Calculation of Discharge in Estuaries

Current State and Advancements

Queensland Government

Prepared by: Water Quality and Investigations, Department of Environment and Science

The Department of Environment and Science acknowledges Aboriginal peoples and Torres Strait Islander peoples as the Traditional Owners and custodians of the land. We recognise their connection to land, sea and community, and pay our respects to Elders past, present and emerging.

The department is committed to respecting, protecting and promoting human rights, and our obligations under the Human Rights Act 2019.

Disclaimer

If you need to access this document in a language other than English, please call the Translating and Interpreting Service (TIS National) on 131 450 and ask them to telephone Library Services on +61 7 3170 5470.

This publication can be made available in an alternative format (e.g. large print or audiotape) on request for people with vision impairment; phone +61 7 3170 5470 or email <library@des.qld.gov.au>.

Citation

Wallace SJ, Ferguson BJ & Livsey D. 2023. Calculation of Discharge in Estuaries: Current State and Advancements. Brisbane: Department of Environment and Science, Queensland Government.

November 2022

Contents

Background

The Great Barrier Reef Catchment Loads Monitoring Program (GBRCLMP) first started monitoring water quality within Queensland's waterways in 2006 as part of the Reef Water Quality Protection Plan 2003 (Reef Plan 2003). Reef Plan has since progressed through multiple iterations, and the GBRCLMP now sits within the Paddock to Reef (P2R) Integrated Monitoring, Modelling and Reporting Program (Reef Plan 2017). The P2R Program monitors water quality status and trends, and reports on progress towards the Reef 2050 Water Quality Improvement Plan (WQIP) targets, objectives and long-term outcomes through the Reef Water Quality Report Card. The WQIP water quality targets for sediments and nutrients are aimed at reducing the end-of-catchment load of these compounds being delivered to the Great Barrier Reef lagoon. The GBRCLMP is the source of sediment and nutrient loads data for the validation of Source Catchments models.

The accurate calculation of discharge from a waterway is a critical component in the calculation of nutrient and sediment loads. Generally, water quality monitoring for these compounds (either lab-based water sample analysis or *in situ* probe) will generate concentration data which is multiplied by discharge to calculate the load of a compound being exported past a point in a waterway. Catchment load is used to calculate the yield of a compound as a weight per unit of catchment area. As discharge is a key component in these calculations, any uncertainty in the discharge data will be propagated through to the final load or yield. To achieve the best possible discharge data integrity, the methods that the GBRCLMP utilise to determine discharge have been continually revised and improved since the program commenced.

This report will provide a summary of the discharge calculation methods that have been or are being used by the GBRCLMP, with the primary focus being on the calculation of discharge in estuaries. These methods include height-discharge relationships, the use of Acoustic Doppler Velocity Meters (ADVMs), velocity indexing techniques, and tidal filtering instances. The report will also introduce several methods that have been recently developed by the GBRCLMP, including a novel tidal filtering approach and a new application of a machine learning algorithm for the determination of discharge based on upstream sources. These new methods have allowed the GBRCLMP to refine the calculation of discharge, with improved interpolation through tidal minimums, and the capacity for calculation of discharge during prolonged data gaps in locally measured discharge.

Discharge Calculation Methods

Height Discharge Ratings

The standard method for determining river discharge is from a height-discharge relationship (rating curve). A rating curve is developed by plotting heights against either gauged or estimated flows, and a line of best fit drawn through them. Streams are gauged by physically measuring water depths and velocities across the stream at specific heights, either with a handheld current meter or acoustic doppler instrument (BoM 2019a). To derive an accurate rating curve the full range of stream heights should ideally be gauged. Where gaugings aren't practical, discharge for a stream can be estimated using Manning's equation, derived from the surveyed slope and channel cross section, and the estimated roughness of the stream bed.

A rating curve is not appropriate for all waterways – a rating curve ideally has a stable cross section, a straight stretch of channel with a clear, free flowing section downstream, and stable hydraulic conditions, where the height-discharge relationship is consistent and controlled by the roughness, slope and channel shape. Persistent or long-term changes (e.g. a change to the cross section from erosion or deposition) can be accounted for by re-gauging the waterway. However, anything downstream that causes variable or short-term backpressure can interfere with the relationship in an unpredictable manner. For example, a low bridge or other obstruction may catch debris, a downstream confluence (Figure 1) may cause flow to back up towards the station (Figure 2), or tidal influence.

Figure 1 Example of a downstream confluence. The pink marker shows a monitoring site on a tributary ~1.6 km upstream of its confluence with the main branch. The blue marker shows the moderate flood level of the main branch. The equivalent elevation contour is marked in

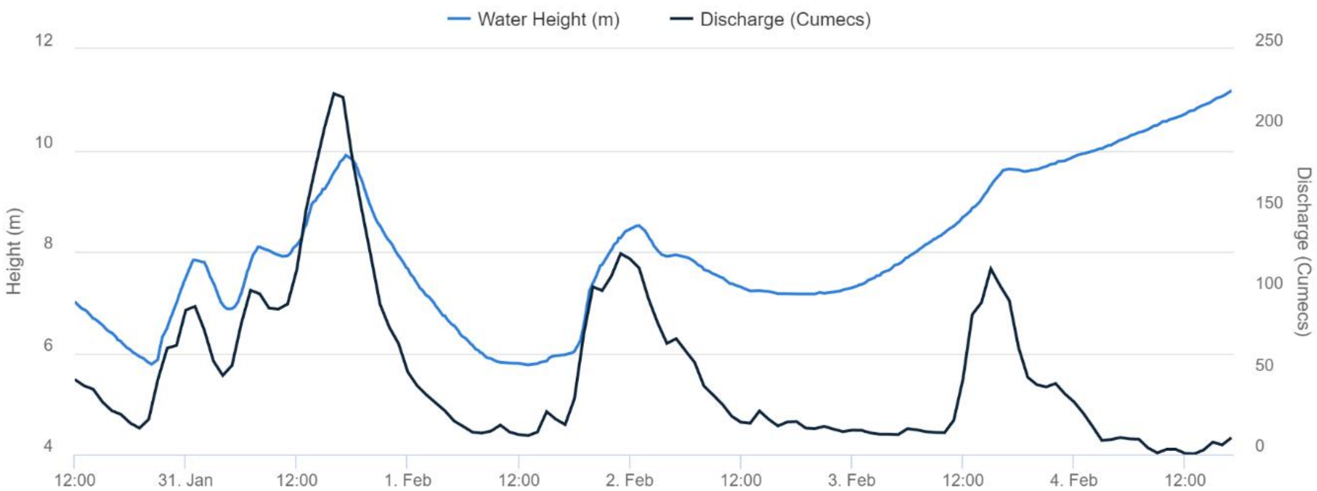

Figure 2 Backpressure from the confluence shown in Figure 1 causing disruption to the height – discharge relationship at a monitoring site

Figure 2 shows that between 12:00 on the 30th of January 2019 and 00:00 on the 2nd of February, the water level and discharge were rising and falling with each other. However, after 00:00 on the 2nd of February, the water level and discharge diverge, with water height increasing without an equivalent increase in discharge – this is caused by the water height of the larger waterway pushing back into the smaller tributary.

Velocity Monitoring and Indexing

To determine discharge at a site not suitable for a rating curve (such as a tidal system, or the site described above), it is necessary to directly measure water velocity in addition to height and area. There are two main methods that can be used for continual velocity measurement; submerged acoustic doppler instruments, such as an Acoustic Doppler Velocity Meter (ADVM), and surface velocity monitoring, such as particle image velocimetry (PIV), radar, and laser doppler velocimetry (BoM 2021). These methods each have their own benefits and drawbacks. The major considerations are generally applicability to specific waterways (due to channel profile/width, water clarity, etc.), cost, on-going maintenance and the associated risks to staff, and the risk of damage to the monitoring equipment. To meet the GBRCLMP's needs and logistical requirements, the GBRCLMP has adopted ADVMs to measure water velocity in real time at tidal sites since 2014. An ADVM is an acoustic doppler instrument, permanently installed *in situ*, oriented perpendicular to flow direction. The data produced by the ADVMs have been used to calculate GBRCLMP loads and yields as a major deliverable for the Paddock to Reef Integrated Monitoring and Reporting Program, under the Reef 2050 Water Quality Improvement Plan (Wallace et al. 2016).

High frequency sound bursts (pings) are utilised by acoustic doppler instruments to collect water velocity data from within the water column using paired acoustic transducers that emit and receive pings from different directions. These pings reflect off particles suspended in the water and echo back to the instrument. The water velocity is calculated based off the phase change (the time shift, measured in degrees of one wavelength) between two consecutive pulses with a known delay – the change in delay between the two pulses that are received by the instrument indicates how far the particle has moved between pulses. The time delay between the ping and the echo determines how far from the instrument each particle is. In order to correctly correlate pings and echoes, each ping consists of a series of coded pulses – this also assist with filtering out ambient noise (Teledyne RD Instruments, 2011). The orientation of the ADVM means that it can profile the velocity of water across the full width of the channel, with the data being collated into predefined sections, or bins, across the channel. At least 20 bins are designated across a channel (BoM 2019b).

