Process for Calculation Discharge Values

Documentation for the Add New Event R script

Margot E. Vore, MSc

June 2021

The Add New Event R script is designed to perform all steps of the discharge calculation process associated with automated salt dilution dump events and is currently running automatically once a week through the Hakai Goose server. The script itself can be found in the projects GitHub repository ([https://github.com/ voremargot/AutoSaltDilution](https://github.com/%20voremargot/AutoSaltDilution)). This document contains detailed information on the process executed by the script as well as the thresholds and constraints chosen throughout the calculation process. See figure 1 for a flow chart of the script’s workflow.

Raw data…………………………………………………………………………………………………………………………………………………1

Data selection and prep…………………………………………………………………………………………………………………………..1

Exploration of EC waves…………………………………………………………………………………………………………………………..2

Summerizing stage data…………………………………………………………………………………………………………………………..4

Create excel sheet for Google Drive…………………………………………………………………………………………………………5

Calculating discharge and uncertainty of an event…………………………………………………………………………………..5

Mixing……………………………………………………………………………………………………………………………………………………..6

Updating the database…………………………………………………………………………………………………………………………….7

CF Trends……………..………………………………………………………………………………………………………………………………….8

RECORD OF UPDATES MADE TO THIS DOCUMENT:

| Name | Date | Updates Made |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

1. Raw data

Raw data from autosalt dump events is automatically sent and collated in the Hecate Hakai server. Files in the server include: (1) one csv file per dump event containing the electrical conductivity (EC) probe outputs, (2) a csv file containing a full record of stage data from all dump events, (3) a dump event summary page containing metadata of each event, and (4) data from the backup sensor system which collects data when system communication errors occur. EC data are collected at one second time intervals while stage data are collected at five second intervals.

2. Data selection and prep

To determine what events have not been processed, the code compares records in the database to the event summary sheet from the Hakai server. An event is considered to be new if there is no record of it in the database. For each new event, prior to analysis, we determine if there are enough valid calibration factor (CF) values that can be used for discharge calculations. For a calibration factor to be used it must meet the following three criteria: (1) it must be within the barrel period (i.e., time period between fills of the solution reservoir) that corresponds to the dump event, (2) the stream temperature during the CF measurement must be within ±5°C of the stream temperature during the discharge event and (3) there can be no flags indicating the CF value is unusually high (>2.9\*10-6) or low (<2.2\*10-6). If there are less than four calibration trials that meet this criterion, the event will not be processed until more CF values are collected. See section 9 for more information on how CF values are chosen.

Once we confirm there is a sufficient number of CF values for calculating discharge, the raw EC data file associated with the new event is opened and data used for analysis are chosen. Depending on the types of EC sensors in the stream, these data may represent both temperature and non-temperature corrected values, with sensors that have only one column of raw data representing temperature corrected values. For analysis, columns that are temperature corrected are selected and all other data columns are excluded. If an event file is empty, this likely signals a connection issue between the autosalt dump system and the downstream sensors, which directs us to check for valid data in the backup EC sensor file. If this scenario occurs, the code will produce the flag, *AD*, indicating to the user that the main event file was empty. A thirty second moving average is used to smooth all chosen data to remove any anomalous values.

We also select which sensors are active during the new event. The sensor table in the database contains a record of all sensors deployed at each research location, highlighting their activation and deactivation dates. The sensors that are active during an event are those whose activation and deactivation dates contain the date of the dump event being analyzed.

3. Exploration of EC waves

Once the proper data are selected and prepped, the code determines the start and stop times of EC waves that occur in the record. An EC wave occurs as the dumped salt solution flows downstream past the sensors, increasing the electrical conductivity of the water for a short period of time. There is one EC timeseries per sensor.

To find the start time of the wave, we determine where EC values become significantly different from the background EC values (ECb) (i.e., the EC values that have not been influenced by the salt solution). In the first iteration, starting ECb (SECb) is defined as an array containing the first 45 seconds of data, of which the mean and standard deviation (*S*) are calculated. The minimum *S* value is 0.05. The start of the EC wave is defined as the first value which exceeds the mean SECb by more than three standard deviations.

