths to calculate stream discharges

Gregory S. Springer and Andrew Hall
Department of Geological Sciences, Ohio University, Athens, OH 45701, USA

Abstract: Scallops are extremely valuable indicators of past water flows in caves because they often record events that cannot be safely witnessed nor measured. Qualitatively, the inverse relationship between their lengths and formative water velocities is useful for determining how flow changes along a cave passage, but they are most valuable because they can be used to directly estimate actual water velocities and discharges. We explore the effects of sample size, measurement choices, and other methods commonly applied to the use of cave scallops in estimating cave stream velocities and discharges. We measured 100 scallops on a cave wall and find them to be log-normally distributed. We used Monte Carlo simulations to sub-sample the 100 scallops for sample sizes of 10 to 30. As expected, smaller sample sizes yield widely varying means with precision increasing slowly with sample size. A sample size of 30 results in greater than 50% of simulated means falling within one standard deviation of the mean for all 100 scallops. This is also true of sample sizes as small as 20, so we recommend a minimum of 20 to 30 scallop measurements in the field. The formulas we use to estimate water velocities and discharges explicitly use the Sauter mean of scallop lengths, but some authors use the arithmetic mean. We simulated the use of both the Sauter and arithmetic means and find that the latter yields substantially larger velocities and discharges. We recommend use of the Sauter mean because that is consistent with the original formulations and the arithmetic mean may cause significant overestimation of velocity and discharge.

Keywords: cave, scallops, karst hydrology, stream, water

Received 25 October 2019; Revised 30 December 2019; Accepted 9 January 2020

Citation: Springer G.S. and Hall A., 2020. Uncertainties associated with the use of erosional cave scallop lengths to calculate stream discharges. International Journal of Speleology, 49 (1), 27-34. Tampa, FL (USA) ISSN 0392-6672 https://doi.org/10.5038/1827-806X.49.1.2292

INTRODUCTION

The largest flows or floods often determine or shape stream channel properties (Wolman & Miller, 1960) and large floods are typically the most effective at dissolving and abrading the walls of caves (Gale, 1984; Palmer, 2007). But the largest flows are often short-lived and difficult to measure. Fortunately, caves often sport scallops, which are asymmetrical depressions that record flow velocities. These unique features have been extensively used to understand cave hydrology using scallop lengths measured in the field (Curl, 1974; Palmer, 1976; Lauritzen et al., 1985; Springer & Wohl, 2002; Palmer, 2007; Woodward & Sasowsky, 2009; Despain et al., 2016). However, field-based data always come with uncertainties, but the magnitude and sources of those uncertainties are not always clear or known.

Sampling methods give rise to uncertainties that can be quantified using various statistical measures, including the standard error. The latter, also called the margin of error, can be minimized by taking large samples, but how large is large enough? And there is always the question of sampling biases, which include questions about how individual objects were chosen for measurement and whether the probability of selecting an individual object was the same as the probability of choosing adjacent objects (Davis, 2002). Also, measures of many geological phenomena are not normally distributed, which can limit the uses or statistical analyses of such data, so the shape of the parent population is of interest (Limpert et al., 2001). We explore some of these sources of uncertainty when using cave scallops to estimate floodwater velocities and discharges (Curl, 1974).

Scallops are intricately tiled depressions on solid substrates created by fluid-mediated erosion (Figs. 1 and 2). Scallops are erosional phenomena created by corrosion, ablation, and abrasion. Individual scallops form where a sublaminar jet detaches from a surface, destabilizes into turbulence at some distance downstream, and reattaches to the surface.
the continuity equation, whereby the discharge

discharges of Lauritzen (1982) were consistent with
and hydraulic modeling. The scallop-derived
1985) and by Jeannin (2001) using spring discharges
using scallop-derived discharges (Lauritzen et al.,
Sasowsky, 2009).

the distance,
, between scallop crests. Modified from Curl (1974).

continues in the downstream direction. Scallop lengths are defined as
the distance to the wall at D. Upon impacting the wall, some water flows upstream
and recirculates within the vortex labeled C, while the remainder
stream flows that cannot be safely observed (i.e.,
floods) and ancient (paleo) hydrological conditions
in active and inactive caves in order to understand
flow conditions that cannot be safely observed (i.e.,
floods) and ancient (paleo) hydrological conditions
(Springer & Wohl, 2002, Palmer, 2007; Woodward &
Sasowsky, 2009).