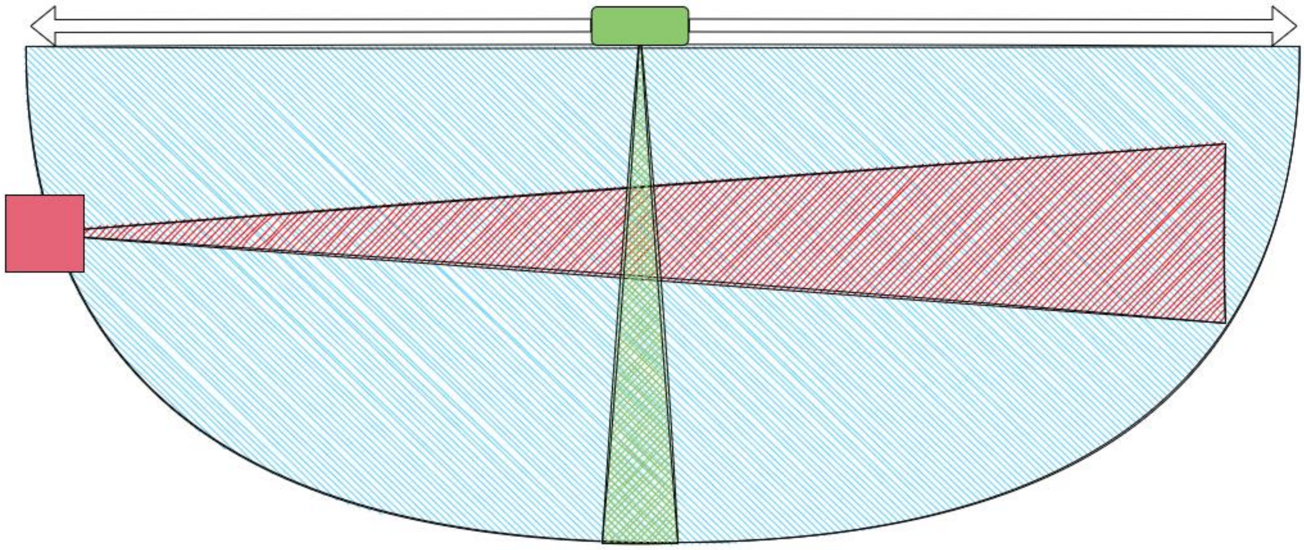

Figure 3 Velocity Indexing – the average velocity of the entire cross section (blue) is measured (gauged), using a boat mounted, downward facing ADCP (green), that is towed backwards and forwards across the waterway. This is paired with the average velocity recorded by the ADVM (red).

One of the limitations of an ADVM is that it only assesses a horizontal slice of the water column that is level with the ADVM – this slice is not necessarily representative of the entire water column. To correct for this, gaugings must be completed across a range of flow conditions. During gaugings, the ADVM will read once per minute while stream discharge is gauged (Figure 3), then gauged discharge is divided by a standard cross section to calculate mean velocity. The average gauged velocity is then regressed against the ADVM measured velocity at the time of gauging to compute a velocity-velocity relationship, or Velocity Index (VI). With sufficient gaugings, a VI can be developed and used to adjust the ADVM velocity to local conditions. Each site has its own unique VI that may vary over time.

Initially, Velocity Indexing relationships were derived using simple linear regression relationships at each ADVM site (BoM 2019b). However, a simple linear regression assumes a consistent velocity index relationship across all conditions – high flow gaugings have confirmed the relationship can vary between tide and event conditions.

To account for these additional factors, the raw channel aligned (x) velocity is decomposed using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) (Torres et al. 2011). CEEMDAN is used to decompose the raw x-velocity into an event velocity component (frequency greater than 24 hours), and two tidal velocity components (Tide A with a frequency of 12 hours, and Tide B with a frequency of 24 hours). When these factors are considered along with water height, cross channel (y) velocities and sufficient gaugings, a more representative relationship across the full range of flows can be determined. The actual factors used in the velocity index depends on the significance of factors, and the availability of gaugings (Figure 4).

Murray River at Bilyana

- 260 Gaugings over 3 years
- Measured flows ranged from -42.28 to 121.88m^3s^{-1}.
- Measured channel velocities ranged from -0.28 to 0.78 ms^{-1}.

Velocity Indexing relationship between the gauged velocities (y) and the H-ADCP velocities ($XVel$ and $YVel$) and water height (*Height*) in these measurements is described by:

$$CEEMDAN_{XVel} \rightarrow TideA + TideB + Event$$

$$Magnitude = Event * Height$$

$$y = 0.3672\,TideA - 0.4124 TideB + 0.1687 Event + 0.01557 Magnitude + 0.8048 YVel + 0.04482$$

(R^2 = 0.92)

Figure 4 Example of a CEEMDAN derived velocity indexing relationship

The use of tidal and event components as independent factors allows for high flow gaugings to better influence the upper stages of a velocity index, without impacting on the tidal ranges of the velocity index. The size of flow events (distinct from the event velocity) are taken into account with “Magnitude”, calculated by multiplying the event velocity (m/s) and river height (m). When high flow gaugings are not available, the “measured” inputs into the velocity indexing relationship can be estimated using methods similar to those traditionally used in estimating high flows for height-discharge relationships, or from combining upstream tributaries. Although these estimates can be suitable in the absence of legitimate gaugings, they might not be as reliable to use once relevant gaugings have been completed. Examples of high flow velocity estimates derived from gauged upstream inputs are shown as red dots in the charts in Figure 5 and Figure 6.

Figure 5 compares the raw and indexed velocities, where the VI regression is limited to low flows (event velocities < 0.4 m/s) and high flow estimates. As can be seen in Figure 5C, the residuals of the gauged velocities from 0.4 m/s to 0.9 m/s are generally improved by the application of this VI. Most of these gaugings are more recent, and can be seen in Figures 5A and 5B, with indexes >170.

Figure 6 shows the same comparison between raw and indexed velocities. However, the VI regression used in Figure 6 included all gauged flows, and excluded the high flow estimates. A marked improvement can be seen between Figure 5B and Figure 6B in the larger gauged events (Index >170). The exclusion of the high flow estimates does have an impact on the residuals of the high flow estimates. This is expected, as the currently gauged velocities peak at 0.9 m/s, whereas the peak velocities recorded in the channel exceed 2 m/s.

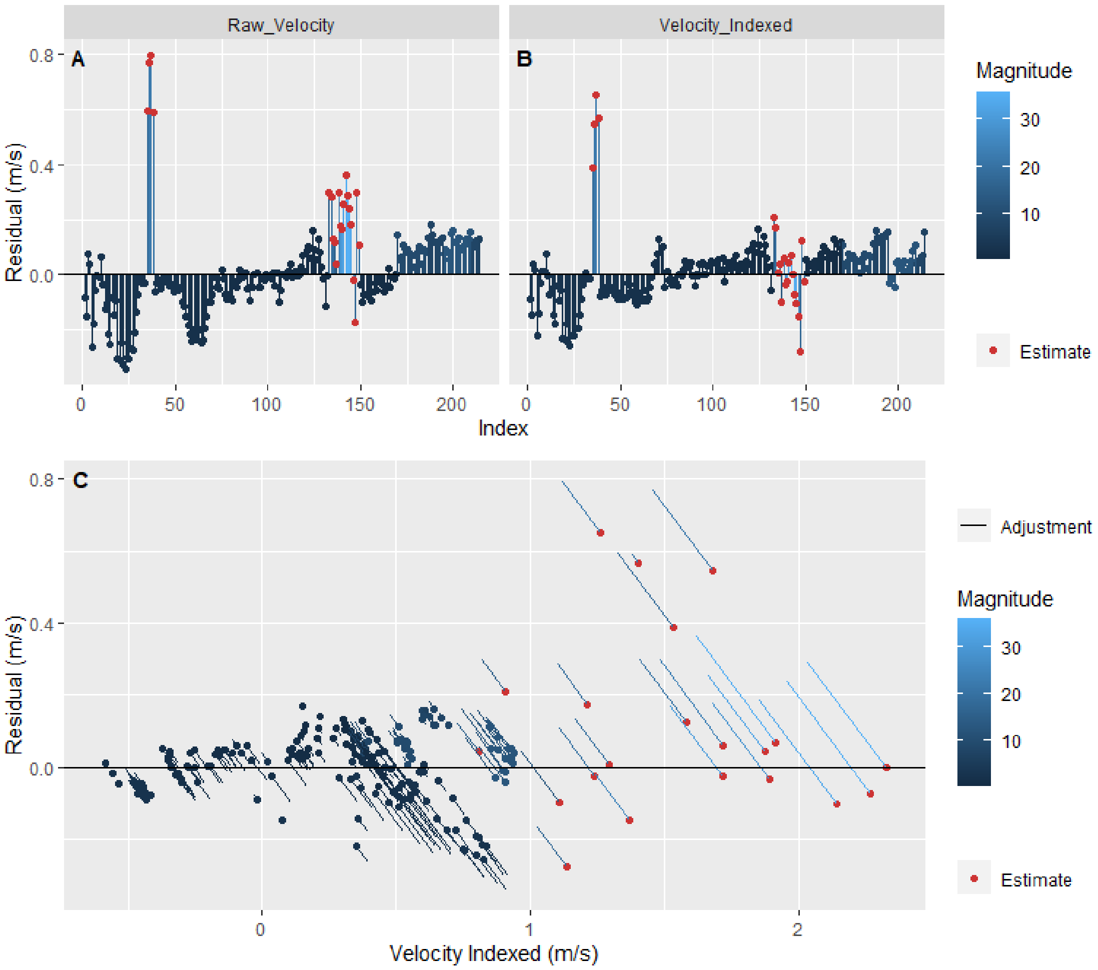

Figure 5 Regression trained to low flow gaugings (raw event velocity <0.4 m/s) and high flow estimates. Event gaugings are a lighter shade of blue.

