

Department of Water WA 4.4

SEMI-AUTOMATING THE DATA VALIDATION PROCESS WITHIN HYDSTRA

AIMS:

- To streamline the import of hydrologic time series data files presently downloaded manually from field sites by developing automated systems. An initial target is ~700 data loggers being installed at Groundwater sites across WA.

- To radically improve the calibration/validation process by semi-automating time calibration, data spike removal, rainfall calibration and the simpler sloping corrections.

- To drastically reduce (and ultimately eliminate) the regular occurrence of data backlogs and to introduce a robust automated preliminary validation system to be applied to telemetered data as it is imported, prior to publication to the WWW.

Where do we start ?

61000003.Raw

Date,Time,Pressure (mmHg)	Temperature (C)	,Level Surface Elevation (,Pressure (mmHg)	,Depth (m)
09/03/2009,18:00:00,486.573,18.37,84.132,752.497,0			
10/03/2009,00:00:00,486.277,18.37,84.128,752.793,0.004			
10/03/2009,06:00:00,487.368,18.37,84.143,751.742,-0.01			
10/03/2009,12:00:00,487.079,18.37,84.139,751.075,-0.019			
10/03/2009,18:00:00,488.182,18.421,84.154,749.791,-0.037			
11/03/2009,00:00:00,487.136,18.37,84.14,750.459,-0.028			
11/03/2009,06:00:00,487.735,18.421,84.148,750.597,-0.026			
11/03/2009,12:00:00,486.781,18.32,84.135,752.03,-0.006			
11/03/2009,18:00:00,487.256,18.37,84.141,752.173,-0.004			
12/03/2009,00:00:00,486.402,18.37,84.13,753.226,0.01			
12/03/2009,06:00:00,486.966,18.421,84.137,752.921,0.006			
12/03/2009,12:00:00,486.529,18.37,84.131,753.179,0.009			

61000009.Raw

Date,Time,Pressure (mmHg)	Temperature (C)	,Level Surface Elevation (,Pressure (mmHg)	,Depth (m)
09/03/2009,16:00:00,656.987,17.612,87.148,753.224,0			
09/03/2009,22:00:00,657.022,17.511,87.148,754.411,0.016			
10/03/2009,04:00:00,657.88,17.561,87.16,753.054,-0.002			
10/03/2009,10:00:00,657.527,17.41,87.155,752.658,-0.008			
10/03/2009,16:00:00,657.935,17.359,87.161,751.102,-0.029			
10/03/2009,22:00:00,657.759,17.561,87.159,751.904,-0.018			
11/03/2009,04:00:00,658.357,17.511,87.167,751.093,-0.029			
11/03/2009,10:00:00,657.467,17.511,87.155,753.181,-0.001			
11/03/2009,16:00:00,658.095,17.511,87.163,753.3,0.001			
11/03/2009,22:00:00,657.022,17.46,87.148,754.722,0.02			
12/03/2009,04:00:00,657.727,17.46,87.158,753.942,0.01			
12/03/2009,10:00:00,657.277,17.46,87.152,755.14,0.026			

61000014.Raw

Date,Time,Pressure (mmHg)	Temperature (C)	,Level Surface Elevation (,Pressure (mmHg)	,Depth (m)
09/03/2009,17:00:00,499.161,17.528,94.151,752.661,0			
09/03/2009,23:00:00,499.046,17.478,94.149,753.351,0.009			
10/03/2009,05:00:00,499.529,17.528,94.156,752.173,-0.007			
10/03/2009,11:00:00,499.187,17.528,94.151,751.799,-0.012			
10/03/2009,17:00:00,499.338,17.528,94.153,750.495,-0.029			
10/03/2009,23:00:00,499.242,17.528,94.152,750.829,-0.025			
11/03/2009,05:00:00,499.797,17.478,94.16,750.73,-0.026			
11/03/2009,11:00:00,498.505,17.528,94.142,752.72,0.001			
11/03/2009,17:00:00,498.7,17.528,94.145,752.883,0.003			
11/03/2009,23:00:00,498.843,17.478,94.147,753.594,0.013			
12/03/2009,05:00:00,498.808,17.528,94.146,753.451,0.011			
12/03/2009,11:00:00,498.288,17.478,94.139,754.029,0.019			
12/03/2009,17:00:00,498.907,17.528,94.148,753.074,0.006			
12/03/2009,23:00:00,498.245,17.478,94.139,754.749,0.028			
13/03/2009,05:00:00,498.439,17.478,94.141,755.51,0.039			
13/03/2009,11:00:00,498.304,17.528,94.139,757.656,0.068			
13/03/2009,17:00:00,497.907,17.478,94.134,757.793,0.07			
13/03/2009,23:00:00,497.134,17.528,94.123,759.583,0.094			

