CATCHMENT DATA

DATA VALIDATION

1. TURBDITY (ISO): VISIO TURB- WTW

Version Number:	3.1
Issue Date:	23/08/2012
Portfolio:	Turbidity and Sediment
Compiled by:	Michaela Cowie

horizons
regional council

VERSION	CHANGES MADE:	DATE ADOPTED	BY
V1	*Basis adopted from Historical Turb sensor processing*	*2011*	*Cowie*
V2	*Refinement of processing towards NEMS*	*03/2012*	*Cowie*
V3	*Processing update: periods of recession*		*Cawthorn*
V3.1	*Updating file paths*	*23/08/2012*	*Cowie*

Introduction

The specification for Turbidity data is contained within the Turbidity and Sediment portfolio documents.

Validation Methodology

The methodology employed in the validation of turbidity data has been adapted in house from the following references:

- ISO 7027
- USGS guidelines for Continuous Water Quality
- USGS Implements new Turbidity Data reporting procedures. 2004
- USGS OWQ Field Manual Chapter 6, Turbidity. 2005
- Lab procedures for how the check data is analysed.

These corrections notes are for reference during the data correction process and assume that you are relatively proficient at the use of Hilltop software.

If you are not fully familiar with the validation of this data source or are unsure of the procedures to follow please read this document in full and /or consult with a Data Analyst or Coordinator, to obtain further assistance.

This Data Validation document identifies the basic procedures to follow when undertaking the processing and reviewing Turbidity ISO Data.

This Data Validation document is currently a work in progress and subject to frequent modification.

1.0 Creating Your Working Directory

Open Catchment Data Tools (CDTools) and fill in the processing register for the data source and correction period. You will need to do this first to obtain a batch number for the new period of processing. Begin the processing from the end of the last batch.

Collect all the chits for the current period of data corrections. Check that all chits are filled in, accounted for and note the final inspection in CDTools.

Create a working directory in Hydrology Data Correction. The file path should be:

\\Ares\Hydrology\Environmental Data Validation\Turbidity\Turbidiy ISO\<SiteName>\<BatchNumber>\...

The above processing folder needs to contain the following:

- File Details Template.doc (FDT)
- Turbidity ISO Register.xls
- Audit.mdb
- URF_Turbidity ISO.xls

Copy the processing files from:

\\Ares\Hydrology\Environmental Data Validation\Turbidity\Turbidity ISO\Docs

Rename the File Details Template and Turbidity ISO Register so that they contain the Batch Number in front i.e. 102 File Details Template

Rename the Audit file (Audit.mdb) to the batch number i.e. 102.mdb

2.0 Configure Hilltop processing files

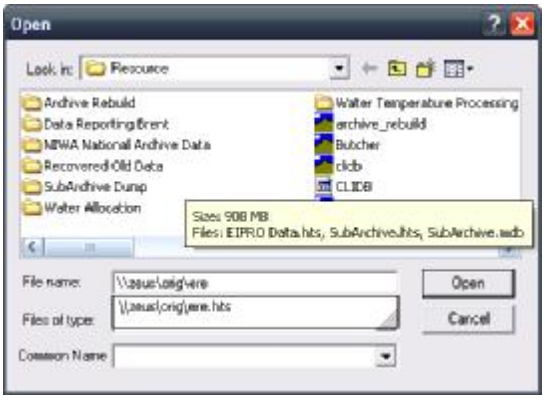
In Manager, create a new .hts file and call it the Batch Number i.e. 102.hts, save this to your working directory. If you have done this correctly, Hilltop will indicate that there is an audit trail in the background. If it does not, then you need to make the audit.mdb and label the audit.mdb the same as the .hts file. If an audit file fails to generate a connection then contact an Analyst.

If you complete your Validation/Corrections and there is no audit trail, there will be problems during the review process and the URF may be declined.

The reason for indexing the batch number and the folder/file structure is that the system databases (Hilltop, CDTool, Audit trails) are now linked. This enables connections and reporting between the different systems.

