# Catchment Data

## Data Correction Notes

Water Level

horizons
regional council

*These corrections notes are for reference during the data correction process. If you have not corrected the following data source previously or are still unsure of how it all works, it would be a wise idea to get the analyst to spend some time going through the process with you.*

## Creating Your Working Directory

Create a working file in/on your local Hera drive. The file path should be

\\ares\Environmental Data Validation\Water Level*[Site]**Batch#**Batch#.hts*

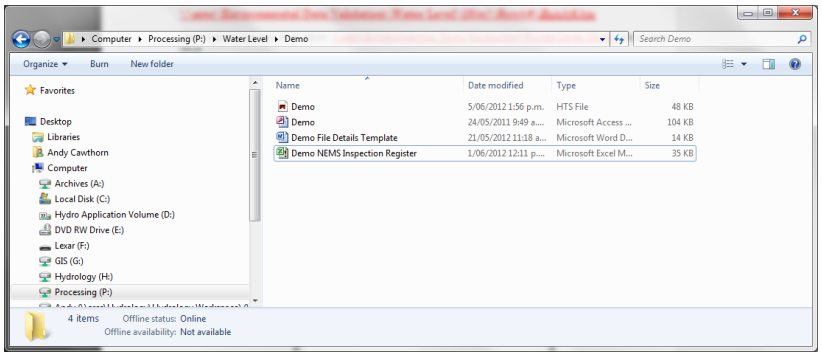
Copy the processing files from \\ares\Environmental Data Validation\Water Level\Docs\. From here grab

File Details Template

Inspection Register, and

Audit

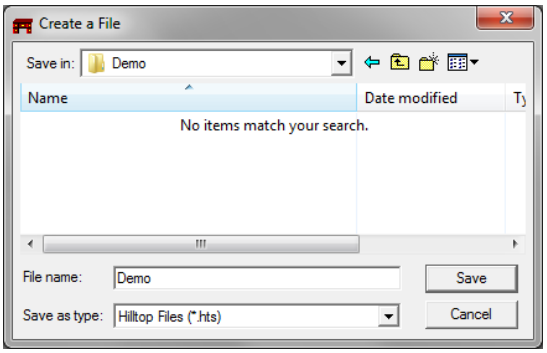
Copy the above into your working directory




Rename the File Details Template and Inspection Register so that they contain the batch # in front i.e. 102 File Details Template as shown above. For this example the file name is Demo

Rename the Audit to the Batch number i.e. 102.

Create a new hts file and call it the batch number i.e. 102. If you have done this correctly, hilltop will indicate that there is an audit trail in the background. If it does not, then you need to make the audit.mdb the same as the hts file. If you complete your corrections and there is no audit trail, there will be problems during the review process




COLLECT ALL THE CHITS FOR THE CURRENT PERIOD OF CORRECTION AND PHOTOCOPY THESE

OPEN CDTOOLS AND FILL IN THE PROCESSING REGISTER FOR THE SELECTED DATA SOURCE.

FIND THE SURVEY INFORMATION AND CALIBRATION INFORMATION. THESE TWO ITEMS ARE CRITICAL FOR DETERMINING THE FINAL QUALITY OF THE WATER LEVEL DATA

# GETTING DATA TO YOUR FILE

IN YOUR *WORKING* FILE, COPY THE ORIGINAL DATA FROM \\ZEUS\ORIG\<SITECODE>.HTS TO YOUR *WORKING* FILE AND CALL IT <SITE_RAW>.

**Comment [AC1]:** This needs changing. Brent still has not sorted the correct file paths and permissions

COPY ANY BACKUP DATA IF AVAILABLE. IF YOU ARE ANTICIPATING A DIFFICULT BATCH OF DATA, THEN COPYING ADDITIONAL SITE DATA MAY ALSO HELP I.E. UP AND DOWNSTREAM SITES, OR ADJACENT SITES THAT HAVE COMMON CHARACTERISTICS.

COPY IN THE WATER LEVEL STATISTICS ALSO. YOU WILL NEED TO COPY THE WATER LEVEL STATISTICS:SD TO SERIES DATA IN THE FILE, AS IT IS STORED CURRENTLY AS A VIRTUAL MEASUREMENT (VM)

IF SURVEYING HAS BEEN COMPLETED DURING THE CORRECTION PERIOD, GO AND FIND THE LEVEL BOOKS AND PHOTOCOPY THE INFORMATION FROM IT. IT MAY ALSO HELP TO COLLECT THE LAST ONE OR TWO SURVEYS ALSO.

