26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Estimating streamflow discharge using the ‘wisdom of the crowd’

C Petheram[a], M Gibbs[b], J Hughes [c], L Seo[c]

[a] CSIRO Environment, Hobart, Tasmania, [b] CSIRO Environment, Adelaide, South Australia, [c] CSIRO Environment, Canberra, Australian Capital Territory
Email: cuan.petheram@csiro.au

Abstract: The estimation of streamflow discharge is fundamental to water resource and environmental management. However, streamflow gauging stations are expensive to install and maintain and consequently large parts of Australia, and the world more broadly, remain effectively ‘ungauged’. Cost effective methods for estimating streamflow discharge would help enable streamflow measurement networks to be expanded. One potential approach integrates citizen science initiatives with the use of time-delay trail cameras, a technology that is becoming increasingly affordable and accessible for remote monitoring.

The Wisdom of the Outer Crowd (WotOC) is a concept that large groups of people can collectively make better decisions and predictions than a single expert. The underlying rationale is that aggregating the opinions of a large number of independent and diverse individuals reduces bias and noise. In this study we test the performance of WotOC for estimating streamflow velocity. Additionally, we examine a cognitive variant known as the "Wisdom of the Inner Crowd" (WotIC). This approach attempts to simulate diversity within a single mind by prompting participants to adopt different perspectives, such as estimating from the viewpoint of another person, thereby eliciting a ‘crowd’ of judgments from a single source.

For this study we surveyed a ‘wise’ crowd of 31 people during a live experiment at the 2025 MODSIM conference in Adelaide. With the permission of conference organisers, the results were written up as an extended paper and reviewed retrospectively. Participants were shown video footage of streamflow from two contrasting sites with known discharge: the North West Bay River and Huon River in southern Tasmania. To test the utility of WotIC, participants were asked to estimate the mean cross-sectional velocity from i) the perspective of four specific individuals; and ii) based on videos captured at different positions and angles of the same river. WotOC and WotIC estimates were then compared to measured velocities at the sites. Streamflow velocity was also independently estimated using space-time velocimetry software Hydro-STIV, utilising the same video imagery shown to the MODISM crowd.

This experiment demonstrated that individual guesses of streamflow velocity are highly unreliable and subject to large errors. However, the geometric mean of these guesses proved to be a robust method for estimating streamflow velocity. Aggregating the estimates of 31 people using the geometric mean resulted in a 77% reduction in error at the North West Bay River site and a 76% reduction at the Huon River site, relative to a single observer. This suggests that WotOC is a viable method for estimating streamflow velocity irrespective of river size. In both cases error was roughly halved with a crowd size of five, with most of the accuracy gains occurring within the first ten participants. While WotOC was effective, space-time velocimetry performed closer to the measured values at both sites.

This study did not show any benefit of employing the WotIC techniques described for estimating streamflow velocity. This is primarily because it is challenging for an individual to escape their own cognitive biases — a hurdle that a diverse group of individuals can more easily overcome.

While WotOC was not as precise as space-time velocimetry, the slight overestimation observed in the crowd's results may be in part explained by some participants intuitively estimating surface velocity rather than the cross-sectional average. Because the cross-sectional average velocity is typically approximately 85% of the surface velocity, this perceptual shift likely accounts for some of the variance. The WotOC approach may have significant applications in developing countries and citizen science initiatives where ground control points cannot be established or where technical expertise in video processing is limited. Furthermore, these results suggest that WotOC may be applicable to estimating other hydrological parameters that are difficult or expensive to measure, such as transmission losses.

Keywords: *hydrology, velocimetry, streamflow gauging, water resource management, citizen science, collective intelligence*

1. INTRODUCTION

The estimation of streamflow discharge is fundamental to water resource and environmental management. However, streamflow gauging stations are expensive to install and maintain; consequently large parts of Australia, and the world more broadly, are effectively 'ungauged'. One such area is northern Australia, a relatively uninhabited region of 3 million km^2, with a very low streamflow gauging density (Petheram et al. 2009). Northern Australia is characterised by an extended dry season and consequently many of the rivers across the north are ephemeral with highly seasonal flow (Petheram et al. 2008).