*(1)*

Three standard deviations from the mean encompasses 99% of data in SECb. We determine that any EC value greater than three standard deviations is significantly different from SECb and is assumed to be influenced by the downstream flow of the salt solution signifying those values are part of the salt wave.

However, as allude to, this process for determining the EC wave’s start time is iterative due to the variability in the salt wave’s location in the record. We place a 400 second threshold between the start time and maximum EC value of the wave to prevent values from being incorrectly identified as part of the EC wave rather than SECb. This means that if the calculated starting EC value is more than 400 seconds away from the maximum EC value, the code will recreate SECb by adding 15 seconds of data to the previously used SECb array and recalculate the starting time. This process continues until the start time is within 400 seconds of the maximum EC value or the end of the record is reached.

The ending time of the EC wave is found in a similar manner. We subset the data from the maximum EC value (which is assumed to be the end of the rising limb) to the end of the record and take a 20 second moving average of the subset to determine where EC values return to the range of SECb.

*(2)*

Finding the end time of the wave is an iterative process based on the duration of the wave, with the duration being the number of seconds between the starting and ending time of the EC wave. We set a wave duration threshold of 1000 seconds to ensure that we are not incorporating an excess of tail values in the EC wave. Starting with *x*=4, we determine an EC value for the ending point. If the ending time is not within 1000 seconds of the starting point, we rerun the calculation with *x*=*x*+1 (the *x* value’s threshold is 70). In some scenarios, the background EC values do not return to pre-dump levels after the dump event. However, with the iterative increase in the standard deviation multiplier, we are able to find the waves ending point if it is within, at a minimum, 3.5 µs/cm of (this value increases as *S* increases).

The code also identifies trends in ECb values of each EC wave. A linear regression is run with all data that are not classified as part of the EC wave (i.e., all ECb values). If the resulting slope is greater than [less than] 2.5\*10-4 [-2.5\*10-4] and the R2 value is greater than 0.75, ECb is rising [falling]. If both these conditions do not hold true than ECb is constant. If every wave in the event has been classified as having the same trend, then that trend is representative of the whole event. If different sensors show different ECb trends, then the overall ECb trend is mixed.

If a rising or falling trend is detected in the salt wave’s ECb values, the end time of the salt wave will be recalculated by adjusting the SECb value to account for the increasing/decreasing trend. From the original calculated end time, we determine what the expected base line value is using the linear regression equation. We then take a 20 second moving average of the data from the maximum EC value to the end of the record and find where the EC values fall within 3 standard deviations of the new expected background value. The selected end times of salt waves that display increasing/decreasing trends should be taken as estimates.

Throughout this process flagging codes are produced to highlight the following common problems:

1. Low EC values (L): If the average EC value recorded for an event is less than 5 µS/cm then the sensor has low EC values.
2. No wave (N): If no starting time can be established through the iterative process described above or the maximum EC value falls within 3 µS/cm of the starting EC value (i.e., the EC values show minimal change) then no salt wave exists in the record.
3. Partial Wave (Pw): If the starting time occurs after the ending time or the ending time is within 45 seconds of the starting time, the record’s wave form is incomplete, and calculations should not be performed with these data.
4. Spike (S): If there is a dip or jump of more than 3 µS/cm in the EC wave that rebound to previous levels within 6 seconds we determine there is a spike in these data. These threshold values were chosen based on visual inspection of dump event data, with no specific data analysis methods being used.
5. Noisy data (Sd): We identify noisy data with standard deviation thresholds. If the standard deviation of SECb or ending ECb (EECb) exceeds 0.60, we identify these data as noisy. To determine high noise levels within the salt wave, we sperate salt wave data into two subsets: the rising limb and falling limb. Within both subsets we find the difference between consecutive EC values and calculate the standard deviation of the differences. If the standard deviation in either subset of the salt wave exceeds 2.1, the EC wave is identified as being noisy.
6. Extreme values (E): If the maximum EC exceeds 130 µS/cm, then extreme values are present in these data.