Curl’s (1974) methodology was empirically validated
using scallop-derived discharges (Lauritzen et al.,
1985) and by Jeannin (2001) using spring discharges
and hydraulic modeling. The scallop-derived
discharges of Lauritzen (1982) were consistent with
the continuity equation, whereby the discharge
downstream of joined passages equaled the sum of
the discharges calculated for the individual passages.
The total discharge was equivalent to the 85th to 95th
percentile of flows measured at a spring (Lauritzen et
al., 1985). Jeannin’s (2001) results were consistent
with these percentiles. Lauritzen (1982) concluded
that his scallop-based discharges have a precision of
±5.8% and, notably, he reported approximate error
ranges based on the standard deviation of scallop
lengths (further discussed below).

The reliability or validity of other reported scallop-
derived velocities and discharges are difficult to
assess because of differences in the assumptions and
methods used. These differences extend to such basic
decisions as which scallops to measure and what
statistical mean to use in the equations of Blumberg
& Curl (1974) and Curl (1974). The latter authors
used the Sauter mean (Ls), which is calculated as:
\[
L_{32} = \frac{L_i}{L_{32}^{e^{s_{32}} - L_{32}}} + \left( L_{32} - \frac{L_{32}}{s_{32}} \right) (1)
\]
where \( L_i \) is the length of the \( i \)th scallop and \( s_{32} \) is the
the corresponding standard deviation (Lauritzen, 1982).
The empirically-based formula for \( s_{32} \) is:
\[
s_{32} \approx \left[ \frac{13}{n(n-1)} \sum_{i=1}^{n} (\ln L_i - \ln L_{\text{mean}})^2 \right]^{.5} (2)
\]
where \( n \) is the sample size. Lauritzen (1982) states
that the equation for \( s_{32} \) is approximate within 10-20%.
Some workers have used the arithmetic mean instead
of \( L_{32} \) (Table 1). Sauter means are greater than arithmetic
means because cubing large values effectively weights
them heavier relative to lesser lengths, which was the
intent of Blumberg & Curl (1974) and Curl (1974).
Given the inverse relationship between scallop lengths
and velocity, Sauter means yield lower velocities and
discharges than arithmetic means. Blumberg & Curl
(1974) directly measured velocities in their flumes,
so their choice of the Sauter mean was presumably
directly informed by their experiments and use of
arithmetic means is not appropriate.

The arithmetic mean is itself skewed by extremely
large and small values, but its use is most justified
where data are symmetrically distributed around a
central value such that largest and smallest values
counterbalance one another. Hence, the arithmetic
mean is commonly used for normally distributed
values. Palmer (1976) justified use of the arithmetic
mean on the basis that large scallops were the most
complete and assumed that only the lengths of
complete scallops were representative of formative
velocities. Hence, she only measured scallops she
regarded as complete, which she observed were most
commonly the largest scallops. Palmer (1976) reports
velocities calculated using the Sauter and arithmetic
means. In all cases, as expected, the arithmetic
means were smaller and velocities higher – often being twice
that obtained from the Sauter mean. The effect of
sampling bias is addressed below.

Hypothetically, if scallops formed only at a single
velocity and were not modified between flows they
would have a uniform distribution, but scallops
truncates one another as they migrate into and along walls and localized flow fields and substrate defects can create scallops across a range of sizes even when discharge is fixed (Vilien et al., 2005). As a result, fields of scallops should be viewed as stochastic phenomena that – if we assume the hydrologic regime is unchanging – are representative of the mean flow field across a portion of the wall during the formative floods. If this is true, no single scallop length (e.g. the largest or most complete) is uniquely representative of the formative velocity and descriptive statistics are required when using natural scallops. This is consistent with the methods used by Blumberg & Curl (1974) irrespective of their choice of the Sauter mean.

If descriptive statistics are required, what is the minimum number of scallops (n) that should be measured? Workers have used as few as two (Harman, 2012) and in excess of 100 (Lauritzen, 1982) and published sample sizes are mostly in the range of 10 to 30 (Table 1). When it comes to sampling, more is always better, but the actual choice of n generally reflects a balance between time, expense, and (in caves) the discomfort created by the environment (e.g. standing in meltwater). It may be possible to photograph or laser-scan a large number of scallops and use software to create a 3-D representation suitable for digitally measuring “all” scallops (e.g., Lundberg et al., 2017), but such methods have steep learning curves and come with their own time constraints, difficulties, and expenses. We explore the effect of sample size using Monte Carlo simulations to sub-sample 100 scallop measurements with n ranging from 10 to 30. An upper limit of 30 measurements was chosen because such samples are generally log-normal and, therefore, presumably reflective of the population they were drawn from (discussed below) (Hall, 2019).