IF YOU HAVE NOT PROCESSED THE CURRENT SITE BEFORE, OR ARE UNFAMILIAR WITH THE SITE, COPY THE LAST BATCH OF PROCESSING TO THE FILE. THIS WAY, YOU SHOULD HAVE SOME INDICATION OF WHAT WAS DONE IN THE PAST OR ISSUES OBSERVED.

# FILE THE FILE DETAILS TEMPLATE

FILL THE FDT WITH THE SITE AND CORRECTION PARAMETER

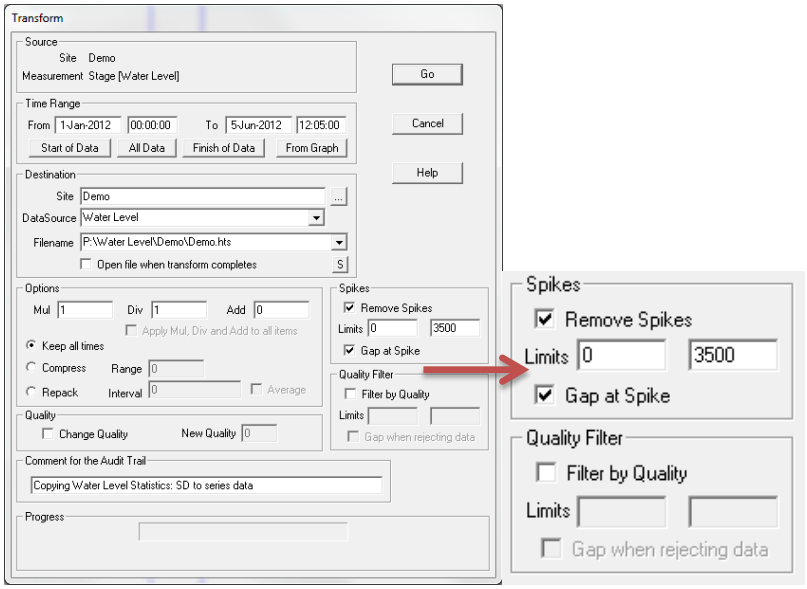
UNDER INITIAL FILE DETAILS > RIGHT CLICK THE <SITENAME_RAW> AND SELECT *DETAILS* AND COPY THAT INFORMATION IN

RIGHT CLICK *GAPS* AND LIST, IF ANY, THE GAPS IN THE RAW FILE AND LIST ANY GAPS DELETED IN THE *GAPS DELETED FIELD*. *IF THERE IS NONE, THEN ENTER ZERO GAPS*

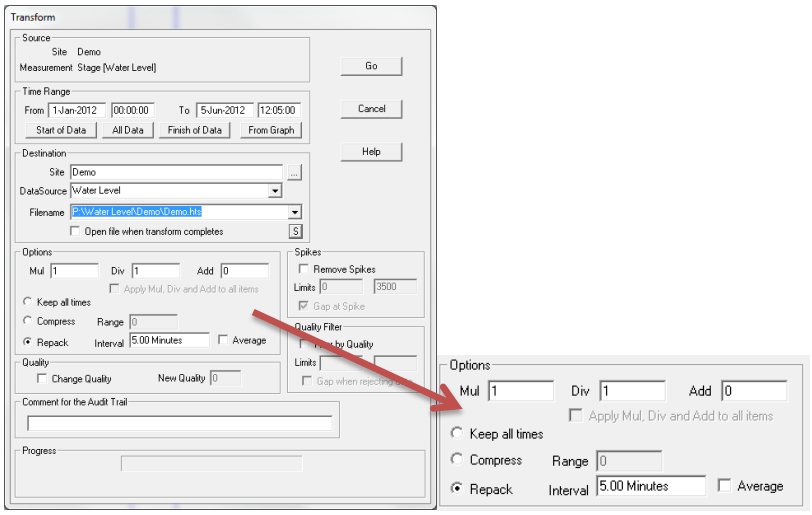
# Spike filtering the FDT

Copy the "Site_Raw" data to another object within the collection tree called "Working". This "working" data is where you enter check data and minor editing to the data. However, rather than copying the data, use the transform function and filter out the logger/sensor recording faults by inserting a gap.