During the northern dry season, it is possible to measure cross-sectional dimensions of dry riverbeds and secure pressor transducers and loggers to continuously measure water level depth. However, accessing sites in northern Australia to measure water velocity is often problematic; sites are remote, many can become inaccessible during the wet season, and the large distances make timely access to measure flow events using traditional methods implausible.

In recent years, methods for estimating streamflow discharge based on estimating streamflow velocity using video footage captured by side-mounted cameras, drones and satellite (Pearce et al. 2020, BoM 2021, Gibbs et al. 2025) have become increasingly popular. However, these methods require sophisticated software and ground control points (GCP) of known distances on both banks within the video image to transform raw video pixels into real-world physical coordinates. In some circumstances GCP can be difficult to identify and establish and typically require specialised equipment to achieve the necessary accuracy.

A simpler, cost-effective method with potential utility in developing countries and in remote, inaccessible regions would be for individuals to estimate streamflow discharge by viewing remotely attained video imagery collected using mobile phones or increasingly affordable time-delay trail cameras. However, individual estimates of physical characteristics are subject to considerable inter-observer variability due to heuristics and inherent subjective and systematic cognitive bias. In this manuscript, we explore the use of 'collective intelligence', specifically 'Wisdom of the Crowd' (WotC), as a means of neutralising random heuristic errors and cancelling out systematic biases when individuals estimate streamflow velocity.

The philosophical foundation of WotC is credited to Aristotle (Surowiecki 2004). However, the first scientific study was undertaken by Francis Galton in 1906. Averaging estimates of the weight of a slaughtered ox made by 787 people at a country fair in Plymouth, Galton observed that the average value of 1,197 pounds was only one pound less than the actual weight (Galton 1907). Since then, the concept has been famously used to find a missing nuclear submarine (Richardson and Stone 1971) and serves as a justification for the modern trial by jury, where collective deliberation is used to minimise individual error (List and Goodin 2001). Today, the principle even underpins modern medical platforms such as 'CrowMed', which leverage a 'crowd' of retired physicians, medical students and researchers to evaluate anonymous patients with mysterious illnesses (Kurvers et al. 2016).

However, crowds are not always 'smart', as noted by Mackay (1852), and sometimes exhibit 'groupthink' or 'herd mentality' characteristics. Famous examples include the misidentification of the Boston Marathon bomber by a 'crowd' of Reddit users (subreddit r/findbostonbombers) and users of other social media platforms such as twitter. In a study of twitter users, Starbird (2014) found that while corrections to misinformation did emerge, they were muted compared with the propagation of the original misinformation; the net effect was the creation of an echo chamber and loss of independence. Similarly, a groupthink mentality emerged in the lead-up to the 2008 stock market crash where market participants were incentivised to ignore 'red flags' because everyone else was doing the same. (Bénabou 2013).

Consequently, it is important to understand the conditions when WotC is a powerful tool rather than a liability. Pointing to spectacular failures such as those mentioned above, Surowiecki (2004) argue that for a crowd to be 'wise', the following four criteria must be met:

1. Diversity of opinion — each person should have some private information.
2. Independence —people's opinion should not be influenced by those around them.
3. Decentralisation —people can specialise and draw upon local, specific knowledge.
4. Aggregation — a mechanism exists for turning individual private judgements into a collective decision.

Aggregating information from a crowd of people is more specifically referred to as Wisdom of the Outer Crowd (WotOC). However, researchers have also had success in aggregating multiple, staggered guesses from single individuals, a concept referred to as Wisdom of the Inner Crowd (WotIC). Fundamental to the success of WotIC is the ability of an individual to make guesses from different perspectives. This can be induced using a time delay between successive guesses, as documented by Van Dolder and Van den Assem (2018) who found that participants who made a second guess weeks later were significantly more accurate than their first attempt; furthermore, averaging the two guesses reduced the error by nearly 30%. An alternative technique is 'dialectical bootstrapping' where after a first estimate is made, the individual is asked to purposefully argue against it to create 'cognitive diversity'. This can be achieved by asking "Why might my first guess be wrong", "What evidence did I ignore", or asking the individual to guess by taking the perspective of someone with whom they regularly disagree. Herzog and Hertwig (2009) found that averaging two 'dialectical' guesses is consistently more accurate than the first guess alone.