A total of 100 scallops were measured in adjacent "patches" with n = 50 for each sample. The two samples are combined because they are continuations of each...
other. The scallops are within and adjacent to the area shown in Fig. 1. Scallops were measured without regard for their lengths or inferred “completeness”; choosing measurements based on completeness assumes all functional scallops have the same geometry, which may not be true, is inconsistent with Blumberg & Curl (1974), and introduces explicit user biases into the process. Lengths were measured to the nearest millimeter using a ruler with length defined as the distance separating the upstream and downstream scallop crests (Fig. 2).

Water velocities were calculated using Curl’s (1974) equations, which are valid for straight passage segments with parallel walls and elliptical cross sections. Each equation contains multiple constants, some of which are temperature dependent, and two unknowns: measures of scallop lengths ($L_{32}$) and passage dimensions (width or hydraulic diameter). The equation for an elliptical cross section is:

$$\bar{u} = \frac{\mu}{\rho L_{32}} Re^{*}[2.5 (\ln \frac{D_h}{2L_{32}} - \frac{3}{2}) + B_L]$$

where $\bar{u}$ is mean velocity, $\mu$ is dynamic viscosity, $\rho$ is fluid density, $Re^*$ is a Reynolds number based on scallop length(s) and friction velocity, $D_h$ is hydraulic diameter, and $B_L$ a constant related to velocity profiles. Based on experiments, $B_L$ equal 2200 and 9.4, respectively (Blumberg & Curl, 1974). We used $B_L$ values appropriate for water temperatures of 5ºC. $D_h$ is equal to $4 \cdot (\text{cross sectional area} / \text{length of the ellipse perimeter})$.

Data were processed and analyses performed using the free software R (R Core Team, 2019). As shown below, scallop lengths are log-normal, so lengths were log10 transformed for all statistical tests requiring normally distributed data. A statistical significance threshold (alpha) of 0.05 was used for all tests. Monte Carlo simulations were run in R using for statements and the built-in sample function. The R scripts can be obtained from the corresponding author and the data are available as a supplemental table in the appendices of Hall (2019).

We report the interquartile ranges (IQRs) of simulated $L_{32}$ and discharge values and compare them to standard deviation bounds on the Sauter mean of all 100 scallops. The width of the IQR is used as a proxy for precision instead of the standard error because the many means calculated using the sample function are neither normally nor log-normally distributed. Also, absolute values of standard deviations calculated from sample data are asymmetrical around the mean when back-transformed from log space. We note that ±1 standard deviation encompasses ~68% of the data, whereas the IQR only encompasses 50% of the observed values.

RESULTS

Statistical distribution

Lauritzen (1982) notes that scallops appear to be log-normally distributed. The 100 Boarhole scallops failed a Wilks-Shapiro test for normality ($p << 0.001$), but their log10-transformed values yield $p = 0.06$ and their distribution is very similar to a normal distribution (Figs 3 and 4). However, the upper tail is heavier than the lower tail, may be the result of an unintentional sampling bias. We also measured 60 samples of scallops in active stream passages of three nearby caves. Sample sizes were from 30 to 40 scallops and 57 of the 60 sets are log-normal ($p > 0.05$) (Hall, 2019). But given the wide range in scallop lengths commonly observed in field data a large sample size is recommended to adequately test for log-normality because the Wilks-Shapiro test is sensitive to the tails of a distribution.

![Fig. 3. A) Histogram of scallop lengths. Note the left skew; B) Histogram of log10-transformed scallop lengths (unitless). The distribution passes a Shapiro-Wilks test for normality ($p = 0.06$).](image)

![Fig. 4. Cumulative percent smaller diagram for all 100 scallops. The dashed line passing through the data represents a truly log-normal distribution, but the upper tail of the observed distribution deviates from the line and may be the reason the data marginally passed a normality test.](image)
Monte Carlo simulations were run in R using the 100 Boarhole scallops to obtain 10,000 Sauter means for each $n$ in the range of 10 to 30 (Fig. 5). Discharges were calculated simultaneously by multiplying by passage area (Fig. 6). The Monte Carlo subsampling is equivalent to sampling the wall with no preference for scallop size or location, which is consistent with unbiased sampling techniques (citation). However, the large number of iterations per $n$ and random chance led to extremes wherein subsamples were dominated by very small or very large scallops, which is a rough simulation of biased sampling. For small $n$, subsamples dominated by large scallops yielded Sauter means outside the boxplot whiskers representing 1.5 times the interquartile range (IQR). The IQR is between the 25th and 75th percentiles. The outliers, as defined by R, are shown as circles (Fig. 5F). Larger sample sizes did not yield outliers when plotting boxplots.

