Reviewer 1

This is an interesting comparison of a variety of techniques for discharge estimation with a view to evaluating the CSVG stereophotogrammetry method for deriving discharges from surface velocity measurement, including the use of an (unspecified) adaptive learning algorithm. I do think, however, that the paper could be significantly improved, in part because the details of the CSVG method are kept almost deliberately vague as if to not give too much away (without actually saying so, though implied by the software code not being made available). However, this makes it really frustratingly difficult to understand what lies behind some of the results. It is suggested, for example, that the method can produce comparable discharge estimates to traditional rating curve and ADCP methods – but only really if a local water level measurement is available (not that this is really a problem these days when low cost methods are available). In the points below I have suggested many places where more detail is needed – if not in the paper directly then referenced to material in the supplementary file.

Reply: *We thank the reviewer for their time and detailed feedback, particularly with regards to improving the clarity of the methods. While we have not intended to provide deliberately vague details, we believe the size of the paper, including the weight of the methodology section and the scope of site data results, necessitated focusing less attention on the specific details of particular functions. We instead provided a balance of the detail required to describe the conceptual approach applied, and further to this detailing the most important components of the application of the approach. The software code is not available, as it is mostly comprising the integration of the hardware recording and analysis data structures with cloud databases and services embedded with the application of the methodology that is described for the analysis approach. We have answered the specific points below, including minor revisions based on valuable feedback from the reviewer for improving the clarity of the manuscript.*

Some specific points are follows:

I would suggest that the results are reordered somewhat so that each site is considered in turn as because the Paterson site is so different from the others – from the photo it would appear that this is only site where a downstream (rather than cross-channel) camera view is used with a flow that does not seem to have developed a uniform flow profile. The reasons for the failures here need more discussion (as shown in Fig S7).

Reply: *Thank you for your comment – we believe by your description that you are actually referring to the irrigation channel site with regard to the downstream facing camera view, whereas Fig S7 is from the Paterson site (as you say), which is shown in Figure 2b with a cross-channel camera view. We have now highlighted how different this site is from the others in the results section on L424. Furthermore, the description of the results in Figure S7 has been provided on L541.*

Added to L424: “It is important to note the irrigation channel site differs substantially from the other case study sites with a downstream field of view and highly turbulent flow conditions discharged through an engineered channel.”

L83 para. Yes, but what technological advances do you mean? Those in the current study? Those to come (in which case more detail needed). Might be better moved to end.

Reply: *Thank you for the insight for this clarification and suggestion for improvement in the flow of ideas. This statement does refer to the technological advances as applied and tested in relation to the advances that facilitated this work.* *In light of your feedback, we agree and have moved the statement to the end of the introduction section.*

L118. Not clear how this 40m relates to the 10m on L156, and how the camera resolution and the 120 degree field of view create the 0.1m analysis resolution?

Reply: *Thank you for raising this question. The 40 m mentioned on L118 is relative to the position of the camera and referring to the system hardware limitation for estimating water levels using stereophotogrammetry requiring the water near edge to exist within the vision of the camera within a 40 m range from the camera's physical location, while the 10 m on L156 is describing the region of the water surface used for stereophotogrammetry water level estimation relative to the near bank interception with the water edge. The camera resolution and field of view has no relation to the 0.1 m analysis resolution, which is simply the fixed size of the grid that the estimations are projected on (while in practice this grid size is actually an adjustable parameter, all results presented and all deployments of the system have used this default analysis grid resolution of 0.1 m).* *Further explanation of the optical flow resolution calculation step has been added to L121 to describe how the fixed 0.1 m analysis grid resolution is not directly connected to the camera resolution (which has these camera hardware limitations accounted for prior to reaching the analysis grid).*

Added to L121: "The optical resolution of the flow in meters per pixel is calculated based on the water surface projection in order to filter any motions in the area of the field of view beyond the limits of acceptable optical flow resolution accuracy (normally limited to a maximum of 0.05 meters per pixel up to 0.2 meters per pixel)."