In the field of hydrology, Strobl et al. (2020) undertook a citizen science experiment where citizens were asked to provide estimates of discharge both directly and indirectly (by estimating the width, depth and surface velocity) of 10 rivers in Switzerland. Participants struggled to estimate discharge directly due to unfamiliarity with the units; most were more comfortable estimating discharge via width, depth and velocity, as they were more familiar with the units of those parameters. Although the authors focused primarily on the ability of individual citizens to accurately estimate discharge in real time and in person, they observed that the median of the citizens' estimates was close to the measured values.

The focus of this study is to explore the utility of WotC techniques for estimating streamflow velocity from side-mounted video imagery at fixed locations, nominally locations where access for velocity measurements is problematic. In practice, video imagery recorded by fixed cameras during a northern Australia wet season can be downloaded the following dry season and easily shared with many people, lending itself to WotC approaches.

Data for this study were gathered via a live experiment involving a 'wise' crowd during our presentation at the 2025 MODSIM conference in Adelaide. The crowd was considered 'wise' because the individuals were not conferring when making their guesses and the MODSIM conference draws a diverse attendance from a wide range of scientific disciplines and geographic and cultural backgrounds. This manuscript and its results have been written up retrospectively with permission of the conference organisers.

2. METHODS

This section describes the methods for estimating mean cross-sectional streamflow velocity using 'collective intelligence' techniques. Individuals were incentivised to estimate streamflow velocity accurately through the offering of prizes. These estimates were compared to mean cross-sectional velocities measured by the Tasmanian Department of Natural Resources and Environment at two operational streamflow gauging stations. Although these measurements are likely affected by inherent measurement errors (McMillan et al. 2012), they are assumed to be the 'true' values for the purpose of this study. Streamflow velocity was also independently estimated using space-time velocimetry software Hydro-STIV (Fujita et al 2020) following the approaches described by Gibbs et al. (2023) and utilising the same video imagery shown to the MODISM crowd for the WotC experiments.

2.1. Wisdom of the Outer Crowd

Two contrasting sites with streamflow gauging stations in southern Tasmania were selected to test the utility of the WotOC for estimating streamflow velocity: the North West Bay River upstream of the Margate Water Supply Intake (Figure 1) and the Huon River at Judbury (Figure 2). At each location 30 seconds of video footage was captured at 720p High Definition at 30 frames per second using an iPhone 15 Pro Max. WotOC estimates of streamflow velocity at the North West Bay River and Huon River sites corresponded with Questions 1 and 6 respectively (see Section 2.3).

Figure 1 Frame from the video taken at the North-west Bay River upstream of the Margate Water Supply Intake

Figure 2 Frame from the video taken at the Huon River at Judbury

2.2. Wisdom of the Inner Crowd

Two WotIC experiments were conducted. In the first experiment, the authors sought to create cognitive diversity between an individual's initial guess and their subsequent estimates by asking participants to estimate the streamflow velocity at the North West Bay River site from the perspective of four specific people (questions 2 to 5): Francis Chiew (surface water hydrologist), Ryan McAllister (economist), Rebecca Doble (groundwater hydrologist) and a young male waiter. These four individuals are referred to herein as the 'target individuals'. Each target individual was asked to estimate the streamflow velocity at the North West Bay site prior to the experiment.

The crowd were shown each target individual's name, occupation and photograph before the corresponding question. It is noted that some members of the crowd may have known one or more of the target individuals. The fourth target individual, the waiter, was the only one not attending the MODSIM conference; he worked at a restaurant where the authors dined the night before the experiment. A random online photograph of a waiter of similar gender, age and attire was shown to the crowd rather than an actual photograph of the individual.

The second WotIC experiment was undertaken at the Huon River site. At this location, three additional 30 second videos were captured using an iPhone 15 Pro Max. While the videos were taken at the same site, they were filmed from three different positions (different riverbanks) and angles (i.e. looking upstream and downstream) to promote cognitive diversity. These videos corresponded with questions 7, 8 and 9 (see Section 2.3). The crowd was informed that the three videos depicted different rivers. At the conclusion of the experiment, participants were asked whether they recognised that the three video setups were all the Huon River site; only one person claimed to have realised they were the same location.