Keep in mind that the process of choosing the start and stop times of EC waves is combined with user observations. After the start and stop times of a waves are chosen through the process described above, the user is expected to double check the selected points. This ensures that the data are being looked at by the user and their judgements can be applied to unique scenarios which are not recognized by the code.

4. Summerizing stage data

After the EC waves have been analyzed, the stage data are manipulated, and summary statistics are produced. The code subsets the stage data to include values spanning the duration of the EC sensor record and calculates the average, minimum, maximum, and standard deviation of the subset. The code also identifies changes in stage that occur during the event by taking the difference between the average of the first 30 sec of the record and the average of the last 30 sec of the stage record. If the difference between these values is greater than 0.5 the stage is rising (R) and if the difference is less than -0.5 the stage is falling (F). If neither of these conditions are met, the stage is constant.

5. Create excel sheet for Google Drive

The code summarizes all EC wave and stage data into an excel sheet that will be used during the data cleaning process. These data populate an excel sheet that has been set up to create plots that depict the EC waves. This excel sheet, once populated with data from the event, is uploaded to google drive and the link to the document is identified so it can be entered into the database.

6. Calculating discharge and uncertainty of an event

To calculate discharge for a specific EC wave, we need several CF values associated with the sensor. Please refer to section two where the criteria used to select valid CF values is outlined. If there are more than six CF values per sensor that meet the temperature-based criteria, we choose the six CF values that have occurred closest in time to the dump event. For each CF value a discharge is calculated.

To calculate stream discharge, *Q*, for an EC wave we use equation 3

*(3)*

where *V* is the volume of the dumped solution in m3 and *t* is time in seconds with *Δt* representing the time interval between two consecutive EC measurements, in our case *Δt*= 1. *RCt*is the relative concentration of the salt solution (mL/mL)

*(4)*

where *ECt*is the EC value at time *t* and *ECbt*corresponds to the background EC level, which changes throughout the duration of the wave. The *CFx* value represents one of the valid CF measurements for the event. *ECbt* is calculated using equations 5 and 6.

*(5)*

*(6)*

While the calculation of discharge ideally occurs on waves that show little change in background ECthroughout the dump event, the inclusion of *ECb*change throughout time allows the method to be more flexible for background stream chemistry changes that are especially variable during storms on Calvert Island.

The uncertainty of the discharge calculation (*Udis)* in percent error is determined by the uncertainty of the dumped salt volume (*Uvol)*, the uncertainty of the EC sensor measurement (*UEC*), and the uncertainty of the CF measurement (*UCF*):

*(7)*

*(8)*

*(9)*

*(10)*

*Res* is the resolution of the EC sensor used (currently 0.01 for all sensors), *CF* is the CF measurement used in equation 4, and *σs* is the standard deviation of the regression analysis used to determine the CF.

The constant *Verr*represents the estimated error in the salt solution that was dumped during the event in L. For our system, *Verr* is estimated to be 0.0726, based on the dimensions of the gauge and the accuracy of the pressure transducer (used to determine the salt solution dump volume in the automatic salt-dilution set-up).

Please see Maartje Korver’s MSc thesis (located in the google drive, https://drive.google.com /file/d/1qL1VbIzijJHVtEdWg2\_r9oV8uS-9NoiW/view?usp=sharing) for more information on how specific errors were calculated: equations 19-20 and Figure 4 for the determination of *ECF* and section 3.5.3 for the calculation of *Uvol*. Please note that if any of the following field equipment ever gets replaced with similar equipment that has different dimensions or resolutions/accuracies, the uncertainty calculations should be adapted accordingly: EC sensors, salt solution gauge, pipet and flasks used in CF measurements.