L124. What do you mean by adaptive learning (you also refer to machine learning later)? No details are given. And here you do not mention the issue of going from surface velocities to profile or mean velocities (see comment on L221)

Reply: *This section is intended as an overview of the system to describe all the aspects and how they relate to each other as well as how this relates to the practical use of the system. The adaptive learning refers to the process described in the later sections of the methodology in 2.3 and 2.4 which result in an adapting surface velocity distribution (adaptive to new observations if changes occur at a site) and learning from new observations to add to the database of velocity distributions which are then each calculated for contribution to the discharge rating. As for mentioning the issue of going from surface velocities to profile or mean velocities, this is a well-known and studied feature of all methods for estimating discharge based on observations of surface velocities, as described in the introduction and the procedure applied for this is detailed at the beginning of section 2.4. Machine learning is referred to generally as the method presented was built to leverage collected sample data to improve performance in gauging stream flows, hence demonstrating the potential for machine learning approaches to overcome challenges in optical stream gauging using cameras.*

L158. Why the first percentile (indeed what does the first percentile mean)?

Reply: *Using the first percentile in this context essentially allows you to quickly take the near-minimum without taking any sporadic outlying minimum value arising from* *erroneous points in the generated point cloud.* ***We have added this note by modifying the line at L158 to improve the clarity of the reasoning behind the choice of approach.***

Modification of L158: "The first percentile of the elevation points of the stereophotogrammetry cross-section profile within this domain is then estimated as the water level (effectively taking the near-minimum of the surface while reducing the impact of any sporadic point cloud artefacts)."

L168. What is this minimisation problem? Since it will affect the estimates it needs more explanation – at least in the supplementary file

Reply: *Thank you for your question and keen interest in the detail behind the algorithms used in this work. We do not think it is reasonable to reproduce this explanation in detail and would like to direct you to section 3.2 of the reference source material for the Farneback optical flow algorithm cited in this line.* ***We have modified this section from L171 to add some detail of the Farneback algorithm for estimating optical flow and provide a reasonable summary of the background detail to the reader.***

Modification from L171: "Shi et al. (2020) compared three established and widely applied optical flow techniques to breaking surges, noting the advantages of the Farneback algorithm for its relatively high accuracy and dense flow fields, as well as a lower sensitivity to noise with the converging iterative solution for the displacement vector, d, between a pair of images using quadratic polynomials following Eq. (1):

$$d(\mathrm{X_{im}}) = \left(\sum_{\Delta X_{im} \in I_{local}} \mathrm{w} A^T A\right)^{-1} \sum_{\Delta X_{im} \in I_{local}} w A^T \Delta b_f \, , \tag{1}$$

where I is the greyscale image with local neighbourhood regions denoted by I_{local} using the image coordinates x_{im} and y_{im} to form $X_{im} = \begin{bmatrix} x_{im} \\ y_{im} \end{bmatrix}$, where the change in brightness between the corresponding pixels in the pairs of images are denoted ΔX_{im}. Furthermore, w is a weighting function over the local neighbourhood regions, while the polynomials are defined by $f(x_{im}, y_{im}) \cong a_1 + a_2 x_{im} + a_3 y_{im} + a_4 x_{im}^2 + a_5 y_{im}^2 + a_6 x_{im} y_{im}$ with $A = \begin{bmatrix} a_4 & \frac{a^6}{2} \\ \frac{a^6}{2} & a_5 \end{bmatrix}$, $b_f = \begin{bmatrix} a_2 \\ a_3 \end{bmatrix}$, and $c = a_1$.
The approach is a variational method combining the assumptions of local neighbourhood brightness intensity variation between frames with the minimisation of an energy function assuming a slowly varying displacement field for locally smooth velocity gradients (Shah and Xuezhi, 2021)."

L187. Motions out of the water surface? Some hint here of a limitation but these are on a surface, needs more explanation. And filtered how? As NANs, or with some replacement strategy?

Reply: *While we very much appreciate your excellent reviewing mindset towards finding potential limitations, and we appreciate that you would be aware of the many limitations present in the available/established methods for measuring natural open channel stream flows, the key motivation for this step of the procedure is the removal of motions that are optically visible to the camera, but are not part of the measurement of the planar surface velocity contributing to the measurement of the nett discharge of water through the stream section. As such, the concept of filtering is used in the regular sense of the word where unwanted material (motions in the vector field which are out of the plane of the assumed water surface) is removed without any other replacement strategy that isn't already described in the methods.* ***We have added the word 'assumed' for clarity to L187.***

Modification of L187: "From this point, the motions out of the assumed plane of the water surface are filtered out of the analysis to further remove false motions unrelated to the waterway surface velocities (such as animals and swinging ropes which are not moving in the assumed plane of the water surface)."