Once sufficient high flow gaugings are collected, the high flow estimates can be excluded from velocity indexing calculations, as the gauged velocities are more reliable than the upstream estimates. The velocity index relationship did not substantially change for the gauged range once high flow gaugings were used in the regression. Thus, for situations where high flow gaugings are not logistically feasible or possible, estimated high flow velocities can be used in lieu of high flow gaugings. Figure 5 also shows that the high flow event relationship could be adjusted by high flow event gaugings independent of tidal low flow gaugings, as can be seen by the different shifts between the higher and lower magnitude gaugings.

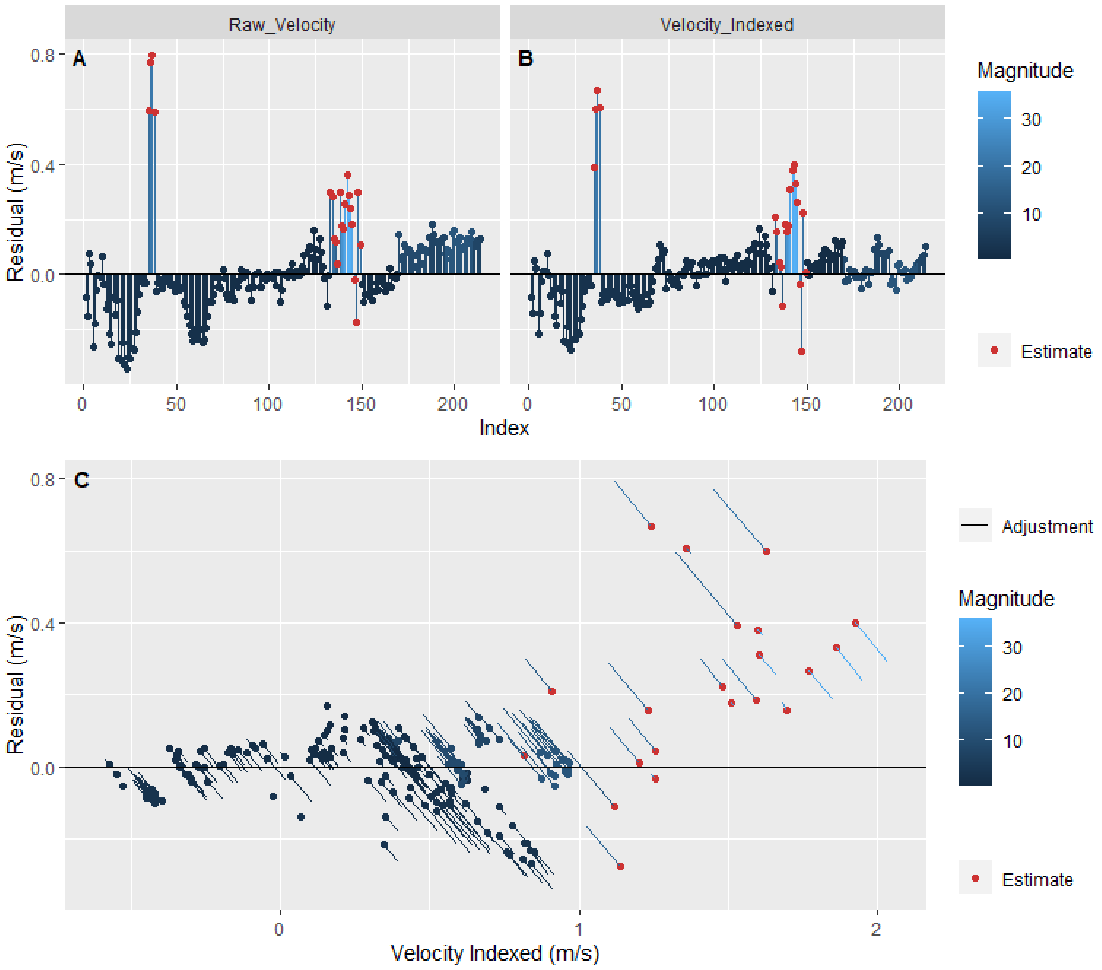

Figure 6 Regression trained only to gauged flows, with estimates excluded. Estimates are plotted in red for reference.

Once the ADVM raw velocity time series data has been velocity indexed (outputting a timeseries of "Calculated Velocity"), the cross-sectional area of the stream at each timestamp is determined. Cross sectional areas are calculated for the full range of heights and a height to area lookup table is created (called a stage-area rating, or a height-area rating). The height is then converted to area from the stage-area rating. For each timestamp, the calculated velocity (m/s) is multiplied by the cross-sectional area of the stream (m^2) to determine timeseries discharge (m^3/sec, or cumecs) (Levesque et al., 2012).

Tidal discharge data

Due to the constant fluctuations in discharge caused by the rise and fall of the tides, tidal discharge is generally only meaningful for loads as a daily average (Ruhl et al., 2005); however, the GBRCLMP requires flow data more frequent than daily to compute loads. To address this problem, flow data were detided using a low pass filter (initially the Godin filter was used for GBRCLMP). Filters such as the Godin filter or Butterworth filter are commonly used to remove a tidal signature from flow data; however, they each have their own strengths and weaknesses (Walters et al. 1982, Ruhl et al., 2005). The unfiltered tidal discharge data shown in Figure 7 is overlayed with the result of three different tidal filters. The Godin and Butterworth filters are pre-existing filters, and the ccInterp filter has been developed by Stephen Wallace (Appendix B).

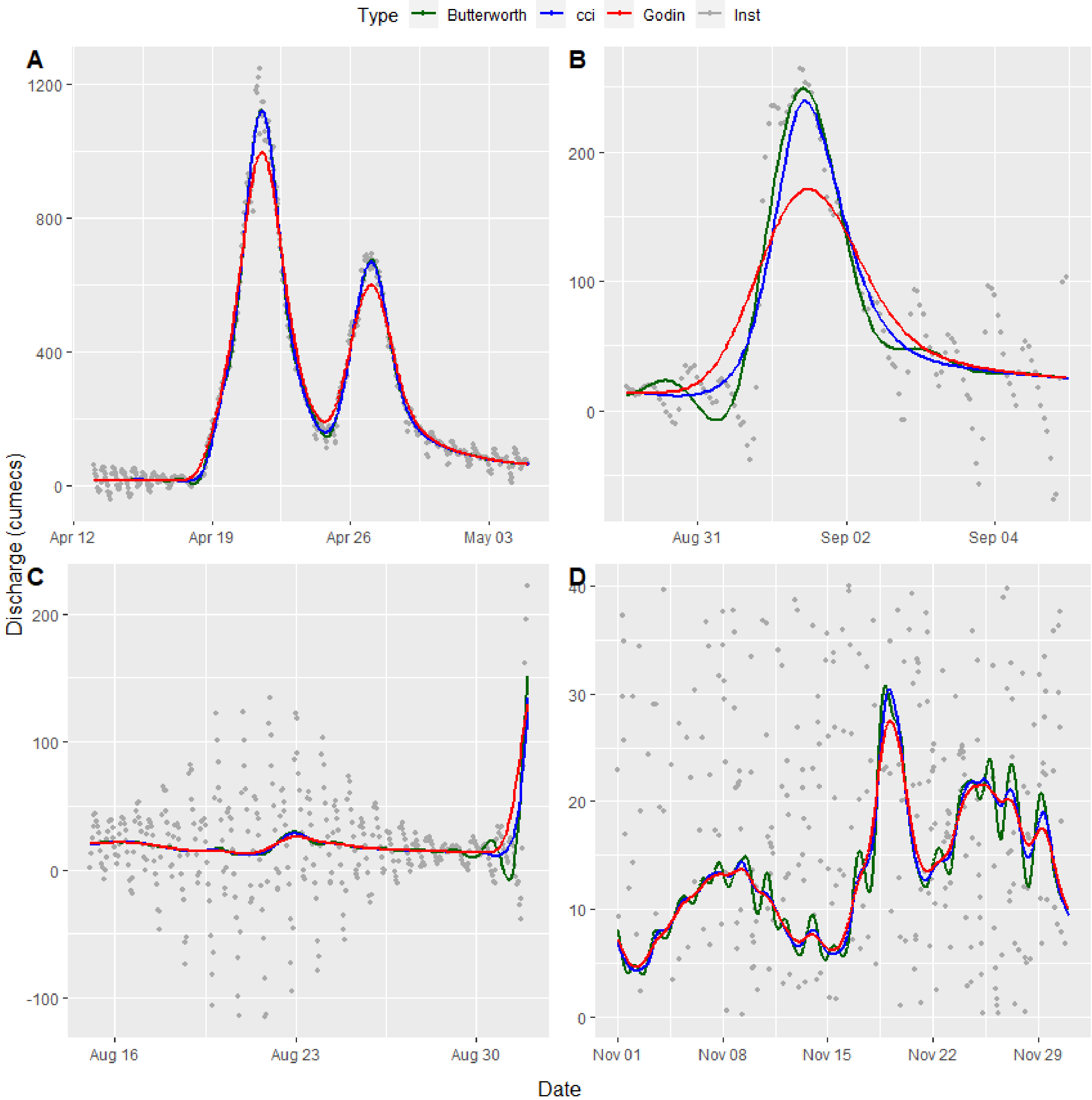

Figure 7 Examples of filters applied to tidal flow data

All filters perform relatively well for large, long duration flow events (Figure 7A), and they all remove small variations in tides well. However, shorter duration flow events (Figure 7B) are overly attenuated by the Godin filter. This attenuation can also be seen across very small events during low flows midway through Figure 7C. Since Godin is a smoothing filter, baseflows are detided well by the Godin filter – the Butterworth is not as effective at detiding baseflows, with some residual tidal signal remaining after processing (Figure 7D). The Butterworth filter is also prone to downward drops before an event (ringing), in response to the sudden rise. The Godin filter does not have the same degree of ringing; however, the smoothing effect can result in Godin filtered discharge apparently rising before the start of the event (Figure 7B,C). The Butterworth filter is heavily impacted by ringing during low flows as well as before sharp rises, as can be seen by the oscillations in Figure 7D.