Date,Time,Pressure (mmHg)

Temperature (C)	,Level Surface Elevation (,Pressure (mmHg)	,Depth (m)
09/03/2009,17:00:00,552.609,17.828,89.26,752.495,0		
09/03/2009,23:00:00,552.138,17.828,89.254,753.338,0.011		
10/03/2009,05:00:00,553.636,17.828,89.274,751.84,-0.009		
10/03/2009,11:00:00,552.764,17.828,89.262,751.484,-0.014		
10/03/2009,17:00:00,553.51,17.828,89.272,750.217,-0.031		
10/03/2009,23:00:00,553.08,17.828,89.266,750.76,-0.024		
11/03/2009,05:00:00,553.991,17.828,89.279,750.517,-0.027		
11/03/2009,11:00:00,552.656,17.828,89.261,752.188,-0.004		
11/03/2009,17:00:00,552.767,17.828,89.262,752.561,0.001		
11/03/2009,23:00:00,552.776,17.828,89.262,753.407,0.012		
12/03/2009,05:00:00,553.246,17.778,89.269,752.99,0.007		
12/03/2009,11:00:00,552.053,17.828,89.252,753.683,0.016		
12/03/2009,17:00:00,552.532,17.828,89.259,753.056,0.008		
12/03/2009,23:00:00,552.187,17.828,89.254,754.555,0.028		
13/03/2009,05:00:00,552.117,17.828,89.253,755.294,0.038		
13/03/2009,11:00:00,551.299,17.828,89.242,757.6,0.069		
13/03/2009,17:00:00,550.925,17.828,89.237,757.826,0.073		
13/03/2009,23:00:00,550.614,17.828,89.233,759.031,0.089		
14/03/2009,05:00:00,551.012,17.828,89.238,758.595,0.083		
14/03/2009,11:00:00,551.331,17.828,89.243,757.904,0.074		
14/03/2009,17:00:00,552.218,17.828,89.255,754.747,0.031		

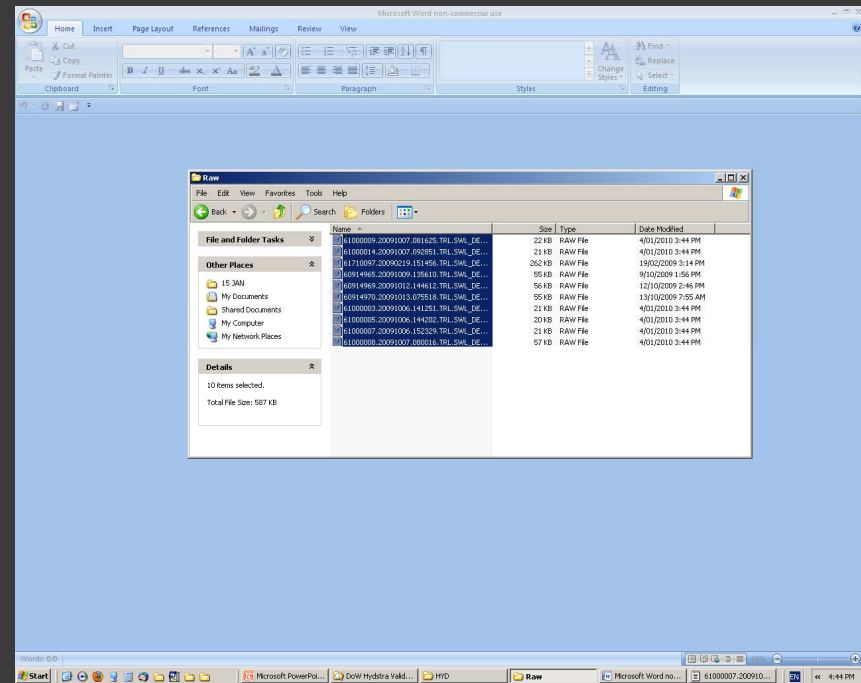
Typically we have ...