L190 Continuity of streamlines imposed how? What assumptions about the nature of the streamlines?

Reply: *Thank you for this comment, we agree that statement is not entirely clear.* *<u>We have modified this section from L189 to be more descriptive of the assumption of the nature of the flow over the analysed section.</u>*

Modification from L189: "Assuming the remaining velocities over the length of the analysis section are velocities related to the motion of the water surface, and assuming a continuity in the uniformity of the analysis section length without transitional flows, the strongest detected velocities are collapsed into a single-dimensional raw cross-section surface velocity profile. The assumed continuity over the analysis section length facilitates the measurement of velocities across spatially inconsistent optical flow measurement/lighting conditions along the length of the analysed section."

L198 "multiple measurements of the same water level over time in different conditions to.combine these measurements into a complete velocity profile" – totally obscure. Different measurements at the same water level should give you an estimate of variability of estimates at that water level, but why does it tell you anything about the profile. In fact you do not seem to consider the profile at all – only using data from elsewhere to estimate a coefficient to convert to mean velocity.

Reply: *We agree that this is confusing, and have attempted to take great care to use the words profile and distribution as clearly as possible when referring to velocities over the stream section. The intent of this statement was not to say anything about the profile of velocities beneath the water surface (which, as you know, are not directly measured by this approach), but to instead refer to the profile of surface velocities across the cross-sectional profile from one side of the stream to the other.* *<u>We have added the descriptor modification in L197 to make clear that the velocity profile being referred to is the surface velocity profile.</u>*

Modification of L197: "This process of developing an adaptive database of surface velocity measurements across the stream at different water levels (adaptive learning surface velocity distributions), allows the system to use multiple measurements of the same water level over time in different conditions to combine these measurements into a complete surface velocity profile, while simultaneously being adaptive to observed changes in surface velocity profiles in non-stationary environments."

L204. Why exponential? Are there not theoretical 2D cross-sectional distributions that you could have tried (though presumably would not be valid for the Paterson site). And in fitting the distribution, what if it is the highest values that are not available?

Reply: *The exponential relationship from the boundary distance factor to the surface velocity profile is just a simple logarithmic relationship of the rearranged form -bx = ln(1-V_s/V_infty). Many alternative relationships and boundary distance factor transformations were tested against optically estimated surface velocity distribution observations at different sites, but the form presented here was the best generically fitting to the data. Thankfully due to the nature of the optical approach, the highest surface velocity values in the distribution are ordinarily the most available owing to a high signal to noise ratio, but cases can arise where these highest values are out of view or obscured by vegetation or other visibility challenges. If the rest of the distribution is intact, then V_infty will be fitted based on the trending asymptote of the observed surface velocities in the transformed boundary distance factor domain. However, if the distribution of the surface velocities is not well-enough observed, then in any case you cannot very well predict the discharge using surface velocimetry unless* *you have learned these through previous observations or surrounding observations at different water levels (as are both included aspects of the approach presented).*

L221 – should not values of a be considered uncertain (and should this uncertainty not be propagated into the discharge estimate (see the cross-section you show in Figure S3)

Reply: *We agree that this is uncertain, and is why we keep track of an envelope which is mentioned in your next point about L230. In fact, we would generally agree that all discharge estimates should be reported as and thought about in terms of estimated ranges (as an indirectly quantified measure). While standard parameters for the calculation of 'a' are configured as part of the analysis configuration, a standard +-15% 'a' range is applied to the independently learning minimum and maximum surface velocity (and hence discharge) envelope boundaries provided with the data reporting. This discussion is then continued in the next point.*

L230. You do not say where these adapted learning distributions come from (and should that not also be associated with an uncertainty estimate using e.g. Bayes updating). You mention an "envelope" but that never appears later in the results.