As might be expected, the smallest subsample size ($n = 10$) yields a wide distribution (Fig. 5A), which is skewed to the right and ranges between 35 and 90 mm. These yield discharges ($Q$) between 3 and 9 m$^3$ s$^{-1}$ (3-9 cumecs) (Fig. 6A). For $n = 10$ to 17, the IQR extends outside the gray zone denoting $L_{12} \pm 1$ s.d. calculated using all 100 scallops. The IQR at $n = 30$ is well within the gray zone and the average Sauter mean (black bars within boxes) is close to the “true” value obtained from all 100 scallops (dashed line). The distribution of Sauter means narrows considerably between at $n = 10$ and $n = 30$ (Fig. 5A-E), but the ends of the boxplot whiskers at $n = 20$ and $n = 30$ differ only by a few mm.

Discharge calculations are much more prone to outliers than Sauter means (Fig. 5F and 6F). At $n = 30$, discharges range from 3 to 8 cumecs, but the IQR is within the gray zone denoting $Q \pm 1$ s.d. calculated using all 100 scallops. All distributions (Fig. 6A-E) are skewed to the right, which shows that a subsample dominated by small scallops is more likely to yield outlier values than a subsample dominated by large scallops (Fig. 6F).

**Choice of means**

Using all 100 scallop measurements, the Sauter mean-based discharge is 5.1 cumecs and the arithmetic mean-based discharge is 6.7 cumecs (Fig. 7). We simulated the effect of mean choice when using $n = 30$ and generated 10,000 discharges using both $L_{12}$ and the arithmetic mean. As would be expected, use of $L_{12}$ yielded lower discharges than use of the
arithmetic mean (Fig. 7) because the Sauter mean places greater weight on large values and these yield lower velocities and discharges. There is comparatively little overlap in the $L_{32}$ and arithmetic mean-based distributions, demonstrating the non-trivial effect of which mean to use.

Fig. 6. Distributions of discharges calculated using $L_{32}$ values obtained from the Monte Carlo simulations for different sample sizes. The range of possible values is still large even for $n = 30$, but the interquartile range of the simulated means falls within the shaded region in (F) corresponding to bounds created using the standard deviation of all 100 samples via equation (2).

(DISCUTION)

In theory, scallop length correlates to the water velocity associated with a dominant discharge (Lauritzen et al., 1985), but scallops on cave walls are rarely observed to be uniformly of one size because scalloping is a stochastic process influenced by many variables. Hence, the 100 observed scallop lengths vary by nearly an order of magnitude and are log-normally distributed, although the upper tail of the distribution marginally conforms to expectations based on the normal distribution (Fig.s 3 and 4). Based on field observations, the log-normality partially reflects truncation of some scallops by their neighbors, but presumably also local variations in flow and a host of other factors. After measuring ~1,800 scallops as part of a larger project, we note that many short scallops are complete and not the result of truncation. So, measuring only the largest “complete” scallops ignores the stochastic, non-linear phenomena responsible for scalloping. The stochasticism can be seen in the scallops reported by Blumberg & Curl (1974) and we recommend measuring scallops of all sizes, so as to
Scallop uncertainties

be consistent with the methods behind the formulas used in calculating water velocities and in recognition of the stochastic origin of scallops.

Arithmetic means calculated using some or all of the lengths will be smaller than equivalent Sauter mean-based values and bias velocities and discharges toward greater values (Fig. 7). The constants reported by Blumberg & Curl (1974) were back-calculated from known (measured) velocities and their calculations used the Sauter mean. Had they used the arithmetic mean, the estimated values of their “universal” constants would have been different. This and the choice of means may partially explain why others have reported different values for the constants reported by Blumberg & Curl (Goodchild & Ford, 1971). The origins of Re*, B*, equation (3) invalidates using the arithmetic mean in velocity and discharge calculations, but even setting that aside, randomly resampling 100 scallop lengths clearly demonstrates that using the arithmetic mean will bias discharges toward higher values. We recommend use of the Sauter mean because it is consistent with the methods used to obtain constants in the relevant equations.