- Use the Transform function in Hilltop. Select the Spikes. This can be used to filter spike noise from the file. Use the Limits to isolate suspect spikes. Ensure that the Gap at Spike is also selected
- List the gaps created in the FDT under: **Gaps Introduced from Spike Filter**




- In Hilltop, list the gaps created and delete them from the record. Be sure to keep gaps in the record that are not just simple linear fills/repacks. Check around the gaps introduced esp. periods of both rising and falling recessions
- List the gaps deleted to the FDT under: **Gaps Introduced from Spike Filter Deleted and Repacked to Recording Interval**
- Transform the Working file back to the recording interval under the Options tab in the Transform dialog box. This repacks the data to the recording interval. Hilltop effectively does the interpolation work for you. Be sure ***not*** to have the Average check box clicked
- At this stage, the edits made for this process do not require Quality Coding, unless the Gaps and subsequent interpolation introduce cautionary or suspect data to the record. *In the past, all gaps removed from the record required a comment and corresponding QC. This was a pain esp. when there were often hundreds of spikes in the dataset. Essentially, all you had was a comment database loaded with spike removal comments and QC codes – Not overly critical, nor useful, to the end-user*




# Loading the Check Data

Once you have photocopied the inspection chits, do the following:

Use the Hilltop Check data Loader.mdb. Ask an analyst if you are unsure on how to use this. The check data loader is essentially the output of the log sheet loader. However, even though the log sheets have been loaded to the loader, they still be checked for errors

Check the Julian date with the date on the chit. If it is correct, put a tick in the date field. If it is incorrect, find out why? Add a non-conformance outlining the issue

Look through the chits for poor field practice i.e. no dates and times on the chit, insufficient information or other things that do not add up and non-conformance issues identified.

Highlight any ESG and/or EPB information and also check inspection comments that could affect water level correction i.e. gas purges, removal of debris, channel changes, logger code changes etc.

In the "working" data, select the check data (the icon that looks like a clipboard). 

enter in the check data information. In most cases, there will be no EPB information, so you can add the ExSG +/- error into the EPB column. For inspections where there is no ESG information, do not add to the check data. Under the old schema, this just created all sorts of confusion with non-hydro staff

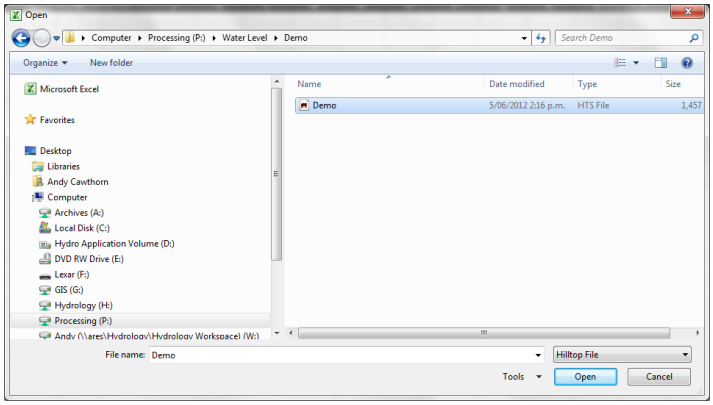
For towers where there is both an EPB and ESG, the values require rearranging i.e. *the EPB values go into the ESG field and the ESG go into the EPB field*. This is because the EPB is the primary reference.

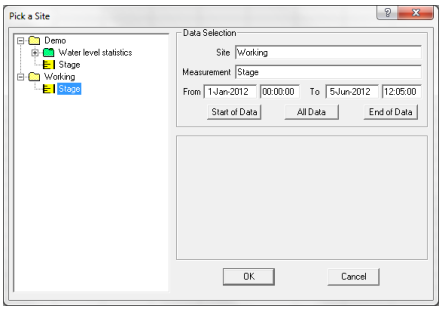
Once you are done, save the file. Actually, saving often is recommended because of Hilltops temperamental nature and does not auto save!