Reply: *The authors appreciate that the origin of the usage of the term adapted learning (surface velocity) distributions may be unclear, and confirm that the adapted learning distributions indeed come directly from observations made by the system aggregated from different points in time where the same water level has been measured as explained in the previous section from L196 onwards and exampled in the supplementary Figure S3.* <u>*We have added the identifying term in L198 to make this connection clear and improve the clarity of the manuscript thanks to the reviewer's feedback.*</u> *The authors have decided to leave out the "envelope" results (which are generally enveloping of all available estimates) to avoid cluttering the results, and focus on the differences between the best estimate provided (particularly given that a single number is ordinarily taken as the best estimate from the gauging stations compared).*

Modification of L197: "This process of developing an adaptive database of surface velocity measurements across the stream at different water levels (adaptive learning surface velocity distributions), allows the system to use multiple measurements of the same water level over time in different conditions to combine these measurements into a complete surface velocity profile, while simultaneously being adaptive to observed changes in surface velocity profiles in non-stationary environments."

L237. We do not need quality codes – we need proper uncertainty estimates. You surely have the information to be able to do so.

Reply: *The authors agree with the reviewer around the need for proper uncertainty estimates, however we consider what constitutes a 'proper' uncertainty estimate to be reasonably debatable, and this paper does not yet seek to present techniques or support of any particular set of uncertainty estimation approach in connection with this work beyond the methodology detailed. The authors do not agree with the comment about quality codes, as this is a data documentation approach required and applied in practice by water agencies across Australia and internationally.* <u>*We have added to the discussion at L636 on uncertainty estimation methods with reference to a recent comparison study (Kiang et al., 2018).*</u>

Added to L636: "Alternative methods for estimating the uncertainty of stream discharge rating curves have been compared in Kiang et al. (2018), finding a wide variation in uncertainty estimates resulting from different methods which demonstrated the necessary careful selection and communication of the assumptions of the uncertainty estimates provided."

L261. Why NSE? That seems inappropriate for a rating curve since NSE scales by the observed variance which is here over the depth values). That is more like a regression so is not a correlation coefficient more appropriate?

Reply: *We appreciate the reviewer's feedback and point of view, but we believe that a correlation coefficient is not more appropriate than the NSE as a statistical metric in this circumstance. Whilst NSE is most notably applied as a skill metric for the fit of hydrological flow timeseries data due to the way it is less skewed by the more frequently observed and perhaps (depending on study objective) less important (and ordinarily easier to predict somewhat closely) low flow data relative to the less frequently occurring flow events which have more significant error margins. In this regard the NSE is more sensitive to extreme values, and the authors consider this to be important for appropriately assessing the rating curves as they have been likewise constructed through time by many data points towards the lower end of the discharge scale and fewer observations towards the higher end (which is also similarly the case for the manually gauged observation distributions). We have added this reasoning in brief to the end of this line at L258.*

Modification of L258: "At two existing government maintained gauging stations, historical manual gaugings have been compared along with CVSG, DischargeLab, and Hydro-STIV measurements relative to the latest published discharge rating using root-mean-square error (RMSE), the mean percentage difference, and the Nash-Sutcliffe Efficiency (NSE) (Jackson et al., 2019) commonly applied for assessing predictive skill for discharges in hydrological settings due to its sensitivity to extreme values."

L280 Table 1 – there seem to be some inconsistencies in presentation here (e.g. water levels of 135m and 0.31m are clearly not both relative to local datum?)

Reply: *Thank you for this comment, the authors recognise that the relevancy of the water levels presented is not in the absolute value, but rather the range of water levels over which observations occurred. In light of this, we have added the note that these water level ranges have been presented relative to local datums in the Table 1 caption while pluralising the relevant Table 1 heading.*

Modification from L280:

Table 1: Field case study sites summary (water level ranges presented relative to local datums).

Site	Period	Distance to stream (m)	Water levels (m)	Reference gaugings	Ground control reference points
Castor River, Ontario, Canada	30 s	-	3.77	1 concurrent (2019)	12
Irrigation channel, NSW, Australia	30 s	-	135.80	1 concurrent (2020)	10
Tyenna River, Tasmania, Australia	56 d	5.9–7.3	0.31–0.87	344 historical ('64 – '22)	9
Paterson River, NSW, Australia	122 d	0-22.5	0.78–10.54	157 historical ('87 – '21)	0

L311. More detail needed on the ADCP for clarity– was averaging over multiple transects or other filtering of anomalies down

Reply: *Thank you for this feedback towards improving the clarity of the manuscript. This detail has been added from L305.*

Modification of L305: "A 30 Hz 30-second video recording (3840x2160 pixel resolution) formed the basis for the surface velocimetry estimations, with a reference measurement provided by a series of four SonTek RS5 moving boat ADCP (San Diego, CA, USA) transects taken between 15 to 20 m downstream of the hydraulic control structure within a timespan of eight minutes and a maximum discharge estimation difference of 8.5% to the most outlying transect measurement."