The ccInterp filter is a best-of-both-worlds filter, where it resembles Godin during low flows (Figure 7C,D), and Butterworth during high flows (Figure 7A,B). The nature of the ccInterp filter is that it dampens rather than accentuates noise unlike the Butterworth filter, while also being responsive at the beginning and peak of flow events.

There are several potential issues that arise from the use of ADVMs in the tidal zone. One of these is the fluctuating salinity over the tidal cycle – the salinity of the water will impact the ADVMs calculation of velocity and needs to be accounted for. This is done by pairing ADVMs with a conductivity sensor and re-configuring the ADVM throughout the tidal cycle as the salinity changes.

Another of the limitations of using ADVMs for velocity monitoring within the tidal zone is that the ADVM must be fully submerged to take a reading. It requires enough water above and below the instrument to allow for the spread of the acoustic beam, as the water surface and riverbed will introduce interference to the return signal. The ADVM units that the GBRCLMP use have a spread of 20:1 (300 kHz model) or 24:1 (600 kHz model), meaning that for every 20 or 24 meters of travel respectively, the height and width of the beam is ±1 m. The ADVM should be mounted high enough to avoid interference from the riverbed, while still being as low as possible to avoid excessive surface interference.

When using the velocity indexing method, the ADVM does not need to measure the full width of the channel. In most circumstances, the velocity of a subsection of the stream can be used as a representative raw input for the velocity index relationship, and the velocity index relationship will correct the partial section to the gauged discharge of the entire stream. In situations where it is not possible to avoid interference in the ADVM signal, it is possible to determine more meaningful velocity index on a subsection of the stream.

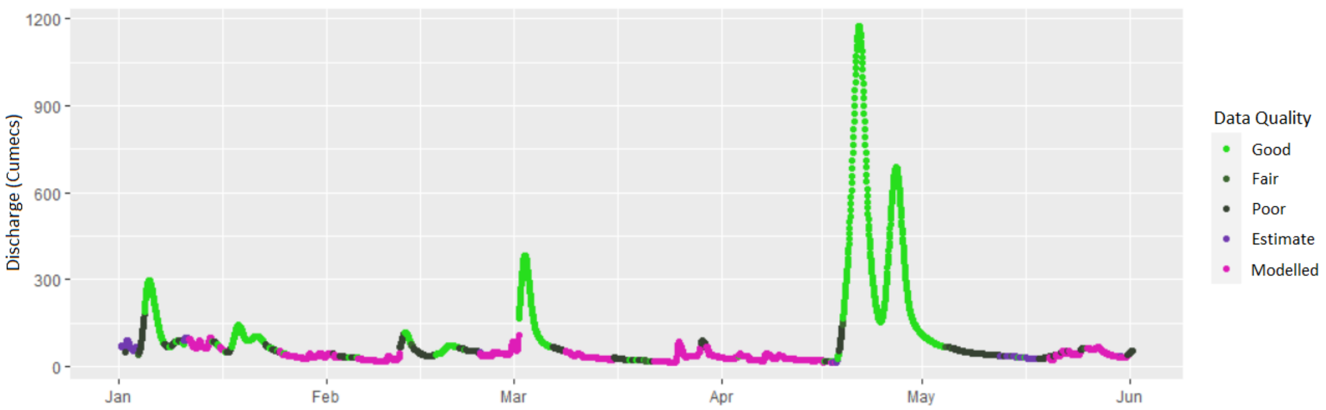

Figure 8 Example of the merged discharge data, highlighting the use of modelled data through low flow periods.

In most cases in estuaries, the ADVM will be mounted at or below the peak of the neap tide. When the tide drops below the instrument, long gaps of up to 12 hours can routinely occur. When this occurs, the data before and after the gap will be coded as QC60 (Estimate), denoting the gap (Figure 8). The filtered tidal discharge data downgrades the quality to the lowest quality that was used to produce that filtered data. Downgrading quality for long gaps has the added benefit of also flagging any potential discrepancies arising from (for example) salt-wedge impacts or non-stabilised internal temperature correction in the ADVM. Velocities below the instrument are interpolated linearly between the last outgoing tide and the next incoming tide. Anything over 12 hours is considered a gap and is not interpolated, although ideally, low tide gaps should not exceed 6 hours at a time for good quality data. Gaps in timeseries discharge data can introduce and accentuate cyclical noise into the signal when filtered, and produce a cumulative bias due to false assumptions about flow conditions between the last outgoing tide and the next incoming tide up to 6 hours later.

During low flow conditions, tidal variability very often exceeds the event magnitude. The result is either a high degree of ringing, or nonsensical filter output (such as negative discharges); therefore, low precision modelled flow is inserted in place of ADVM discharge (Figure 8). Modelled daily interval flow data is supplied on an annual basis by the Water Quality Modelling team, in the Department of Environment and Science (DES) for this purpose (Waters et al. 2014). Although not particularly suitable for event flows, during baseflow periods the lack of fine scale temporal detail in the modelled data is not of a concern for load calculation purposes. This is due to low variability of discharge, and the infrequent concentration data collection (i.e. monthly). However, there are methods available that will produce better results than leaving gaps in the record and interpolating linearly. One such method is an application of Empirical Dynamic Modelling, a machine learning approach, as described below.

rEDM machine learning for infilling gaps

Empirical Dynamic Modelling (EDM) is a machine learning approach that can predict real time tidal or event discharge that can produce a continuous timeseries of discharge data without gaps. When a tidal filter is applied to the discharge data with EDM inserted into gaps where water is below the ADVM or the ADVM wasn't recording, it will produce a better quality filtered timeseries, with less ringing. When the interpolated gaps between dry periods are interpolated with a better quality EDM discharge, the result is less ringing, more realistic discharge data, with less bias introduced than a straight line interpolation, which therefore doesn't have to be excluded to only be replaced with low precision modelled data.

Empirical dynamic modelling infers spatially and temporally nonlinear relationships between multiple timeseries of data in a training set, and then apply these same relationships to predict a past or present timeseries. The absence of defined parametric equations in this approach provides a flexible framework that can adapt to a wide range of varying circumstances within similar subsets of data. A set of indicative predictors are trained to a target discharge, either tidally filtered or unfiltered discharge. When the derivative of the local height is used as a predictor (change in height over time), EDM can be used to calculate tidal discharge in real time. To predict the underlying flow signal without tide, the height derivative is simply omitted as a predictor.

The EDM tools are accessed through an R Package called rEDM, and its use is detailed in Appendix A.

Summary

The use of ADVMs for tidal sites and those with variable backpressure have been an important addition to the GBRCLMP expansion into end-of-system monitoring. Where sites have a good record of gaugings in a range of flow conditions, a good quality velocity index can be calculated for that site. The velocity index for a tidal site will use a decomposition of the ADVM velocity to assess tide and event components separately, which improves the fit during all flow situations without the assumption that the velocity index is the same during event flows and baseflows.

Different tidal filters have been evaluated for detiding discharge data from GBRCLMP sites, and it has been determined that a novel detiding method called ccInterpFilter performed better than the currently available filters. When there is a weak freshwater discharge signal in a strongly tidal stream, as typically occurs during baseflow periods, the resultant filtered discharge is of a lower quality due to logistical and technological limitations. The low quality filtered discharge is replaced with a low precision modelled input (daily timestep) instead.

The degree of uncertainty in calculated discharge data can be improved with the insertion of more meaningful interpolations, infilling gaps and replacing low quality values in the ADVM dataset. The rEDM machine learning method for predicting tidal discharge is an appropriate method for determining these interpolated values, especially when compared with highly biased and presumptive linear interpolations. The rEDM method performs well for events that fall within the measured range of the ADVM, and can accurately predict gaps due to water level falling below the ADVM or instrument malfunction.

References

BoM (2019a) National Industry Guidelines for hydrometric monitoring Part 4: Gauging (Stationary Velocity-Area Method). NI GL 100.04–2019, Bureau of Meteorology, Standards. Available at http://www.bom.gov.au/water/standards/documents/NI_GL_100_04-2019.pdf [Verified 22/11/2022]

BoM (2019b) National Industry Guidelines for hydrometric monitoring Part 9: Application of In-situ Point Acoustic Doppler Velocity Meters for Determining Velocity in Open Channels. NI GL 100.09–2019, Bureau of Meteorology, Standards. Available at http://www.bom.gov.au/water/standards/documents/NI_GL_100_09-2019.pdf [Verified 22/11/2022]

BoM (2021) National Industry Guidelines for hydrometric monitoring Part 11: Application of Surface Velocity Methods for Velocity and Open Channel Discharge Measurements. NI GL 100.11–2021, Bureau of Meteorology, Standards. Available at http://www.bom.gov.au/water/standards/documents/NI_GL_100.11-2021.pdf [Verified 22/11/2022]

Levesque, V.A., and Oberg, K.A., 2012, Computing discharge using the index velocity method. USGS, Techniques and Methods. Available at https://pubs.usgs.gov/tm/3a23/

Reef Plan (2003) Reef Water Quality Protection Plan: For catchments adjacent to the Great Barrier Reef World Heritage Area. ISBN 0 642549 51 6, Queensland Government. Available at https://www.reefplan.qld.gov.au/__data/assets/pdf_file/0013/46120/reefplan-2003.pdf [Verified 22/11/2022]

Reef Plan (2017) Paddock to Reef Integrated Monitoring, Modelling and Reporting Program 2017-2022: Summary. Queensland Government. Available at https://www.reefplan.qld.gov.au/__data/assets/pdf_file/0026/207674/ptr-integrated-monitoring-modelling-report-program-summary.pdf [Verified 22/11/2022]

Ruhl, C.A., and Simpson, M.R., 2005, Computation of discharge using the index-velocity method in tidally affected areas: U.S. Geological Survey Scientific Investigations Report 2005-5004. Available at https://pubs.usgs.gov/sir/2005/5004/sir20055004.pdf.