Add the comments to the Initial File Details

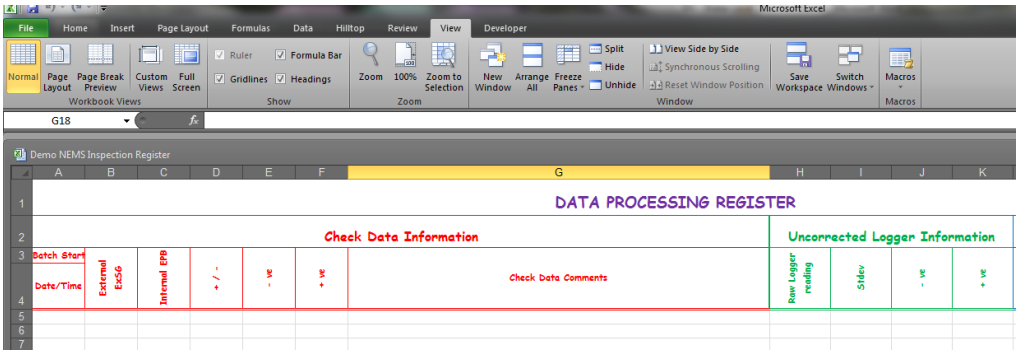
# Populating the Correction Register

- Open the Batch Number NEMS Inspection Register
- The register is used to isolate suspect inspections, trends and other odd bits
- in the 'Dump' Tab use the 'Get Hilltop Check Data' script. this will populate the spread sheet with the check data values you have previously entered into hilltop, and pluck its associated logger value




- Add, by copying and pasting, the check dates/times into Cell A5, the external ExSG reading in B5, the EPB (if any) into C5, the +/- ExSG error into D5 and the Raw (non-edited)logger values into H5. The spread sheet should do all the math for you
- Copy the check data from the Working file to the Water Level Statistics file. Run the Get Hilltop check data script again, but this time for the Standard Deviation. Copy and paste the output into cell I5
- If you do not have the 'Get Hilltop Data Script", ask and analyst to install this for you




- To aid correcting the data, add the Hilltop Inspection Comments to the register in Cell G5

# Editing the Data

The working file is where you are going to make edits to the data. Here are some more popular methods:

***Use of the Register:***

use the register to identify issues or areas that are a bit suspect. Note comments regarding the corrections made with Excels Review Tab. This can put notes and comments on the spread sheet. This is useful reference when annotating the plots

While the time series data is not corrected to the external or internal reference, these should be looked at carefully, esp. when there is clear deviation between the measured and reference.

Have the Comments Tab open in the Register, as any changes that you make to the data, require a corresponding comment irrespective of how mundane/tedious or pointless it may seem. If you cannot find the 'technical' term to describe what you just did, open the comments register and one should pop-up for you or ask the analyst

View a month's worth data and right click edit. have a look around inspections for missing data, spikes caused by the on-site technicians (purges, calibrations etc.) or periods of noisy data generated from sensor fouling or burial. Not everything they do in the field is written on the field chit!

***Minor/Major Edits:***

Only remove spikes that stand out from the rest. Noise in the data is perfectly acceptable. Use the QC schema for its respective quality marker

Sudden dips and rises in stage data should be treated with caution. If they are the result of known processes i.e. logger offset changes, purges, sensor burial etc. edit the data – use either the last good inspection, or event, and ramp to the surrounding data from the change in offset. Make note in the comments on the method used

If suspect rises and falls occur treat these with caution. The general rule is that if there is no evidence to suggest that the data is faulty or suspect, leave the data in the record. River hydraulics is a stochastic process and it would be expected that there are some odd bits in the data that cannot be explained, one of the beauties of hydrology as a science. If you are not sure, then add a comment to the comment register

Periods of slips in the data, drawn and recovery should be left in the data. Add a comment to the comments register Of what is happening. These hydrologic events, should only be edited from the data if the introduce extreme low and/or high flows (however, you won't really know this until you have done the ratings)The reasons for this, is that having all the hydrologic characteristics in the data, whether they be slips or drawdowns/recovery, is reflective of the catchment characteristics. Hydrology sites are not there to measure in-situ site physiognomies, but the catchment characteristics. It is much easier also for scientists to choose what they would like to add or remove from their datasets! And finally, if they decide in 20 years' time that they actually want that data in there on the archive, it is much harder/and time consuming to go back and add the data back in

Radar data is all noise and should be left as such. Just remove the feedback noise spikes from the data (that should have been covered in the earlier spike filter/Transform). Radar is often used primarily for flood warning and backup, and is treated as such

Noisy periods of data can be interpolated using smoothing virtual measurements (refer to virtual measurement documentation). Smoothing should only be applied below the series data upper quartile, unless it is really noisy or along recessions. However, not all noise in the data set should be smoothed. Rivers are naturally noisy, so editing out noise is up to the site characterises, sensor issues and preference. In many cases, it is more ideal to remove obvious noisy periods. In fact, some batches may fail if there is too much smoothing