2.3. Survey design

Survey protocols

This experiment utilised the cloud based online survey software SurveyMonkey, which enabled real-time data collection. The software prevents multiple entries by logging IP addresses and placing a temporary cookie in the participant's browsers. The survey was anonymous and no personal information was collected, although participants could provide their names to be eligible for a prize. Names were used for this purpose only. Participants could withdraw from the survey at any time prior to final submission. Thirty-one participants undertook the survey, with all participants completing the survey.

The survey consisted of nine questions. Participants were instructed to limit their velocity estimates to values between 0.01 and 5 m/s, inclusive, and to provide estimates to two decimal places:

Question 1: River No. 1. What is the cross-sectional mean velocity in m/s.
Question 2. What did Francis Chiew (SW hydrologist) estimate the cross-sectional mean velocity in m/s
Question 3: What did Ryan Macallister (economist) estimate the cross-sectional mean velocity in m/s.
Question 4: What did Rebecca Doble (GW hydrologist) estimate the cross-sectional mean velocity in m/s
Question 5: What did the waiter estimate the cross-sectional mean velocity in m/s.

Question 6: River No. 2. What is the cross-sectional mean velocity in m/s.
Question 7: River No. 3. What is the cross-sectional mean velocity in m/s
Question 8: River No. 4. What is the cross-sectional mean velocity in m/s.
Question 9: River No. 5. What is the cross-sectional mean velocity in m/s.

Method of aggregation

Two common methods of aggregation are the arithmetic mean and geometric mean. The former is typically used when errors are considered linear, while the latter is used when errors are considered multiplicative. According to Weber's Law (Weber 1978), human perception of physical magnitudes is proportional rather than linear; consequently, estimation errors are expected to be multiplicative. Therefore, the geometric mean was used to aggregate the crowd's velocity estimates.

Calibrating the crowd

Research shows that humans are much better at relative magnitude estimation than absolute magnitude estimation (Laming 1984). To leverage this, the crowd was shown two 30 second videos of rivers prior to the survey. The first river had a very low cross-section mean velocity (0.02 m/s) and the second river had a high cross-section mean velocity (2 m/s). This served to anchor the crowd, enabling them to make relative estimates and potentially reducing individual error.

3. RESULTS

3.1. Wisdom of the outer crowd

Figure 3 shows a bivariate response plot of participants estimates for streamflow velocity at the North West Bay River and Huon River sites. The plot also includes the geometric mean of the 31 participants, estimates derived from space-time velocimetry and the measured ('true') values. The WotOC was closer to the measured value than 71% and 58% of the individual estimates at the North West Bay River site and Huon River site respectively. However, when using the logarithmic Euclidean distance as a measure of closeness to the measured values across both sites the WotOC outperformed 90% of individuals (i.e. 28 of 31) who comprised the crowd. This suggests that as the number of sites requiring streamflow estimates increases or the creation of a rating curve from multiple videos of the one site taken at different times, the WotOC will increasingly outperform any single individual. Space-time velocimetry performed closer to the measured value at both sites than the WotOC.