- Raw Data Files
- Field Calibration Readings
- Quality Coding & Standards Criteria
- Data Commenting procedure
- All aimed at delivering datasets to an Archive that is fully QA'd

For data downloaded in the field, most of the above is undertaken manually

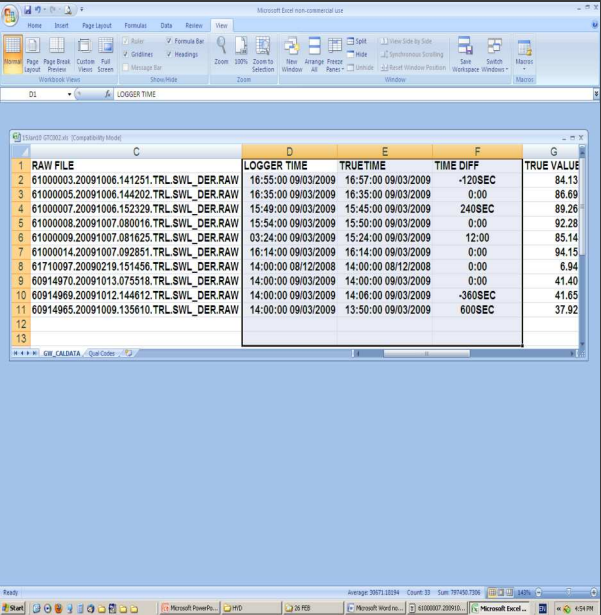
WA 4.4 intending to :

- Import multiple raw files ... (as presently occurs for telemetered sites)



Time Checks

- Automatically calculate logger vs true time differences
- Compare differences to preset criteria
- If criteria is satisfied, make the adjustments

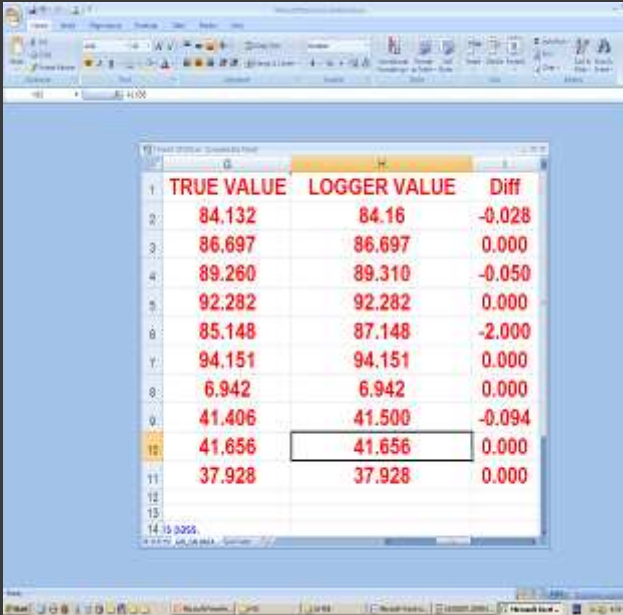


The screenshot shows a Microsoft Excel spreadsheet with the following data:

	C	D	E	F	G
	RAW FILE	LOGGER TIME	TRUETIME	TIME DIFF	TRUE VALUE
1					
2	61000003.20091006.141251.TRL_SWL_DER.RAW	16:55:00 09/03/2009	16:57:00 09/03/2009	-120SEC	84.13
3	61000005.20091006.144202.TRL_SWL_DER.RAW	16:35:00 09/03/2009	16:35:00 09/03/2009	0:00	86.69
4	61000007.20091006.152329.TRL_SWL_DER.RAW	15:49:00 09/03/2009	15:45:00 09/03/2009	240SEC	89.26
5	61000008.20091007.080016.TRL_SWL_DER.RAW	15:54:00 09/03/2009	15:50:00 09/03/2009	0:00	92.28
6	61000009.20091007.081625.TRL_SWL_DER.RAW	03:24:00 09/03/2009	15:24:00 09/03/2009	12:00	85.14
7	61000014.20091007.092851.TRL_SWL_DER.RAW	16:14:00 09/03/2009	16:14:00 09/03/2009	0:00	94.15
8	61710097.20090219.151456.TRL_SWL_DER.RAW	14:00:00 08/12/2008	14:00:00 08/12/2008	0:00	6.94
9	60914970.20091013.075518.TRL_SWL_DER.RAW	14:00:00 09/03/2009	14:00:00 09/03/2009	0:00	41.40
10	60914969.20091012.144612.TRL_SWL_DER.RAW	14:00:00 09/03/2009	14:06:00 09/03/2009	-360SEC	41.65
11	60914965.20091009.135610.TRL_SWL_DER.RAW	14:00:00 09/03/2009	13:50:00 09/03/2009	600SEC	37.92
12					
13					

Value Checks

- Automatically calculate logger vs check reading differences
- Compare differences to preset criteria
- If criteria is satisfied, automatically make the adjustments



The screenshot shows a software window with a data table. The table has three columns: TRUE VALUE, LOGGER VALUE, and Diff. The data is as follows:

	TRUE VALUE	LOGGER VALUE	Diff
1	84.132	84.16	-0.028
2	86.697	86.697	0.000
3	86.697	86.697	0.000
4	89.260	89.310	-0.050
5	92.282	92.282	0.000
6	92.282	92.282	0.000
7	85.148	87.148	-2.000
8	94.151	94.151	0.000
9	6.942	6.942	0.000
10	41.406	41.500	-0.094
11	41.656	41.656	0.000
12	37.928	37.928	0.000

How is this Achieved ?

- ◎ TIME and VALUE CHECKS

- ◎ STORED WITHIN A SIMPLE DATABASE
- ◎ WHERE POSSIBLE, DATA ENTRY TO BE PERFORMED IN THE FIELD
- ◎ THIS AUTO ROUTINE WILL SIMPLY REFER TO THE VALUES

- ◎ QUALITY CODE CRITERIA

- ◎ STORED WITHIN A SIMPLE DATABASE

- ◎ TOTALLY CONFIGURABLE by the USER, or DATA_MANAGER

Process Control

- Database Fields include options for the operator to HALT the auto-validation, based on preset criteria.

PLUS :

The routine will make use of existing HYAUDIT options, through comparisons with your existing Archive

- Absolute MAXIMA & MINIMA ever recorded (or multiple thereof)
- Maximum RATE OF RISE
- Maximum RATE OF FALL
- Use these to auto delete data SPIKES

So far ...

- ⦿ Logger Time within bounds ?
- ⦿ Value checks within bounds ?
- ⦿ If “Yes”, apply adjustments and assign Quality Code
- ⦿ If “No”, HALT auto-processing
- ⦿ Remove single data spikes if they exist

WOULD YOU LIKE TO SEE A
DEMO ?

◎ Review .mdb

◎ Run demo

◎ Review Results

What is intended to be the prime advantage ?

- Validation of time series hydrological data has always demanded very high standards.
- Historically, an experienced operator may be able to complete and archive data from ~10-15 downloads in a day
- For datasets that are “clean” this process is aiming at validation and archival of ~ 30 downloads in 5-10 mins.

Any Questions so far ...

Where to from here ?

- Automate data comments
- Trial other parameters
- Add a “Peak” field to calibration database
- Build a friendly front-end onto calibration database

Additional suggestions welcome