Torres, M.E., Colominas, M.A., Schlotthauer, G. and Flandrin, P. (2011). Complete ensemble empirical mode decomposition with adaptive noise. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. 4144-4147. DOI: 10.1109/ICASSP.2011.5947265.

Teledyne RD Instruments (2011) Acoustic Doppler Current Profiler Principles of Operation A Practical Primer. P/N 951-6069-00. Available at http://www.teledynemarine.com/Documents/Brand%20Support/RD%20INSTRUMENTS/Technical%20Resources/Manuals%20and%20Guides/General%20Interest/BBPRIME.pdf [Verified 23/03/2023]

Wallace, R., Huggins, R., King, O., Gardiner, R., Thomson, B., Orr, D.N., Ferguson, B., Taylor, C., Smith, R. A., Warne, M. St. J., Turner, R. D. R., Mann, R. M. 2016. Total suspended solids, nutrient and pesticide loads (2014–2015) for rivers that discharge to the Great Barrier Reef – Great Barrier Reef Catchment Loads Monitoring Program. Department of Science, Information Technology and Innovation. Brisbane.

Walters, R. A., and C. Heston, 1982: Removing tidal-period variations from time-series data using low-pass digital filters. J. Phys. Oceanogr., 12, 112–115, Available at https://doi.org/10.1175/1520-0485(1982)012<0112:RTPVFT>2.0.CO;2. [Verified 22/11/2022]

Waters, DK, Carroll, C, Ellis, R, Hateley, L, McCloskey, GL, Packett, R, Dougall, C, Fentie, (2014), Modelling reductions of pollutant loads due to improved management practices in the Great Barrier Reef catchments – Whole of GBR, Technical Report, Volume 1, Queensland Department of Natural Resources and Mines, Toowoomba, Queensland (ISBN: 978-1-7423-0999).

Appendix A

rEDM Derived Discharge

In the current study, the rEDM model was trained to velocity indexed, tidally filtered ADVM discharge data, with upstream discharge data, upstream rainfall data, and tidally filtered local height data used as model inputs. The site used for the initial trial was Mulgrave River at Deeral (MRD), a GBRCLMP site with a good history of gaugings (190 gaugings between 2015 and 2022), a well-established velocity indexing relationship (R^2 = 0.97), and an upstream DRDMW Gauging Station (Mulgrave River at Fisheries – MRF – approximately 44 km upstream, Gauging Station Number 111005A).

The predictive inputs into the model were:

- The upstream discharge, as measured at MRF by DRDMW
- Accumulated rainfall at MRF over the prior 6 hours, 24 hours, 72 hours, and 7 days. Each accumulated total excludes the periods shorter than it, to prevent stacking of the data.
- Tidally filtered water height from MRD. The filter used was a Combined, Cumulative Interpolation Filter (ccInterpFilter), with a filtering period of 24 hours, developed by Stephen Wallace for the purposes of tidal filtering – see Appendix B for more detail.

The target for rEDM for detided discharge was velocity indexed, tidally filtered (ccInterpFilter) discharge.

The predictors listed are the ones that were used in this trial of rEDM. There are a number of other inputs that have been considered, and either trialled and disregarded, or passed over for this trial.

- The daily modelled discharge generated by DES (Waters et al. 2014) was trialled as an input but did not improve the output of rEDM, and so was not included. Daily modelled data from DES is also only provided on a yearly basis, and is therefore at best only useful retrospectively and not able to be used as a predictor for real time discharge.
- There is a discharge monitoring station (Mulgrave River at Peets Bridge) that captures more of the MRD catchment (66%, compared to 45% captured at MRF). However, rEDM was tested in this case with a less representative input, as not all sites will have such a representative upstream gauging station.
- There is also rainfall data available from multiple other locations within the Mulgrave catchment that could be included.

The data used for this trial spanned assorted periods between December 2015 and June 2021. This period was selected as it had previously been processed, reviewed, and reported by the GBRCLMP. The training period selected was 01 January 2018 – 01 June 2019. This training period included a combination of low flows and events (including the largest event measured in this trial), as well as two distinct wet seasons.

A four-month subset of the training period from January 2018 to May 2018 (the calibration period) was used to prove the calibration process was effective, and the outputs of the calibration period are displayed in Figure 9. As can be seen (and as would be expected), the rEDM output aligns very well to the training data set (R^2 = 0.964, Figure 13A). The raw rEDM output has also been processed using a six-hour ccInterpFilter. This is not specifically required, but can be useful when a site is trained only to rainfall data, which can be more sporadic in nature than upstream flows, and so has been included here to demonstrate its influence (or lack thereof).

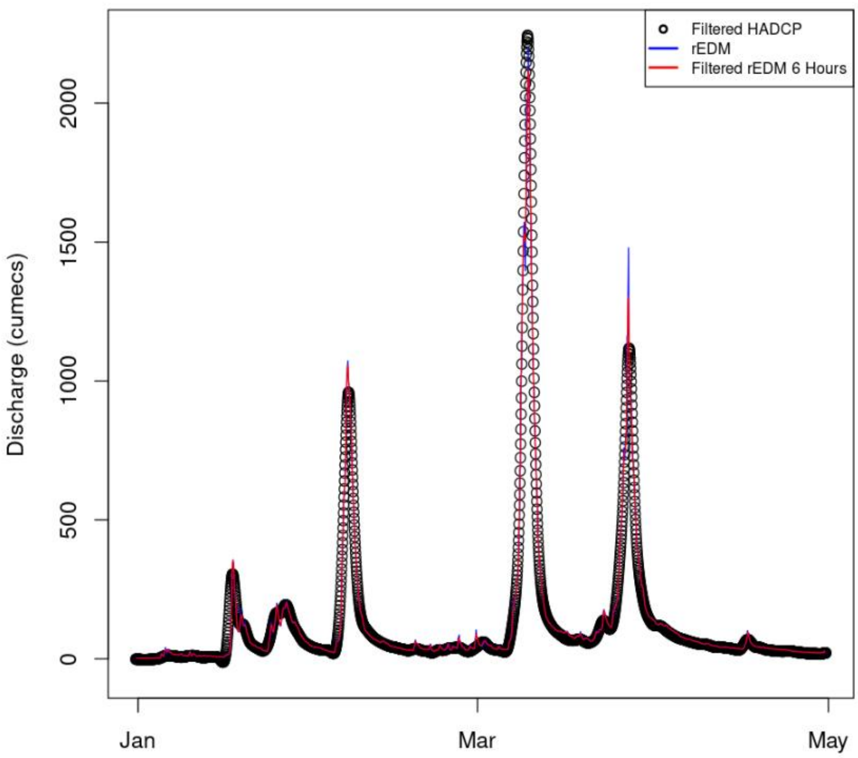

Figure 1 Raw and Filtered rEDM outputs at Mulgrave River at Deeral (01/01/2018 – 01/05/2018) from the calibration period, compared to the measured, tidally filtered ADVM data

The relationships that were determined by rEDM in the training set were then applied to several validation sets of varying characteristics to ensure that the relationships identified by the training set apply to data equally well from outside the training set. It was assumed that the model would be flexible enough to adapt to new data if outputs of the validation data trials fitted the measured discharge well.

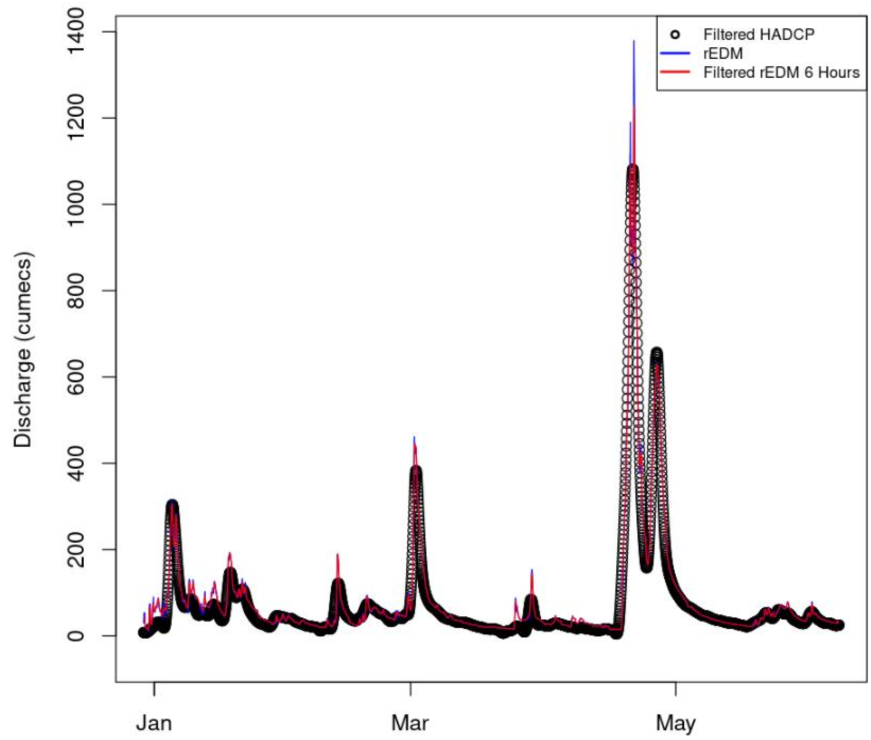

Figure 10 High flow validation at Mulgrave River at Deeral (30/12/2020 – 08/06/2021), comparing raw and filtered rEDM outputs to filtered ADVM discharge.