3.0 Populate Your File with Data

In your Hilltop file, copy the original (raw) data from \\Ares\Original\Hilltop Telemetry\<SiteCode>.hts or \\Ares\Original\Internal\<SiteCode>.hts [i.e. \\Ares\Original\Hilltop Telemetry\WKW.hts] to your Hilltop file and call it *[<SiteName>_Raw]* [i.e. Kiwitea at Haynes Line_Raw]



- Copy the Turbidity (ISO) from the Original file.
- Copy any EPA data including Turbidity (EPA) to the same file.
- Copy the Turbidity and Sediment data from Sampler Provisional and Water Quality Archive to the same file.
- Copy the Water Level data from the site currently being corrected. This can either be from the provisional archive or raw (telemetry) data. It is not critical, but the more corrected the data, the better. Copy the flow data also if it is available.

Tip: using uncorrected data change the quality code to differentiate from the Hydrometric/Provisional Archive data as well as Water Quality Archive/Sampler Provisional data.

- Copy Turbidity, Stage and Rainfall data from up and downstream sites. If these are not available, use neighbouring sites.

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 can be obtained from the Hydrometric & Provisional

Archives). This way you should have some indication of what was done in the past or issues observed.

Or use the previous batch documentation on:

\\Ares\Hydrology\Environmental Data Validation\Turbidity\Turbidity ISO...

4.0 Populate the File Details Template (FDT)

Fill the File Details Template (FDT) with the <SiteName> and Data Source: Turbidity ISO.

In Manager, under Initial File Details > Right click the <*SiteName_Raw*> and select details and copy that information into the FDT.

5.0 Loading the Check Data

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

5.1 Use the Turbidity (ISO) Combined VM to combine the low and high range turbidity data sources. Turbidity (Low Range): Point Sample is the primary, data record maximum at 400 NTU. Turbidity: Point Sample is the high range, over-ranging above 4000 NTU.

Note Turbidity: Combined Point Sample is used for telemetry purposes and runs a different VM.

5.2 Copy the Turbidity (ISO) Combined data to another site within the control tree called <*Working*> and rename as Turbidity (ISO). This <*Working*> data is where you enter check information data and edit the data. If working with large data files, correcting 1 year at a time is a better option. In this case, rather than naming the file <*Working*> within the control tree, add the year being corrected i.e. <2005>.

5.3 Check the chits/highlight important/relevant details to the processing record:

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?

5.4 List the Lab data from Turbidity and Sediment, Turbidity ISO (HRC) for the processing period from the raw file. Print this off and note all the Turbidity check information on the

photocopied chits. **Do not use** Depth Integrated samples as these are taken across the channel and not at the sensor using them may distort any corrections which are subsequently made to the data.

All Sampler Provisional/Water Quality Archive data (previously Qualarc data) needs to be treated with caution and needs to be fully verified in this process (as the system contains numerous entry errors) like the following:

- *Samples taken on weekends (though this can be checked with log sheets).*
- *Samples taken at midnight (system has lost its time format).*
- *Other water quality parameter in the turbidity data source*
- *Regular adoption of the turbidity sampling (Suspended sediment concentration) was not adopted until July 2011. Prior to this results might be inconsistent from the lab, i.e. only turbidity EPA or ISO given for samples.*
- *When samples were collected/picked up/sent to the lab and when the sample was physically taken from the river*

For most sites, more than one council team has taken samples, so there might be occurrences where there are multiple samples taken on the same day or samples without hydro log sheets. Some instances you will be required to resolve these issues by locating sample sheets or reviewing the sample project codes.

Any sample found to be incorrect in the Water Quality Archive and Sampler Provisional should be reported and non-conformanced.

Using the photocopied chits, highlight any other turbidity check inspections not found within Water Quality Archive and Sampler Provisional**,** and comments that could affect Turbidity corrections i.e. sensor cleans, calibrations, new sensors, burial, and logger code changes etc. Only add comments that could affect turbidity data corrections

5.5 In *<Working>*, select the check data (the icon that looks like a clipboard). 

Enter in the check information. Enter in the dates and times into their respective fields. Where there is no turbidity information, enter a -1 into the turbidity check column. This preserves the record and indicates that an inspection was done on that particular date and time, but no sample taken.