After all individual discharge and uncertainty values are determined for each sensor (there should be a discharge measurement for each valid CF measurement), the overall discharge and error is calculated for the event. All *Q* values that are associated with EC waves having less than two flags are averaged to find overall discharge. EC waves that have two or more flags likely have unique data patterns which lack a substantial or noisy EC wave and are thus not included in the averaging.

To determine the overall error of the event, *Utotal*, the percent error of individual discharge measurements is converted to absolute error, *Uabs*, and a range for each *Q* value is found.

*(10)*

The maximum of all the *Q* ranges, *Qmax*, and the minimum of all *Q* ranges, *Qmin*, are determined.

*(11)*

where *Qavg* is the average of all *Q* values which corresponds to the overall discharge of the event.

7. Mixing

We also calculate the mixing percentage which signifies how well the salt solution has mixed within the stream before it reaches the EC probes. If the stream does not mix well (i.e., the mixing percentage is high) this creates discrepancies between discharges from different sensors. Events that have been poorly mixed are either excluded from the rating curve or, in the case of high discharge values, are incorporated into the error analysis.

Mixing is determined by the difference in discharge values between two sensors. Mixing percentages are only determine between discharge calculations that use CF measurements from the same day to reduce the influence of CF values on the mixing calculation. For each pair of sensors with corresponding CF values, two percent mixing values are determined. For a pair of sensor’s A and B, *MA*represents the mixing of sensor A relative to sensor B and *MB* represents the mixing of sensor B relative to sensor A

*(12)*

*(13)*

*Q* is the calculated discharge and *E* is the relative error associate with the discharge. If the calculated *M* value is negative there is no mixing discrepancy, thus *M*=0. For every combination of sensors with corresponding CF values we calculate *M*. The overall mixing percentage of an event is found by averaging all *M* values together, regardless of if the value relates to sensor A or B.

8. Updating the database

After all discharge, uncertainty, and mixing values have been identified for an event, the database is updated with the appropriate details. The tables in the database that are updated are autosalt summary, all discharge calcs, salt waves, and autosalt forms. After the event is added to the database, it should be checked by the user to look for any unique features of the event the code did not identify. To see how the data is post-processed, please see the data cleaning instructions located in the google drive.

9. CF trends

We performed trend analysis with all CF values in relation to time and stream temperature to help us determine how to best select valid CF measurements for each new event. For both analyses, we grouped CF values in three different ways: (1) CF values by site, (2) CF values by sensor, and (3) all CF values (i.e. no grouping).

When comparing CF values to stream temperatures we find a slight positive relationship. Figure 2 shows these relationships at the sensor grouping level with the table showing the linear regression slopes and R2 values. Of the 21 sensors evaluated, we found 7 of the sensors had a trend where R2 > 0.5 and 11 sensors had trends with a slope greater than 0.005. Figure 3 reveals the large range of CF values for each temperature bin however the mean CF values are increasing in relation to stream temperature.

To look at how CF values change over time, we compute the number of days from when the CF value was collected to the start of its corresponding barrel period. This is then compared to the CF values to determine if the length of time after the salt reservoir is filled affects the CF calculations. We find no significant relationship with only 1 of 21 models having a linear relationship where R2 > 0.50 (Figure 4). We also see high variability in CF value by station relative to time, with no definitive trends in the mean values (Figure 5).

From these findings, we chose to select valid CF values based on the temperature of the stream at the time of the dump event relative to the stream temperature during the collection of CFs. To capture the variability in CF values within a ±5°C temperature range we choose four to six CF measurements per sensor and average the resulting discharges together. If there are more than six valid CF measurements, those collected closest in time to the dump event will be used.

The ±5°C threshold was chosen based on the slopes of significant linear regression models (R2 >0.50) (Figure 2). When averaging the slopes of significant models together, we find that the average change in CF values due to stream temperature is 0.01. To prevent stream temperature from influencing the CF values by more than 0.05, the temperature during the CF collection and the new dump event cannot differ by more than 5°C.