The first validation trial used data from a high flow period. The period selected for this validation dataset was from the 30 December 2020 to 08 June 2021. This period included the second largest measured event from MRD, as well as several other significant events. The outputs of the high flow validation dataset can be seen in Figure 10. The rEDM outputs are quite a close match to the observed discharge data ($R^2 = 0.941$, Figure 13B), and in some cases the rEDM output was more responsive than the filtered ADVM discharge, which is impacted by smoothing from the filter.

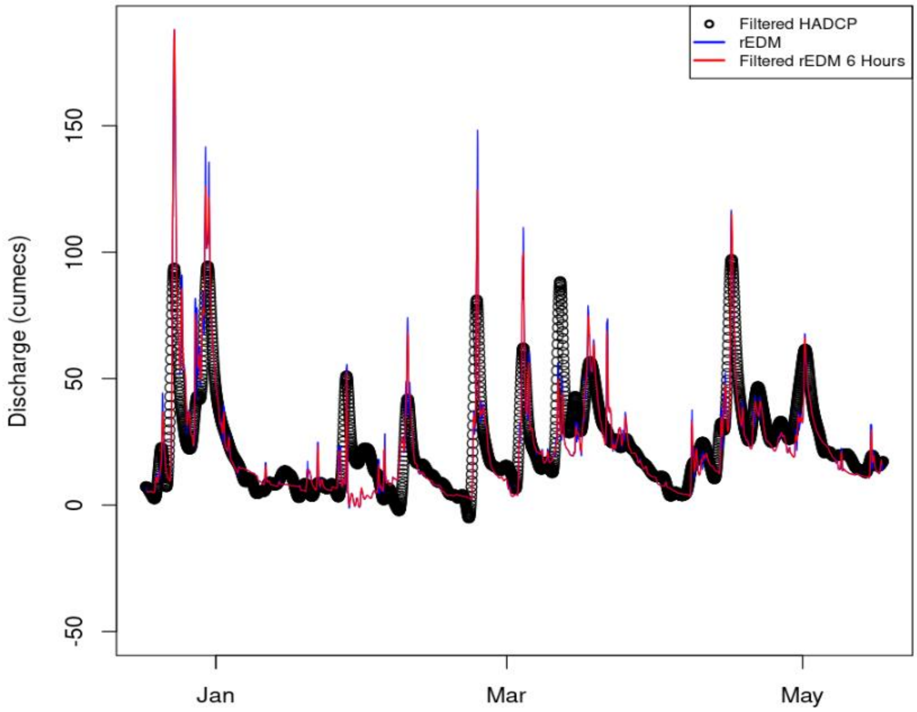

Figure 2 Low flow event validation at Mulgrave River at Deeral (18/12/2015 - 18/05/2016), comparing raw and filtered rEDM outputs to filtered ADVM discharge.

The second validation data set to be trialled was a period of low flow events. The time period that was selected for this validation period was from the 18th Dec 2015 until the 18th May 2016 due to the numerous small events that occurred between these dates. The smaller range of discharge also allowed for the examination of potential issues in extremely small events and base flows. The outputs of this validation dataset can be seen in Figure 11.

There are multiple sources of uncertainty with the velocity indexed, tidally filtered, ADVM data that are exacerbated at lower flows, resulting in a large portion of low flow data being replaced with daily modelled data from DES (Figure 8). The true detail of very low flows is often obscured by artifacts introduced by data gaps, and changes in water composition and density (e.g. salt wedge ingress), and the smoothing effects of the tidal filters – in effect, the signal to noise ratio. Filtering of the tidal data can potentially remove existing detail or introduce non-existent detail through cyclic noise arising from gaps in the record. As a result of this, although the rEDM outputs do not match the low flow events as recorded by the ADVM exactly, they could in fact be more accurately representing the effective discharge past the site. In the case of particularly short events (<12 hours), the rEDM output is likely to be more accurate, as tidal filters would treat them as a tidal signature and remove them.

As an indication of the issues arising from low flow uncertainty, Figure 12 shows the same time period and rEDM outputs as Figure 11, comparing against raw, unfiltered, ADVM data. This shows events derived from rEDM can be smaller than the tidal fluctuation. This makes it difficult to distinguish legitimate small events from tidal discharge alone and is a significant advantage of the rEDM method.

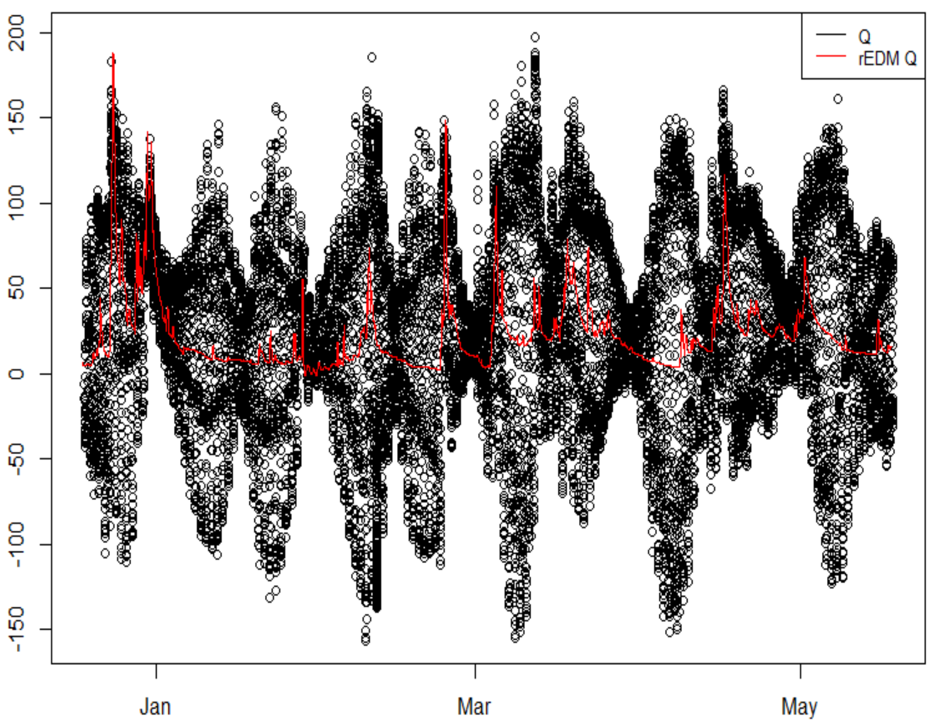

Figure 3 Unfiltered ADVM discharge data at Mulgrave River at Deeral (18/12/2015 - 18/05/2016), compared to the rEDM outputs from the low flow event validation period.

When the rEDM is trained to the full range of expected flows (peak discharge ~ 2300 Cumecs, Figure 9), it performs similarly well during calibration ($R^2 = 0.96$, slope = 0.86, Figure 13A) and validation ($R^2 = 0.94$, slope = 0.979, Figure 13B). However, when the rEDM is trained on a smaller event (peak discharge ~ 1100 Cumecs, Figure 10, Figure