Enter the check inspection comments into the comments field

Once completed, save the file. Actually, saving often is recommended because of Hilltops temperamental/dated nature and no auto save feature

5.6 List and add the check inspection comments to the Initial File Details (FDT)

5.7.1 Apply VM Spike filter to remove all over-ranging values within the data (*see Turbidity Processing Appendix*). Using the spike filter VM does not affect data quality.

Transform and repack the spike filter VM into Turbidity (ISO)

5.7.2 List out all gaps and copy these into the File Details Template note under Gaps in Initial File: Spike filter gap removal.

5.7.3 In Hilltop list and delete the minor gaps: 1hour and less in duration. Transform the data series back into 15minutes. Quality data is not required for this step.

If there is any missing data (gaps) for the period that you are validating, you will need to locate this period of data or determine if the gap is valid; contact the Regional Coordinator to obtain the missing data and/or determine if it is in-fact a period of missing data. This data should be updated to the site original files first then copied to your working file.

6.0 Populating the Turbidity ISO Register

Populate the Turbidity ISO Register with the check data and the raw data file. There are a range of tools to do so, if you are unsure ask an Analyst.

Complete the Initial data stats within the Register.

Difference: Logged-Checked

Difference %: Logged/Checked

7.0 Editing the Data

The *<Working>*, (or *<200X>*) file is where you are going to make minor edits to the data.

Have the Comment Sheet open (within the Turbidity ISO Register), as any changes that you make to the data, require a corresponding comment. If you can't find the 'technical' term to describe what you just did, ask an Analyst.

Here are some more popular methods:

7.1 Removing point spikes

In the case of singular spike data, use the spike filter VM to remove these spikes. Using the spike filter VM does not affect data quality

7.2 Making Minor Edits

Editing the data is simply a copy and paste affair and using common sense when applying the minima Virtual Measurements to which data is considered real and that to replace and/or remove.

7.2.1 Create a series of Virtual Measurements within the object tree under the <*SiteName_Raw*> data. These should contain the following:

- 3 hour minima.

- 6 hour minima.
- 12 hour minima.
- 1 day minima as last resort (these are likely to be removed and left as missing record)

7.2.2 Transform the VM's into series data within the file i.e. 3 hour Minima.

View a month's worth data at a time with check data loaded and right click edit. Have a look around inspections for missing data, spikes caused by the on-site technicians or periods of noisy data generated from sensor fouling or burial.

Follow the flow chart when correcting noisy or suspect periods of data from highest to lowest attempting to reduce the amount of noise within the data. It is important that you preserve as much data as possible. Do not just default to the largest spatial minimums as a default. The order of preference if from best to least quality.

Try not to re-run the VMs on top of each other. However, it may be necessary during extreme fouling periods.

If the flow chart fails to solve the problem, use visual interpolation, or if this fails, delete the data and leave it as missing record for now.

It is important during the copy process that the 'comment for the audit trail' and 'quality data' is added. However, depending on which version of Hilltop you are running, this may or may not work.

When copying Data from a VM use the Transform Data option as copying a VM will only copy the rating, not the series data generated by the VM.

To aid in the copy process, overlay the VM data with the raw data. Zoom into the period of data that you think requires smoothing and use the "copy from graph" (rounding to the nearest 15 minute) option in the Transform pop-up menu. This speeds up the copy process rather than manually entering the to-and-from dates to the copy period (the copy to-and-from dates/times appears in the audit trail). The smoothing VMs only require QCs.

The copy process from the VM data to the working file does not require commenting. The copy process shows up in the audit trail. The critical part is the addition of the quality data during the copy process. There is no point regurgitating what is in the audit trail!

Any correction that has been 'visually' interpolated, such as a spike, greater than an hour is synthetic record and needs to be listed in the Comment Sheet.

Missing record over large events should not be attempted with synthetic record. For now, leave it as missing record and should be listed in the Comment Sheet.

For periods with an upward recession, this data may require deletion. Leave the data missing for now or try the 3 and 7 day minima VMs (esp. if the data is during recessions. The aim is to recover as much information as possible).

Ramp correcting to the adjacent date is also an option.