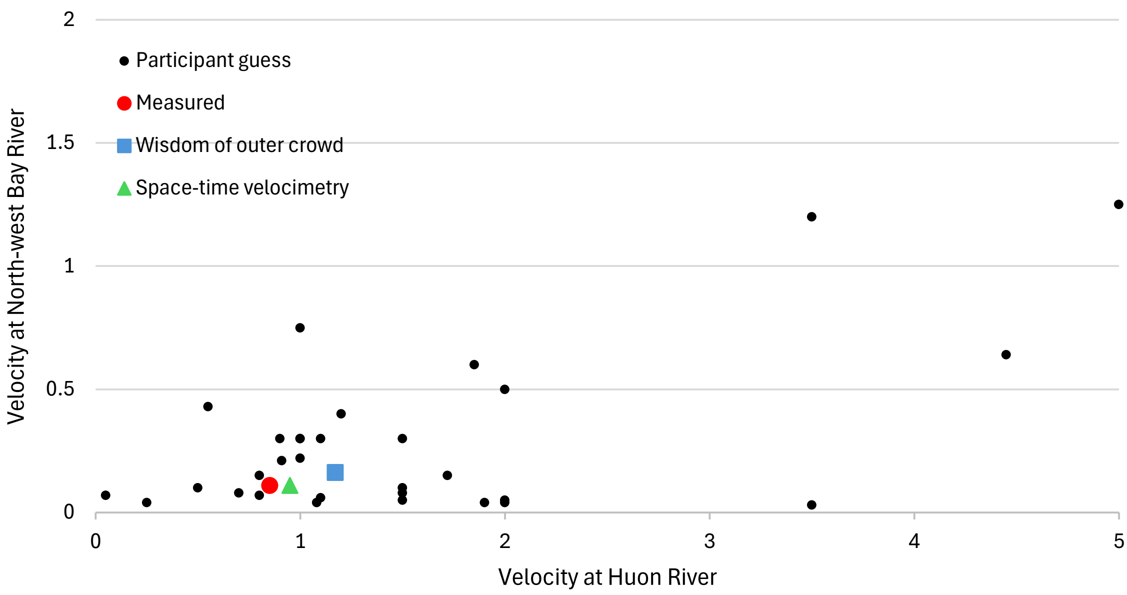

Figure 3 Bivariate response plot of velocity for North-west Bay River site and Huon River at Judburry

Figure 4 demonstrates that aggregating the estimates of 31 people using the geometric mean resulted in a 77% and 76% reduction in error relative to a single observer at the North West Bay River and Huon River sites respectively. This suggests that WotOC is a robust method for estimating streamflow velocity irrespective of river size. In both cases the error is roughly halved with a crowd size of five, with the majority of accuracy gains occurring within the first ten participants.

When the data are transformed to reflect the multiplicative nature of the error, low and non-significant log-Pearson (0.30) indicates that for the majority of the crowd, an individual's error on the first video was not related to the error on the second.