The choice of sample size has significant impacts on the range of possible L20 values and hence precision. The fact that scallops are log-normally distributed means that small sample sizes may not adequately represent the full range of lengths at a given location and this is why the ranges of simulated L20 values in Fig. 5 become increasingly narrower with increasing sample size. Interestingly, the interquartile range of simulated L20 values for n = 20 is similar to ±1 standard deviation around the Sauter mean of all 100 scallops (Fig. 5F), which leads us to recommend sample sizes of no less than 20 scallops. Using a sample size of 30 will always be better than using a sample size of 20, but measuring those 10 additional scallop lengths will, on average, only marginally improve the precision of an L20 estimate (Fig. 5F).

The precision of velocity and discharge measurements improves significantly as n increases from 5 to 20 (Fig. 5A-C). For n ≥ 20, the IQR of possible discharges is similar to ±1 standard deviation around the discharge calculated using the Sauter mean of all 100 scallops, but in this case the ranges represented by IQR and ±1 standard deviation are ~1 cumec. This implies that the choice of which scallops to measure is significant enough that reported velocity and discharge values should at best extend to one or two significant digits. Reporting discharges similar to ours as precise three or more significant digits is overly optimistic and we encourage reporting them to only 1 or 2 significant digits. In our case, this would correspond to uncertainties of as much as 20%. Accepting this limited precision or large standard error is important if inferential statistics are to be applied to the velocities or discharges.

CONCLUSIONS

Scallop measurements should be made when studying them because their distributions are log-normal and smaller sample sizes yield poorly reproducible results. Scalars of all sizes should be measured in recognition of the stochastic origin of scallops and to avoid user biases. In fact, the choice of which scallops to measure can significantly affect the final outcomes of a study and the uncertainties are such that results should probably only be reported to one significant digit. Reporting greater precision masks the underlying uncertainties associated with the wide range of scallop sizes commonly observed and the possibility of sampling biases.

ACKNOWLEDGEMENTS

We thank three reviewers for constructive critiques that improved the manuscript. We thank the West Virginia Association for Cave Studies for logistical and field support. This research was partly funded by a grant from endowments established by alumni of the Department of Geological Sciences at Ohio University, Athens.

REFERENCES

Bird A.J., Springer G.S., Bosch R.F. & Curl R.L., 2009 – Effects of surface morphologies on flow behavior in karst conduits. In: White W.B. (Ed.), Proceedings of the 15th International Congress of Speleology, Kerrville, TX, 3, p. 1417-1421.
Blumberg P.N. & Curl R., 1974 – Experimental and theoretical studies of dissolution roughness. Journal of Fluid Mechanics, 65: 735-751. https://doi.org/10.1017/jfm.1974.5
Burrill K.A. & Cheleuget E.L., 1998 – Corrosion of Candu Outlet Feeder Pipes, AECL, Japan, 31.
Charlton R., 2003 – Towards defining a scallop dominant discharge for vadose conduits: some preliminary results. Cave and Karst Science, 30: 3-7.
Checkley D. & Faulkner T., 2014 – Scallop measurement in a 10m-high vadose canyon in Pool Sink, Ease Gill Cave System, Yorkshire Dales, UK and a hypothetical post-glacial canyon entrenchment timescale. Cave and Karst Science, 41: 76-83.
Claudin P., Durán O. & Andreotti B., 2017 – Dissolution instability and roughening transition. Journal of Fluid Mechanics, 832: R2. https://doi.org/10.1017/jfm.2017.711
Curl R., 1974 – Deducing flow velocity in cave conduits from scallops. The NSS Bulletin, 36: 1-5.
Curl R., 1966 – Scallops and flutes. Transactions Cave Research Group of Great Britain, 7: 121-160. https://doi.org/10.1090/S0002-9947-1966-0185084-5
Davis J.C., 2002 – Statistics and data analysis in geology. Wiley, New York, 656 p.
Despain J. & Stock G.M., 2005 – Geomorphic history of Crystal Cave, Southern Sierra Nevada, California. Journal of Cave and Karst Studies, 67: 92-102.
Despain J.D., Tobin B.W. & Stock G.M., 2016 – Geomorphology and paleohydrology of Hurricane Crawl Cave, Sequoia National Park, California. Journal of Cave and Karst Studies, 78: 72-84. https://doi.org/10.4311/2013ES0114
Gale S.J., 1984 – The hydraulics of conduit flow in carbonate aquifers. Journal of Hydrology, 70: 309-327. https://doi.org/10.1016/0022-1694(84)90129-X