Over-ranging: *It is not missing* - we just do not know what the data is doing beyond this point.

7.3 Filling Missing Record with Backup Data

Filling the missing record is as simple as identifying the gaps and copy and pasting the Turbidity EPA record if it is available for the site. If not, leave the gaps for the present time.

7.3.1 Copy the Turbidity EPA dataset from the Provisional Archive into the *<Working>* directory.

Note: Using uncorrected Turbidity EPA make sure to over plot the flow rate and water temperature for the pump to determine if and when the pump was running as this will determine if you are able to use this data or not. If the pump is not running you cannot use the Turbidity EPA data.

Determine where the data is suitable to fill the gap.

Once the gap identified has been filled with the Backup data, the data requires ramping and pulling to the adjacent start and end times of the gap markers (this is done in step 8)

As always, avoid adding Backup data to fill in the missing peak information if it introduces unrealistic peaks or over an extended period of time.

All Backup data inserts require a comment (beginning and end of when the data has been inserted) in the Comment Sheet and are Quality Coded as 200.

Depending on Hilltops bugs, there is an option to add comments during the transformation and/or copy process, which adds comments into the audit trail. Quite a handy tool for periods of substantial editing! In addition the ability to potentially assign a QC of 200 to copy across as you are editing the information.

7.4 Gaps which cannot be filled due to missing record, particularly over events should be QC as 100.

8.0 Ramp Corrections

Once you have made all the smoothing corrections to the data i.e. spike removals, smoothing etc., it is time to ramp the data, if needed, and more obviously during sensor cleans, replacements etc.

Copy your <*Working*> file into <*Corrected*> file in the Control Tree. Any ramp changes required are made in the <*Corrected*> file.

Make sure that when ramp correcting Turbidity, that the actual ramp correction does not force negative values into the data, nor use the lab samples for the ramp corrections. The data can only be corrected to a cleaned sensor value and/or recession information or the insertion of synthetic data.

When all ramp corrections have been made copy completed <*Corrected*> into [SiteName] file.

9.0 Quality Coding the Data

Quality coding is required for all the changes that have been made to the data and should have been added during to the copy from VM data series to the <*Working*> (Synthetic Data), when copying in the Backup data (Turbidity EPA), and any minor edits (Data Correction)

The quality data should be added during the copy process from the VM series data to the working file in the "Copy Data Source" dialog box

9.1 Populate the Turbidity ISO Register with the corrected data from [SiteName] directory. Complete the Register using the corrected data adjusted difference to assign the data quality. Enter the Quality value in two steps:

1) Use the Relative Difference (%) to assign the Quality Code. See Table below: Check for values which have a large difference or below 5NTU.
2) Correct the quality code for values below 5NTU which have been assigned a low QC. Use the absolute difference to assign QC values. See Table below:

Table 1: Quality Code values adopted from NEMS

Relative Difference		**Absolute Difference (0-5NTU)**	
Quality Code		Quality Code	
600	≤ 5%	**600**	0-1 NTU
500	≤ 10%	**500**	1-2 NTU
400	≥ 10%	**400**	> 3 NTU
Universal Quality Codes			
300	Synthetic		
200	Cautionary/unverified/Backup		
100	Missing		
0	Raw		

9.1.1 Use the Comments Sheet to correct the synthetic data within the period for quality coding. The Turbidity ISO Register QCs should be the primary quality codes for the period.

Check the data transferred from editing steps is in accord with the Comment Sheet.

Print off gap data from Hilltop file and QC the remaining gaps within the data.

9.2 Load the quality data to [SiteName] and list it in the File Details Template.

When inserting quality codes, **the lowest quality prevails. Quality codes are applied from the previous inspection**…i.e. the quality code you calculated for each inspection is applied to the previous inspection when you enter the information into Hilltop from the Turbidity ISO Register.

Save the Quality Data when you have finished.