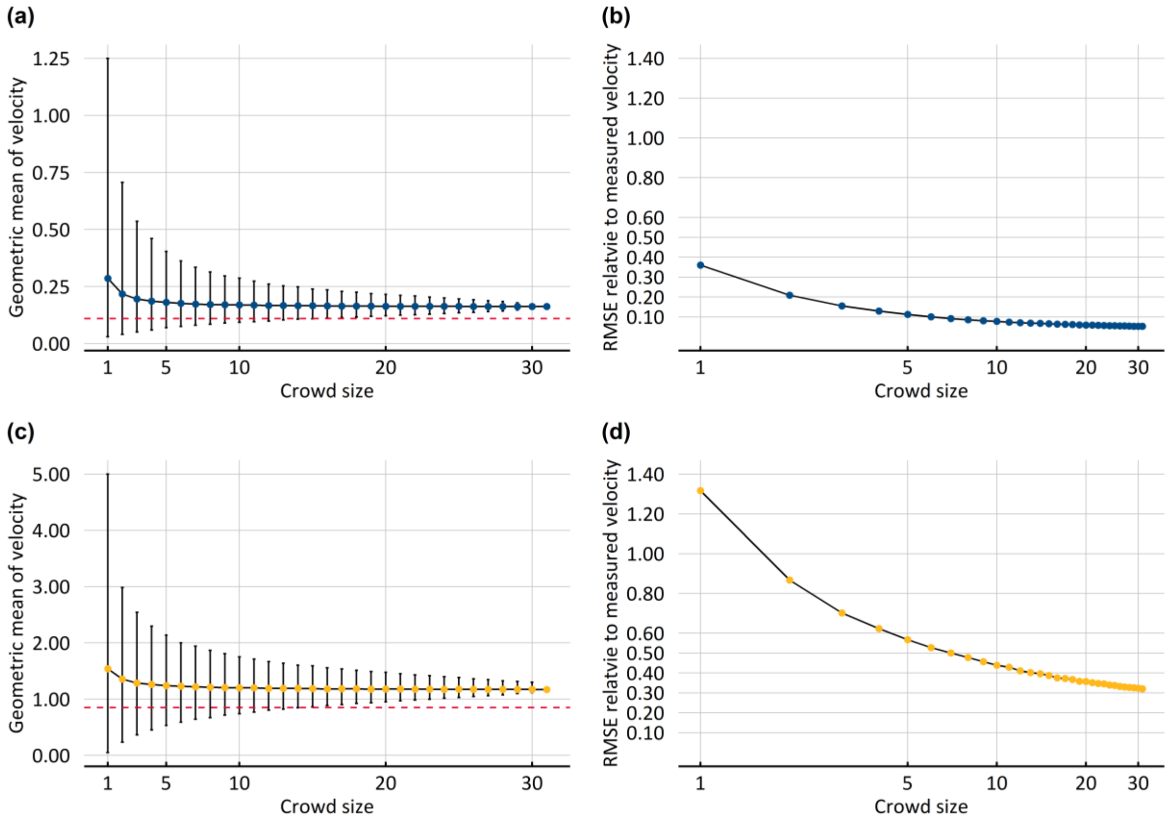

Figure 4 Sensitivity of streamflow velocity estimation accuracy to crowd size for the North-West Bay River (panels a and b) and Huon River (panels c and d) sites. Panels (a) to (d) were developed using a bootstrap resampling approach (sample size of 10,000), sampling without replacement. Dashed red lines signifis the measured ('true') values.

3.2. Wisdom of Inner Crowd

The WotIC experiment at the North West Bay River site resulted in only 14 of the 31 participants improving upon their first estimate. For the Huon River site the WotIC technique resulted in 17 of 30 participants improving upon their first estimate. Using the logarithmic Euclidean distance as a measure of closeness to the measured values across both sites the WotIC only outperformed individual estimates for 51% of participants (i.e. 16 of 31).

4. DISCUSSION AND CONCLUSION

This experiment demonstrated that individual guesses of streamflow velocity are highly unreliability and subject to large errors. However, the geometric mean of individual guesses was found to be a robust method for estimating streamflow velocity. Using the logarithmic Euclidean distance as a measure of closeness to the measured values at the two sites, the WotOC outperformed 90% of the individuals who comprised the crowd. Notably, there is no reliable way of knowing *a priori* which individuals may beat the crowd. Consistent with previous research, it was found that as the group size increases the marginal gain drops towards zero; for the purposes of estimating streamflow velocity there was little benefit of having a 'wise' crowd greater than 10-15 people.

The Huon River site exhibited a greater spread of data than the North West Bay River site. This is consistent with Webers Law, which states that human error scales with magnitude; the Huon River site had a considerably higher velocity (0.85 m/s vs 0.11 m/s). Furthermore, individuals viewing video footage may not merely be perceiving speed but potentially perceiving energy. Since kinetic energy is proportional to the square of velocity, higher-velocity environments create more turbulence and white-water, potentially further scaling the potential for human error.

This study did not show any benefit of employing the WotIC techniques described for estimating streamflow velocity. This is primarily because it is challenging for an individual to escape their own cognitive biases — a hurdle that a diverse group of individuals can more easily overcome. Inducing a time delay of several weeks between successive guesses may have resulted in more accurate WotIC estimates. Nonetheless Van Dolder and Van Den Assem (2018) demonstrated that even an infinite number of inner-crowd guesses is often less accurate than the average of just two independent individuals. Fiechter and Kornell (2021) demonstrated that as people become more accurate (i.e. experts) there was no benefit to WotIC.

WotOC was not as effective at estimating streamflow velocity as space-time velocimetry techniques. However, if participants intuitively estimated the surface velocity, as opposed to the cross-sectional average velocity, this may explain some of the slight overestimate of WotOC, where typically the cross-sectional average velocity is approximately 85% of the surface velocity (Rantz 1982). The WotOC approach may have application in developing countries and citizen science initiatives where GCP cannot be established or expertise in processing video is limited. The results of this study suggest the WotOC may also be applicable to estimating other hydrological parameters that are difficult to measure, such as transmission losses.

ACKNOWLEDGMENTS

We would like to thank Sam Fisher from Tasmanian Department of Natural Resources and Environment for providing cross-sectional information for the North West Bay River and Huon River sites, and Dr Shaun Kim and Dr Julien Lerat for reviewing this manuscript.