***Synthetic and Missing Record:***

Missing record over hydrograph peaks should not be attempted with synthetic record. Leave it as missing. However, this is open to interpretation. An easy way is to copy stage data from a nearby site in to the period of missing record. While not overly accurate, the data is ramped to the surrounding data. Horizons use a no missing record criteria. Data corrected this way over peaks may end up being left as missing record following review

Periods of missing record: Close the gap and use hilltops Transform and repack to the recording interval. Double check for suspect data following the repack

Adjustments to controls that affect water levels should not edited using ramp to recession techniques. The data should be 'cleaned' to reflect the adjustment to water level following the modification on the control. Changes to controls should be adjusted using a modification to/from ratings

Check the data with the calibration reports. Note also that even though a calibration may 'fail' it may be calibrating in the operating range of the senor. DOUBLE CHECK the calibration data

Check the survey information also. Unfortunately, issues with surveys are found over time, and not instantaneously during the survey

# Using the Register

Once you feel that you have made all the changes to the data, copy the *"Working"* data to Site Name i.e. Manawatu at Hopelands. In the correction register, run the '*Get Hilltop Check Data'* macro again. Note any changes to the logger value in the register. Add the adjusted logger readings into cell L5 of the register

For large differences between logged and external water levels, you need to find out why? This is OK during flood events, as it is expected that there should be some degree of difference. Not so much during lower flows/stage levels. if you suspect that there is enough evidence to warrant some degree of caution, then QC the data accordingly

Check the +/- error associated with the ESG reading. For example, if the ESG is 574 +/- 3 mm and the logger is 643, and there is no reason for the discrepancy, you should find out why? it may simply be a mis-read esp. if the next reading is at similar stage height and the difference is within tolerance

Conversely, avoid all high stage ESG readings if the final difference is outside the +/- error. Never correct to the situation where the final difference exceeds the associated error during high water level events. For example, an inspection has an ExSG reading of 3260 +/- 50 and the logger value was recorded at 3550. Do not correct the data, unless the sensor is buried or other odd situations that would warrant adjustment. Such corrections often result in the warping of data and unnatural looking rising limbs and recessions of the hydrograph

Look for poor readings especially during periods of low flows i.e. a difference of 23 mm during low flows at Mangapapa at Troup Road without indication of any issues should be treated with caution. The water surface would be very still at these levels. Data such as this, affects stage to flow ratings and MALF statistics. You need to look for why the situation occurred. If it is a poor reading,

Look for trends in the data set. If there are clear trends in the dataset that warrant some degree of caution, make note of these.

# Quality Coding the Data

here we apply the quality code schema

Quality coding is required for all the changes that are made to the data, and are as follows:

All data that has not been modified and has been calibrated and surveyed to the operational standard as defined in the NEMS standard QC = 600

All minor corrections ≤ 3 hours QC = 500

All Major corrections to the data ≥ 3-24 hours QC = 400

Synthetic data QC = 300

This code refers to data that has actually been made up to achieve required goal: Using data from other sites, unreliable interpolation to fill gaps/noise, regression based data etc. Generally editing that goes beyond simple spike removals and smoothing. Smoothing ≥ 1 Day

Unverified Data QC = 200. Flood warning sites (5 year survey info contains the baseline QC schema for each site). Data out of calibration/survey. Suspect deviations from inspection information. Upwards recessions. Areas generally that warrant caution to the final user. All unreferenced data sources i.e. Soil Moisture and Temperature. Backup only Rainfall Sites

Missing record QC = 100

Add the quality data to the file details template

# Final Steps

Create another file called 'Audit' as a Virtual Measurement and use the VM. This shows the changes made to the data during the correction process and helps in the review of data

Fill in the rest of the file details Template: Final Details and Gaps

Open the URF form located in \\Hera\councildata\Hydrology\Hydrology Data Correction\URF2010.xls

Copy the Corrected data to the Sub Archive. Make sure you have the *'Gap at Start'* check box ticked!

# To Print

Print the URF2010

Print the File Details Template

Both the Correction register and the comments spread sheets

graph showing the quality data

graph showing the before and after i.e. raw and final

3 month plots of the above showing the Audit VM. You may also want to annotate these plots

Open the audit trail in excel and print this off. Format the date/times to DD/MM/YYYY hh:mm:ss

Fill out CDT for the date the correction batch was completed

Hand in the corrected data file to the analyst for review