L373. But Figure 3 does not really support this – either there appears to be little difference or for Paterson it seems disadvantageous.

Reply: *This line forms part of the overall results introduction which summarises the results section in its entirety before detailing the results of each section. The previous line of this section is applicable to Figure 3 with the caveat of 'under suitable conditions' noted. The authors have expanded this line further to clarify the subject of this statement.*

Added to L370: "The results of this work found broadly comparable gauging results using the raw data of the different measurement technology approaches employed, predominantly falling within a relative error of 15% under suitable conditions when comparing between the results of both the detailed surface velocity distribution case studies and longer deployment timescales evaluated."

L373. But is it not the large percentage that is greater than 0.5 m that is more significant (as clear in Figure 5)? It is unclear why a stereophotogrammetry method can be >0.5m in error for so much of the time. Is this a result of the particular camera system used? It is off the shelf but has only 120mm separation between the lenses.

Reply: *Figure 5 presents the percentage of water level error <0.5 m, whilst the >0.5 m error percentages are the remaining 2% and 38% for the Tyenna River and Paterson River sites respectively. We are not sure if the reviewer has personal experience with measuring water levels from a distance in natural riverine environments, or if the reviewer is aware of any previously published and evaluated datasets with comparable distances and timescales, but these are the results of the particular camera system we used while applying the method described to estimate water levels through stereophotogrammetry. Since this data was analysed, further development has improved this accuracy somewhat using a calculated ambient environment correction factor, but the results presented in this manuscript were produced and analysed prior to this additional development.*

L425. Well yes (look at the photo)! So should you not present this as a "test to failure" type of site? You would not actually have had to go much further downstream to have been more successful.

Reply: *We are pleased that the reviewer agrees with the statement made by the authors. However, moving the site downstream would represent a different site, which is not the site of the case study. Figure S1 in the supplementary materials has been revised demonstrating the relative locations and discharge estimations resulting between each of the methods along the section reach length. The authors previously described the optimisation of the analysis region for each measurement technology's* *most suitable region of interest reaching the best available result for the irrigation channel site in the methodology section on L314.*

Modification of Figure S1:

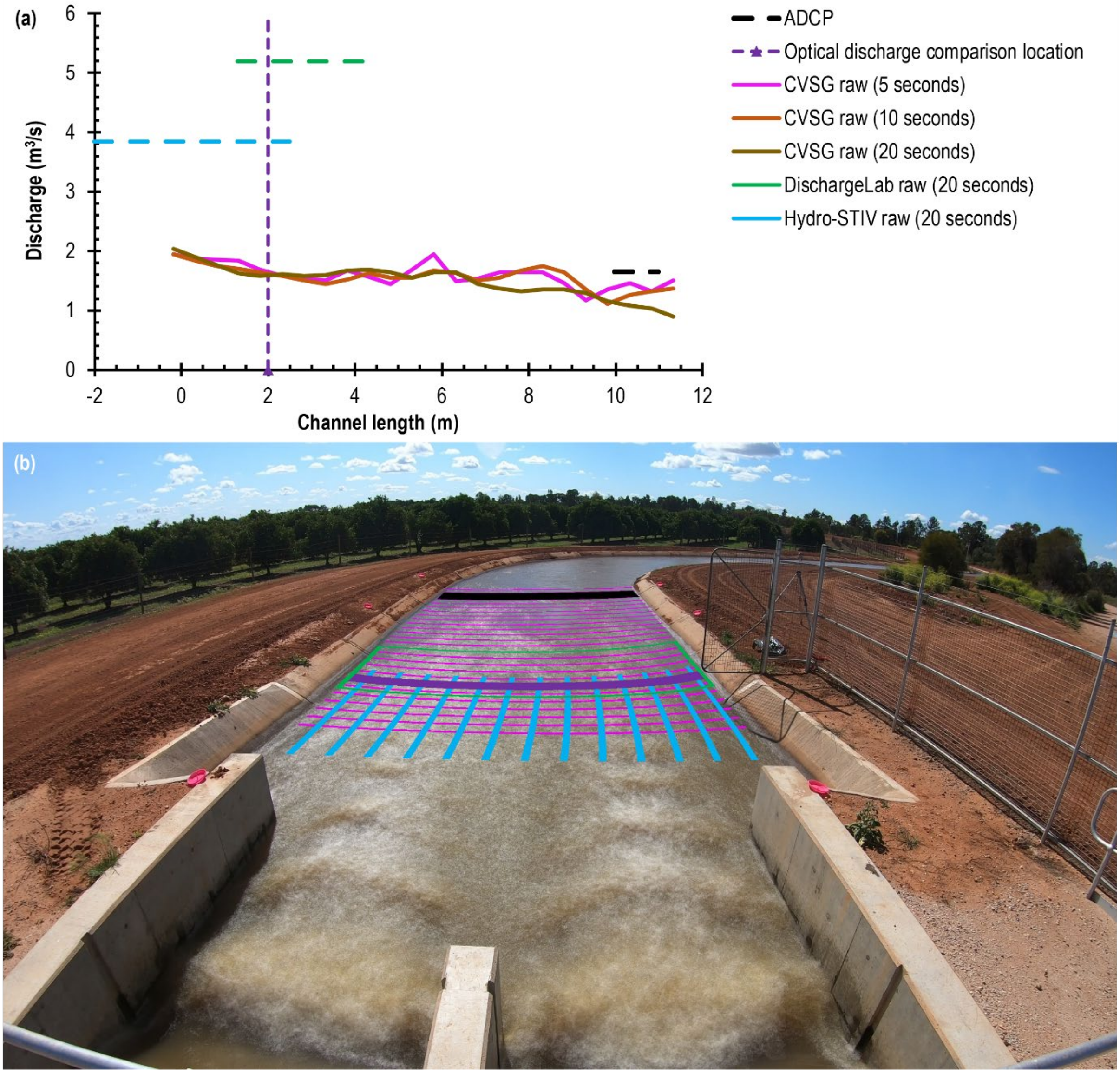

Figure S1: Raw discharge measurements using different technologies along the length of the irrigation channel in NSW, Australia.

L470. This appears to be a combination of trend in cross-section/rating as well as statistical observational variability for that depth. So when you refer to the "latest rating curve" – what period of observations is used to define that curve? (Also Figs 7,8, Table 6, etc later)

Reply: *The authors appreciate the reviewer's observation of this fact. We would like to clarify that the latest gauging station rating fit refers to the best estimate published by the government agency using a best fit of past manual gaugings by a professional hydrographer (representing the best available estimate using the currently applied technology and data for each site).* *__We have taken care to add the necessary__* *clarification detail to what this ‘latest rating curve’ means in the methodology from L255.*

Modification from L253: “Additionally, historical ADCP derived estimates of discharge used to develop discharge ratings were utilised as a reference. Whilst the most up to date discharge rating fits published by government agencies based on the professional judgement of hydrographers using the applied technology and data available prior to the deployment of optical methods at each site were used to represent the best available estimates.”

L502 Why do you refer to correlation plots without giving correlation coefficients?

Reply: *We do not feel that correlation plots necessitate the presentation of correlation coefficients unless this provides a relevant insight. The error has been broken down visually and quantified into different classification groups, presented in a way that the authors believe is more insightful and relevant to the context of the problem and data being evaluated. If there is a relevant reason for providing these coefficients, then the authors can add these to the figure or figure caption.*

L521What do you actually mean by learned discharge rating curve? Is it purely a filtered estimate over time that will average out error, or is other data input to the process (you have not said how it works earlier). Clearly if you input the actual levels (or weight by error relative the the measured level) you are going to get much closer to the “latest rating curve” as shown in the other plots (and Figs S5, or even S7).

Reply: *The learned discharge rating curve is described earlier in the methodology section 2.4 (titled ‘Adaptive learning discharge rating’). It has been applied precisely in the way it is explained by leveraging a fit across all of the adaptive learning surface velocity distributions that is described in the methodology section 2.3 for each of the observed water level increments. There is no additional input of actual (gauging station measured) water levels to assist in improving the discharge ratings derived from the stereophotogrammetry estimated water levels. The use of gauging station measured water levels is only applied completely independently in-place of stereophotogrammetry estimated water levels where the discharge from each are compared for evaluation purposes.*