REFERENCES

Benabou R (2009) Groupthink: collective delusions in organizations and markets. Review of Economic Studies 80, 429–462.
Bureau of Meteorology (2021) National Industry Guidelines for hydrometric monitoring–Part 11: Application of surface velocity methods for velocity and open channel discharge measurements. Bureau of Meteorology: Melbourne, Australia.
Fiechter JL and Kornell N (2021) How the wisdom of crowds, and of the crowd within, are affected by expertise. Cognitive Research 6, 5.
Fujita I, Watanabe K, Iguchi M and Hasegawa M (2020) Improvement of STIV technique by using deep learning. Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering) 76(2), 991–996.
Galton F (1907) Vox Populi. Nature 75, 450–451.
Gibbs M, Guglielmo M, Gardner B, Hughes J and Petheram C (2025) Streamflow gauging from space – Space Time Image Velocity from satellite video. In: Asia Oceania Geosciences Society Conference; 27 Jul 2025 to 01 Aug 2025; Singapore.
Gibbs M, Hughes J and Petheram C (2023) Space-time velocimetry to estimate discharge during the 2022 River Murray flood. In: 25th International Congress on Modelling and Simulation (MODSIM2023); 09 to 13 Jul 2023; Darwin.
Herzog SM and Hertwig R (2009) The wisdom of many in one mind: improving individual judgments with dialectical bootstrapping. Psychological Science 20(2), 231–237.
Kurvers RH, Herzog SM, Hertwig R, Krause J, Moussaid M, Argenziano G, Zalaudek I and Wolf M (2016) Boosting medical diagnostics by pooling independent judgments. Proceedings of the National Academy of Sciences (PNAS) 113(31), 8777–8782.
List C and Goodin RE (2001) Epistemic democracy: Generalizing the Condorcet jury theorem. The Journal of Political Philosophy 9, 277–306.
Mackay C (1852) Memoirs of extraordinary popular delusions and the madness of crowds, Volume 1. Select Rare Books, 13.
McMillan H, Krueger T and Freer J (2012) Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrological Processes 26(26), 4078–4111.
Pearce S, Ljubičić R, Peña-Haro S, Perks M, Tauro F, Pizarro A, Dal Sasso SF, Strelnikova D, Grimaldi S, Maddock I, Paulus G, Plavšić J, Prodanović D and Manfreda S (2020) An evaluation of image velocimetry techniques under low flow conditions and high seeding densities using unmanned aerial systems. Remote Sensing 12(2), 232.
Petheram C and Bristow KL (2008) Towards an understanding of the hydrological factors, constraints and opportunities for irrigation in northern Australia: A review. CRC for Irrigation Futures Technical Report No. 06/08. CSIRO Land and Water Science Report No. 13/08.
Petheram C, McMahon TA and Peel MC (2008) Flow characteristics of rivers in northern Australia: implications for development. Journal of Hydrology 357(1-2), 93–111.
Petheram C, Rustomji P and Vleeshouwer J (2009) Rainfall-runoff modelling across northern Australia. CSIRO Water for a Healthy Country Flagship, Australia.
Rantz SE (1982) Measurement and computation of streamflow: Volume 1. Measurement of stage and discharge. U.S. Geological Survey Water-Supply Paper 2175, 131–139.
Richardson HR and Stone LD (1971) Operations analysis during the underwater search for Scorpion. Naval Research Logistics Quarterly 18(2), 141–157.
Starbird K, Maddock J, Orand M, Achterman P and Mason R (2014) Rumors, false flags, and digital vigilantes: misinformation on twitter after the 2013 Boston marathon bombing. iConference 2014 Proceedings.
Strobl B, Etter S, van Meerveld IJA and Seibert J (2020) Accuracy of crowdsource streamflow and stream level class estimates. Hydrological Sciences Journal 65(5), 823–831.
Surowiecki J (2004) The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. Doubleday.
van Dolder D and van den Assem MJ (2018) The wisdom of the inner crowd in three large natural experiments. Nature Human Behaviour 2, 21–26.
Weber EH (1978) The Sense of Touch (Ross HE and Murray DJ, Trans.). Academic Press.