10.0 Audit

Within the .hts file create a Site Audit and audit data stream VM and VM script:

Combine the data sources from each site:

Get “{<site>_Raw} Measurement [<Data source>]” as x

Get “{<site>} Measurement [<Data source>]” as y

Z = x-y

Put z

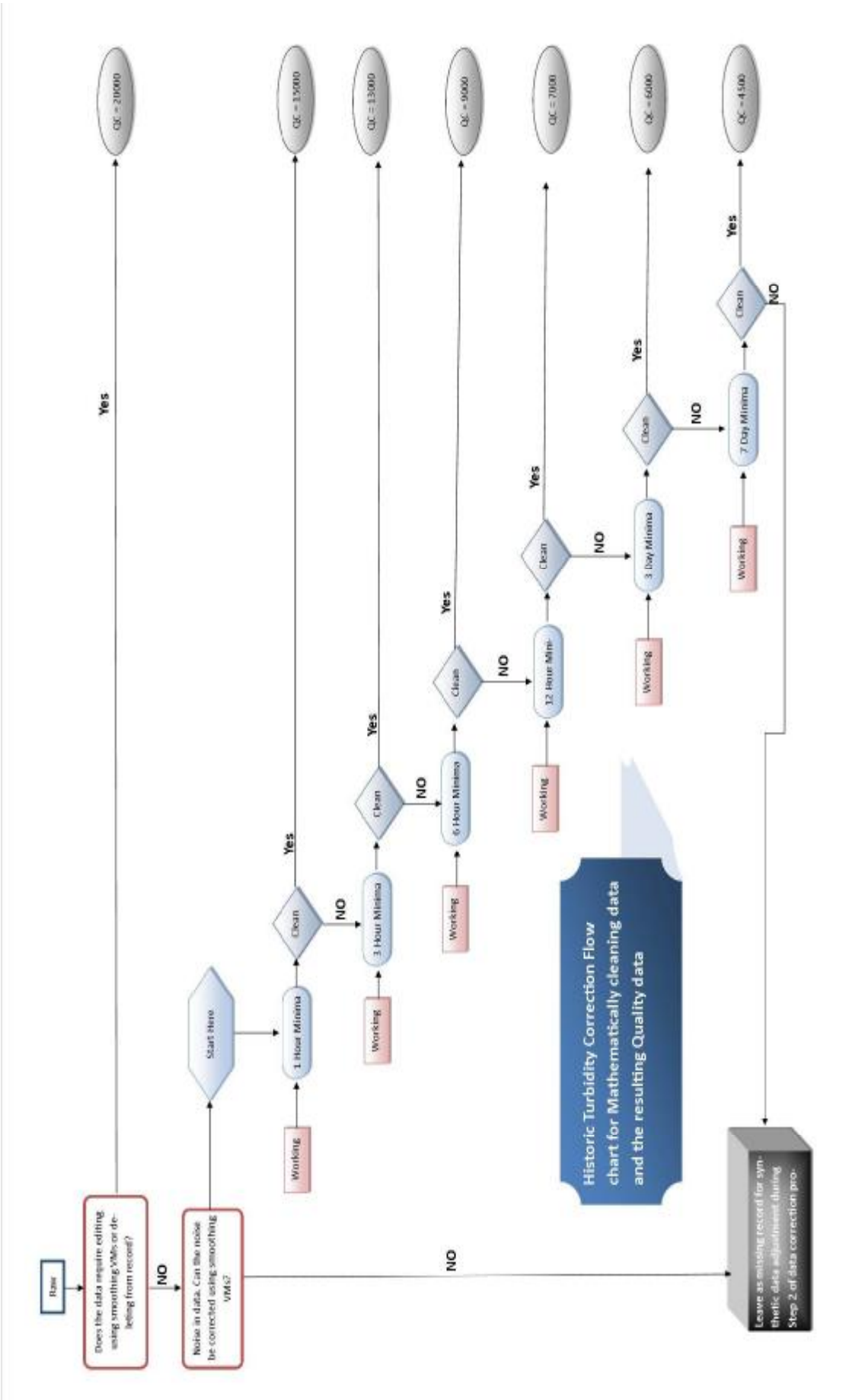
i.e.