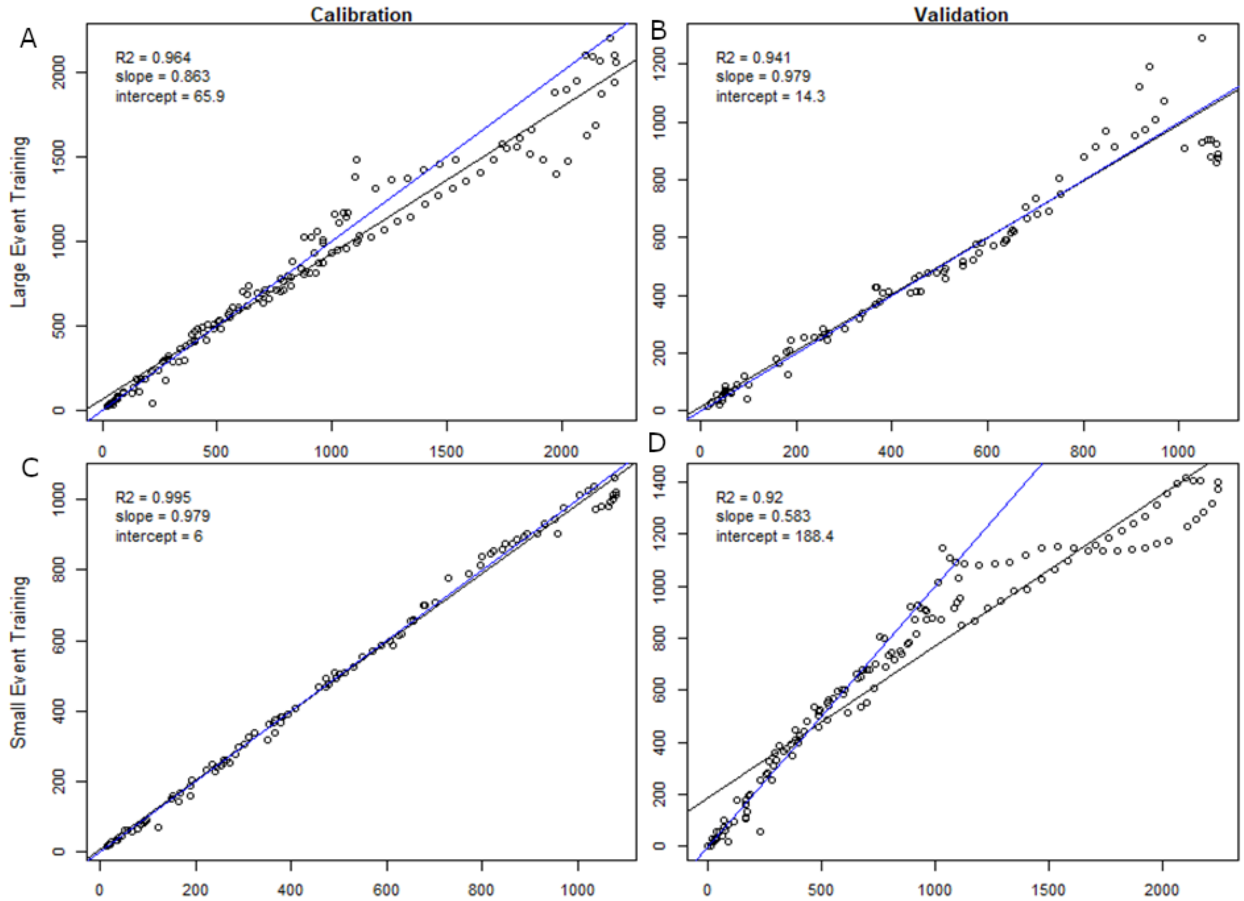

Figure 4 Comparing validation to calibration, X axis tidally filtered discharge (cumecs), Y axis rEDM derived discharge (cumecs).

13C) and then extrapolated to predict a larger event, the rEDM output is stunted to around the magnitude of the calibration period, in this case around 1100 cumecs. The validation shows rEDM performs well up until the max measured range, and then begins to plateau (R^2 = 0.92, slope = 0.58, Figure 13D). The comparison of Figure 13A and Figure 13D shows that rEDM is able to match the high flows much better when trained to high flows than when trained to low flows.

In order to evaluate the relative performance of the rEDM outputs of discharge and the daily modelled discharge data sourced from DES, the two datasets are plotted side by side against tidally filtered ADVM discharge data in Figure 14. As can be seen, the daily modelled data is acceptable at low flows, when compared with the filtered ADVM discharge. However, once the discharge moves away from baseflow (more than 100 cumecs in this example) the daily modelled discharge and the filtered ADVM discharge data begin to diverge.

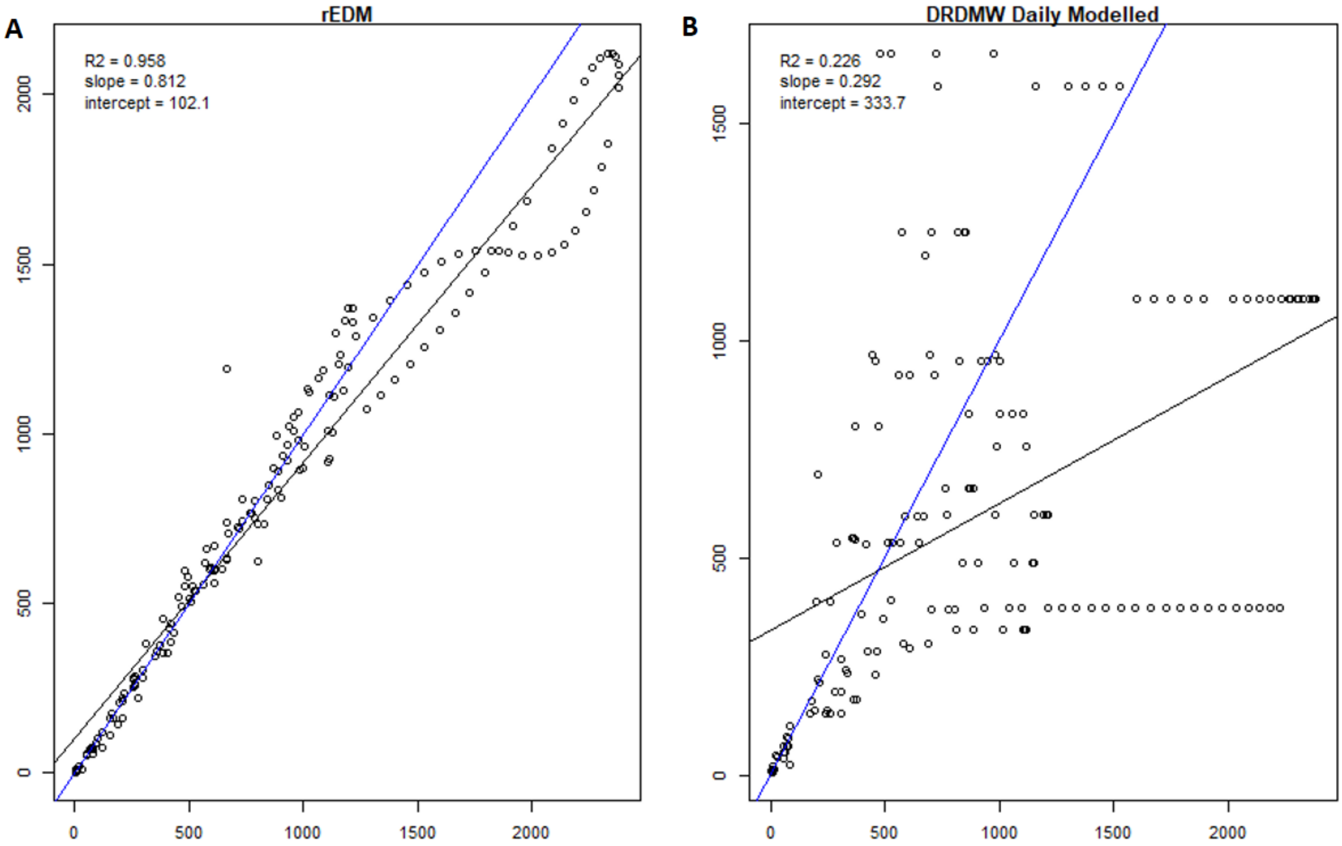

Figure 14 Model performance of A) rEDM and B) daily modelled discharge (y-axis) against tidally filtered ADVM discharge data (x-axis).

The poor correlation in Figure 14B is at least partially the result of comparing hourly data to a daily average. This is demonstrated in Figure 15 which shows similar characteristics when comparing daily ADVM to modelled, although there is still a significantly better performance of daily ADVM over daily DES modelled (as should be expected).

Limitations

As demonstrated in this paper, there is a relationship between discharge and height, upstream discharge, and rainfall in the catchment. The rEDM method has been demonstrated to be able to derive the relationships between these factors and then apply them effectively to additional data, providing a useful tool for the determination of discharge. However, it is very dependent upon the quality and availability of the inputs, as these will impact the reliability and stability of the derived model as well as its ability to be applied to new data. Large changes in catchment dynamics could have an impact on the effectiveness of the model. It also completely relies on previously measured, velocity indexed discharge data to be used as a training set. For a highly representative velocity index, a large event would need to have been both measured and gauged in the waterway that is to be modelled.

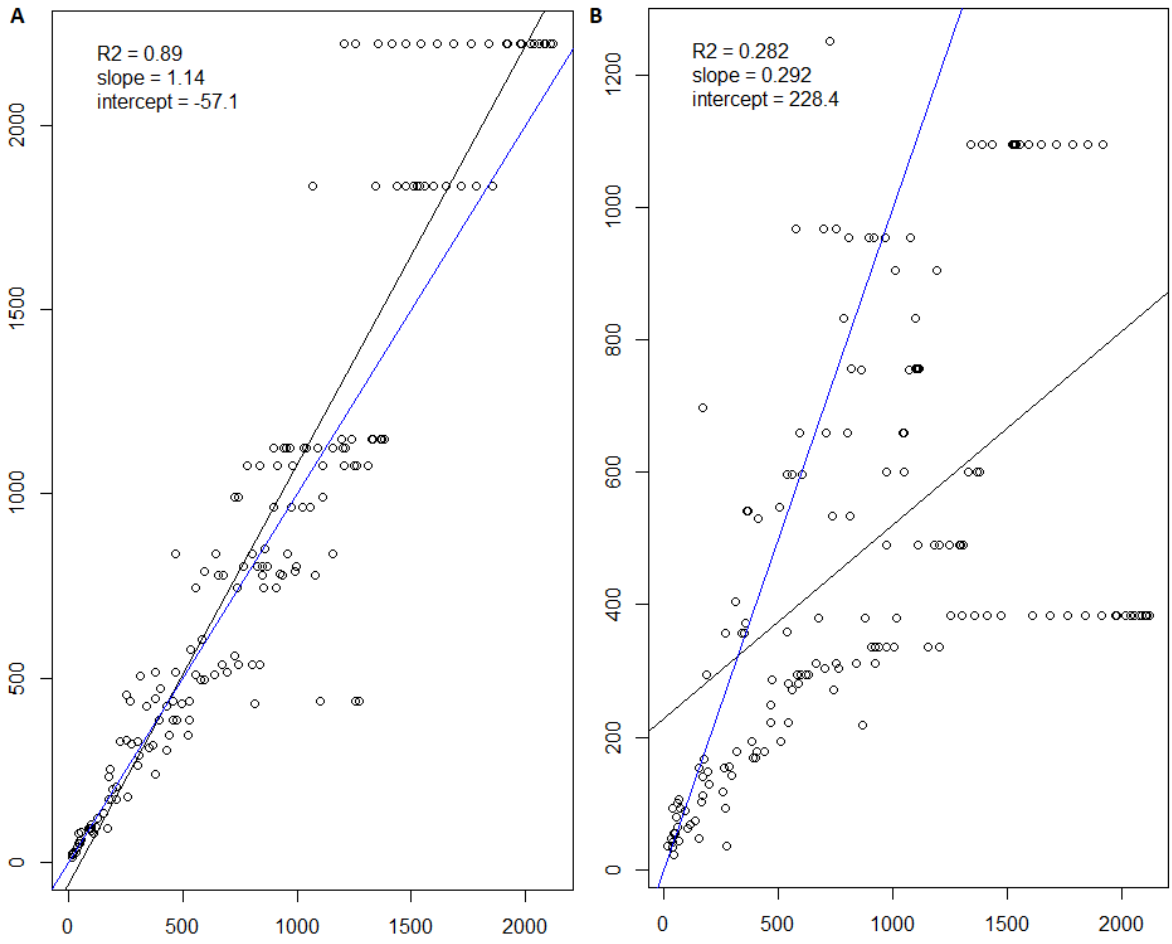

Figure 15 Comparison of A) daily mean discharge from ADVM and B) daily modelled discharge (y-axis) against rEDM discharge (x-axis)