Get "{Mangatainoka at Pahiatua Town Bridge_Raw} Turbidity Combined (ISO) [Turbidity (ISO)]" as x

Get "{Mangatainoka at Pahiatua Town Bridge} Turbidity (ISO) [Turbidity (ISO)]" as y

z = x - y

Put z


Raw
Does the data require editing using smoothing VMs or deleting from record?
Yes
NO
QC = 20000
Noise in data. Can the noise be corrected using smoothing VMs?
NO
Leave as missing record for synthetic data adjustment during Step 2 of data correction pro-
Start Here
Working
1 Hour Minima
Clean
Yes
NO
QC = 15000
Working
3 Hour Minima
Clean
Yes
NO
QC = 13000
Working
6 Hour Minima
Clean
Yes
NO
QC = 9000
Working
12 Hour Min-
Clean
Yes
NO
QC = 7000
Working
3 Day Minima
Clean
Yes
NO
QC = 6000
Working
7 Day Minima
Clean
Yes
NO
QC = 4500
Historic Turbidity Correction Flow chart for Mathematically cleaning data and the resulting Quality data

11.0 Final Steps

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

List out all remaining gaps and copy these into the File Details Template note under Gaps in Final Details and Gaps.

11.2 Fill in the URF Form

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

12.0 To Print to pdf

1) Print the URF

2) Print the File Details Template

3) Print the Turbidity ISO Register

4) Comments Sheets with the missing record

5) Graph showing the quality data – full range plot (QC)

6) Graph showing the before and after i.e. raw and final – full range plot

7) Full range (raw + final) + audit VM

8) Open the audit trail in excel and print this off

9) Attach copy of Log Sheets

10) Attach copy of Turbidity (ISO) sample values listed

13.0 Open CDtools and complete the required information.

14.0 Put a post-it note on the final inspection in your processing batch with your name and the data source you have processed.

15.0 Return chits to their draw if you have not done so before.

Collate these printouts and the copied chits in order and "Hand in" for review and updating.

Reviewing the data

Reviewing the data is a relatively simple process if the above has been followed. The review process of data is to insure that the correction methodologies found in this Data validation Document are followed. If areas of improvement are identified, make note of these. A quick way for the reviewing of the corrections made is use the auditing VM (located in the CDT VM reference library) and have the quality data open to check against the changes made. The below gives an indication of what to look for when reviewing the data:

1. Check the URF form on the front. Is all the information needed found there? Cross check CDT. If errors exist, make the changes and note them on the URF and/or CDT
2. Check the initial File Details Template. Is it all correct? Have the check comments been made, spelt correctly or missing? Note the mistakes/errors on the review template
3. Are all the changes made to the data complete and accurately reflect the correction processes used in the audit trail? Have the comments a corresponding quality code?
4. For periods that utilise 3 and/or 7 day minima's make note of the periods that require deletion and filled with synthetic data
5. Does the audit show all the changes made. If not, then add comments to the review template. Access 2007 and 2010 contain a range of tools for analysing the audit trail and finding the necessary information for reviewing
6. Are all the graphical printouts within the processing? Are there any visual inconsistencies'?
7. Fill out the non-conformance register with all errors found in correction process
8. If minor errors exist, correct them and make note of the changes made
9. *Keep in mind when filing the review sheet, that correction batches only fail if they require complete or substantial re-correction, which to be honest, given the rushed first attempt, may be the case…*

Appendix: Turbidity Processing

VM to combine – Turbidity (ISO) Combined:

Get "Turbidity: Point Sample" as TurbH
Get "Turbidity (Low Range): Point Sample" as TurbL

If TurbL < 399 then
Put TurbL
Else
Put TurbH
Endif

Apply VM Spike filter to remove over-ranging within the data:

Get Turbidity (ISO) as x

If x < 4000 then
Put x
Else
Putgap
Endif

Virtual Measurements which appear within the Turbidity Processing VM and form a set of accepted VMs:

- Flow to sample EPA
- Flow to sample EPA Recession
- 3 Hour Minima
- 6 Hour Minima
- Baseline Filter
- Spike Removal Filter
- Data Removal Tool
- 12 Hour Minima

- 1 Day Minima

If these original VMs are edited beyond their titles/descriptions add this into the Comment Sheet or FDT details.