In this trial, a period of 18 months has been used as a training dataset, with an additional 12 months of data used for validation. This trial has also restricted the input data to a less representative data set in order to test the limits of the rEDM method. With a training set that included the largest event recorded at the site (and despite restrictions placed on the predictive inputs), the rEDM method produced reliable outputs for a range of conditions that equalled or exceeded the accuracy of alternative methods. Additional improvements in the reliability of the output could be sought by increasing the representivity or number of the predictive inputs or by increasing the length of the training data set. However, there is a risk of over-fitting the model to the training data – if this were to occur, the model could lose the flexibility to respond to a wide variety of upstream conditions. In some instances, it may be necessary to use multiple upstream rainfall and discharge inputs, especially where there are multiple significant tributaries feeding into the one system, or larger geographical areas that may have spatially variable rainfall.

One major limitation of rEDM is that it is not particularly suitable for the extrapolation of discharge in events that are larger than those in the training set, as the models tend to stunt the peaks of events outside training parameters. If a larger event does occur, assuming there is confidence in the velocity index extrapolation, the rEDM could be recalibrated on the larger event.

Appendix B

Combined Cumulative Interpolation Filter

The author has developed a method of filtering tidal signals from time series data that can be used on both height and discharge data, called a Combined Cumulative Interpolation Filter (ccInterpFilter). This method utilises Monotone Interpolation, through the function *spinterp*, part of the *pracma* R package.

https://www.rdocumentation.org/packages/pracma/versions/1.9.9/topics/spinterp

The ccInterpFilter (period of *n* hours) takes all data in the timeseries and offsets it to ensure that there aren't any negative values. It then accumulates those data, creating a monotonic (constantly increasing) data set. The filter generates *n* timeseries of cumulative data, with a cumulative period of *n* hours centred on the timestamp, and an offset of 1 hour from the previous timeseries.

Time series generation working example:

> A filter period of 24 hours on a discharge dataset from 00:00 on the 1st Jan until 23:00 on the 31st Jan will generate 24 separate timeseries of daily cumulative data, each with 30 timesteps. The first timeseries will be daily cumulative totals from 00:00 – 00:00+1 day the second timeseries will be daily cumulative totals from 01:00 – 01:00+1 day up to the 24th timeseries, which will be daily cumulative totals from 23:00 – 23:00+1 day.

Each of these timeseries is then run through the monotone interpolation (*spinterp*), returning hourly values. The interpolated timeseries are averaged, returning an hourly accumulated interpolated timeseries. The derivative of this accumulated timeseries is then determined (to de-accumulate the timeseries), and the offset is removed. An example of the filtering can be seen in Figure 16. The R script required to process the data can be found in Figure 17.

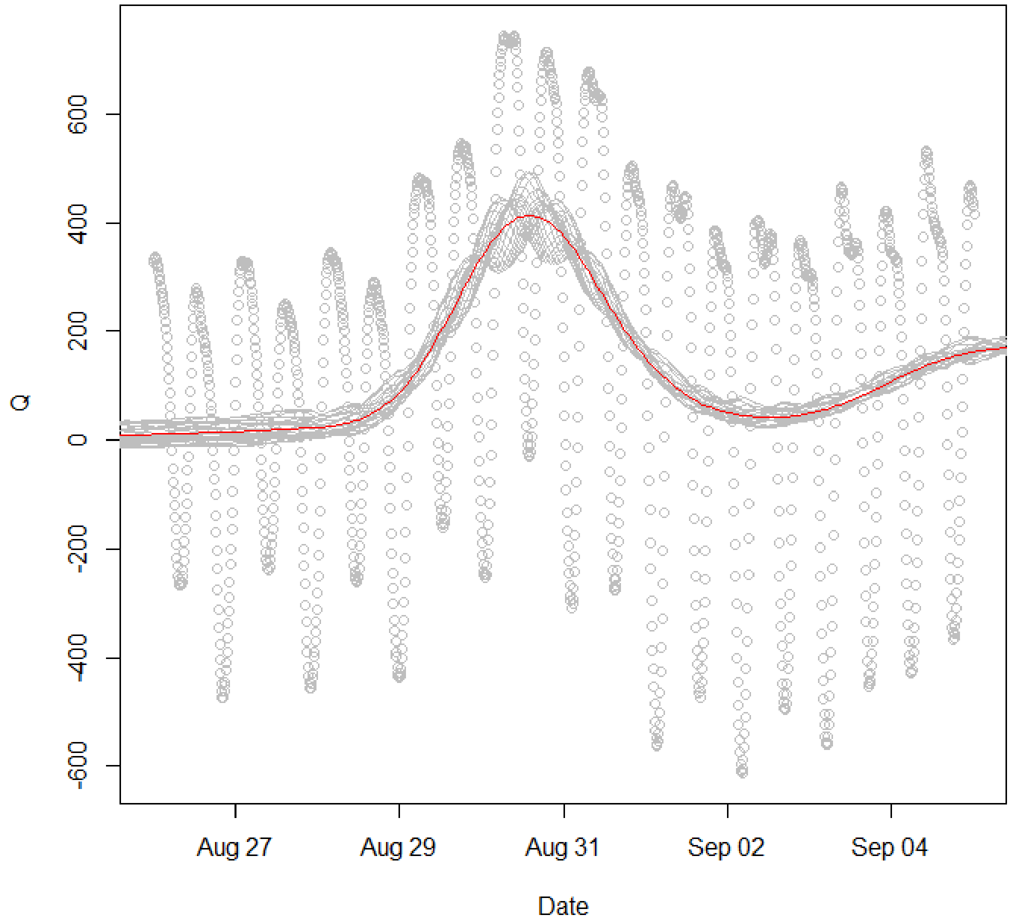

Figure 16 ccInterpFilter components visualised, with a 24 hour filtering period.

```
ccInterpFilter <- function(ts, hours = 24, type="spinterp")
{
  roundedtime <- round(ts[1,1] , units="hours")+1*60*60  # Trim times up, won't use part hours
  f.round <- approxfun(ts[,1], ts[,2])
  startvalue <- f.round(roundedtime)
  df <- data.frame(t = roundedtime, startvalue )
  colnames(df) <- colnames(ts)
  ts <- rbind(df, ts)

  # converts time series to daily, and cumulates, interpolates, and then gets the derivative.
  # does this for each hour, i.e. for 24 hour averaging it will do it 24 times offset an hour each time
  # the result is the average of all hours interpolated.
  # ts = dataframe of posixct time, and a value.

  spinterpData <- 0 #initialize
  for(i in 0:(hours-1))  #Loop through all daily interpolations
  {
    # for each 24 hour interpolation, increment the offset by an hour each time
    daily <- changeInterval(ts, Interval=hours*60, offset=i*60)
    offset <- min(daily$FMean)
    daily$FMean <- daily$FMean - offset

    daily$FMean[daily$FMean < 0] <- 0
    daily[is.na(daily)] <- 0
    numdate <- as.numeric(as.POSIXct(daily$Date, format=format)) /60/60/24    # convert to numeric
    spinterpSegment <- spinterpConvert(numdate, daily$FMean, type=type)    # create hourly spinterp data ( interpolation
of cumulative daily discharge )
    spinterpSegment$Data <- spinterpSegment$Data+offset    # remove offset again
    if (i>0){# after first row
      pad <- nrow(spinterpData) - length(c(rep(NA, i), head(spinterpSegment$Data, -i)) )
      if ( pad > 0)
      { # pad with na's
        spinterpData <- cbind(spinterpData, i=c(rep(NA, i), head(spinterpSegment$Data, -i), rep(NA, pad)   ))
      }else{
        spinterpData <- cbind(spinterpData, i=c(rep(NA, i), head(spinterpSegment$Data, -i)))
      }
    }else{# first row
      spinterpData <- spinterpSegment
    }
  }
  setcolnames <- paste("Col", 0:(hours-1), sep = "")
  setcolnames <- c("Date", setcolnames)

  colnames(spinterpData) <- setcolnames
  spinterpData <- na.trim(spinterpData)  # trim na's
  spinterpData <- cbind(spinterpData, avg = rowMeans(spinterpData[-1]))

  spinterpData$Date <- as.POSIXct(spinterpData$Date*60*60*24, origin="1970-01-01")
  return(spinterpData)
}
```

Figure 17 ccInterp filter script - R programming language