From the Datalogger to Hydra 2: Load Cell Raw data

1) Data extraction form the datalogger at the SSL

The data of the load cells are saved in a datalogger located in the Svartisen Subglacial Laboratory. The data are extracted from the datalogger using the local computer and an old version of the software called PC200w.

The output filename is user-defined and typically looks like *trykk_20mar001.dat*. The user must every time enters in the PC200w the date of download in the above format (i.e. *20mar*).

The datalogger produces sometimes a few files and sometimes many more that succeeds each other as follows: *trykk_20mar001.dat, trykk_20mar002.dat, trykk_20mar003.dat…*

Transfer all files to a transportable hard drive and leave a copy on the local machine.

2) Transfer to NVE server

All the raw data from the load cells are then copied to the NVE local server at
> \\nve\fil\h\HB\Bredata\breprosjekt\Engabreen\Engabreen Brelabben\labb-trykkdata
>   AND are gathered in a folder for each year (the year correspond to the year of data retrieval).

3) Document the time span of each file

The first step is to document what each file contains in the local *Data_list_YEAR.xls*. All files must have a description of their time span (extracted manually from the file) and are ordered chronologically. If possible, indicate in a note beside: file corruption (see note below), duplicates or that the file contains only NaN values.

File name	Day start	Date start	Time start	Day end	Date end	Time end	Data interval
trykk_20mar001.dat	25	25/01/2013	23:23	28	28/01/2013	09:57	2

Note: The date is computed with the excel function: "=DATE(2013,1,0)+B2". B2 is the start day. In this example, it is equal to "25", which gives "25/01/2013".

FILE CORRUPTION
You have to scan the file yourself, as it may be corrupted:
- Lines may be truncated or shifted and contains two dates in one line
- Large blocks of data may suddenly jump back in time (no continuity in the data)
- Compression or font problems causes blocks of data to be unreadable.
You will find some examples of data corruption in the files: trykk_20mar002.dat, trykk_28aug001.dat, and trykk_29nov002.dat

4) Selection of the data files

4.A) Files that cover continuously the span of the year are then selected whereas **4.B)** duplicates are disregarded. This process is documented in the Excel sheet *Data_list_compiled.xls*, which has a sheet made for each year, like the one below.

The final selection is placed on the top of the sheet and filenames are highlighted in green. Files that have overlaps (i.e. duplicates) are placed at the end of the sheet and a note must indicate which selected file (green file) overlaps with the duplicate.

Periods with no data must also be written down (see in the selection part for the period 274-334 in day of the year) and commented.

4.A)

2013							Should ONLY consider files in GREEN	
	Day start	Date start	Time start	Day end	Date end	Time end	Data interval [min]	
trykk2_06jun001.dat	1	01/01/2013	00:01	108	18/04/2013	19:13	2	2012-2013
trykk_06jun002.dat	108	18/04/2013	19:15	156	05/06/2013	23:09	2	
trykk_06jun003.dat	156	05/06/2013	23:11	157	06/06/2013	01:27	2	
trykk_17sep001.dat	157	06/06/2013	01:29	230	18/08/2013	22:53	2	
trykk_17sep002.dat	230	18/08/2013	22:55	259	16/09/2013	12:03	2	
trykk_29nov001.dat	259	16/09/2013	12:05	265	22/09/2013	08:11	2	
Kopi_av_trykk_29nov002.dat	265	22/09/2013	08:13	274	01/10/2013	07:51	2	Corrected - see Note
	274	01/10/2013		334	30/11/2013		NO DATA	< I don't believe that there is
Engabreen_trykk_dec03_new.dat	334	30/11/2013	04:35	336	02/12/2013	15:11	2	not a file with these data.

Added by PiM, 2013.10.04: Original File were edited to keep only data from 2013

trykk2_06jun001.dat	360	25/12/2012	17:33	108	18/04/2013	19:13	2	2012-2013

From the analysis of continuity in the .dat files, we get the following order in the raw data:

For All Files: Code, DayOfYear, HH:MM, LC6 ,LC1e, LC4, LC2a ,**12_1, 12_2**, **LC7**, Battery (once a day)

Note:
> **PiM 2013.12.06: CHECKED! 2min Data in** *Kopi_av_trykk_29nov002.dat* **are from 2013.** But Data from **164 (04:19)** to **204 (19:35)** are **Duplicates** of *trykk_17sep001.dat*.
> **PiM 2013.12.06:** The values between *Kopi_av_trykk_29nov002.dat* & *Engabreen_trykk_dec03_new.dat* show quite a **JUMP.**

Update 2015.02.16: The same one day shift than in 2012 may occur in 2013. **The downloaded files with timestamps like *trykk_17sep002.dat* do not match the last date of the file: 16/09/2013.** In 2012, the datalogger clock was reset wrongly because of for leap years, the day of the year are different. **The data timestamp must be corrected by: +1day and the correction MUST BE APPLIED TO ALL DATA IN 2013.**

4.B)

			DUPLICATES					
trykk_19apr001.dat	320	15/11/2012	17:25	320	15/11/2012	20:11	2	
trykk_19apr002.dat	320	15/11/2012	20:13	46	15/02/2013	16:43	2	
trykk2_19apr001.dat	320	15/11/2012	20:13	107	17/04/2013	327	2	Duplicate IN
Engabreen_trykk_29jan.dat	25	25/01/2013	23:21	28	28/01/2013	10:51	2	*trykk2_06jun001.dat*
Engabreen_trykk_12feb.dat	40	09/02/2013	04:23	42	11/02/2013	14:59	2	
Engabreen_trykk_11mars.dat	67	08/03/2013	00:29	69	10/03/2013	11:05	2	
Engabreen_trykk_26apr.dat	112	22/04/2013	22:03	115	25/04/2013	08:39	2	Duplicate IN
Engabreen_trykk_june04.dat	152	01/06/2013	02:23	154	03/06/2013	12:59	2	*trykk_06jun002.dat*
Kopi av trykk_29nov002.dat	265	22/09/2013	08:13	274	01/10/2013	07:51	2	Some Dupl. IN
								trykk_17sep001.dat
Engabreen_trykk_sep01.dat	241	29/08/2013	03:31	243	31/08/2013	14:07	2	Dupl. IN trykk_17sep002.dat

5) File Backup

Before any editing, the files must be copied to two folders:
1) All the files (selected and duplicates) must be copied in ***Raw_BackUp.***
2) Only the selected files (green files) must be copied to ***Raw.***

6) File Editing

The files that are contained in the folder ***Raw*** have to be manually edited to remove corrupted lines and data that spans over the year of interest. Any editing must be

documented in *Data_list_compiled.xls* either in a note on the line of the edited file or after the selection part (with the green files). Below, examples of editing are given:

6.A) It is common that the first line of the file gets corrupted. This line can be removed or if the data are still consistent but misses the day of the year, can be corrected.

6.B) A classic file correction is the removal of data from the year before or the year after. The selected files (in ***Raw***) must only contain data from their year as shown below.

This implies that a file with data covering two years has two be copied to the folders of the two respective years. For example, to add data from the end of 2013, I had to copy the file *trykk_20mar004.dat* from the folder 2014 into the folder 2013 and then removed data from 2014.

2013			**Should ONLY consider files in GREEN**					
	Day start	Date start	Time start	Day end	Date end	Time end	Data interval [min]	
trykk2_06jun001.dat	1	01/01/2013	00:01	108	18/04/2013	19:13	2	**2012**-2013
trykk_06jun002.dat	108	18/04/2013	19:15	156	05/06/2013	23:09	2	
trykk_20mar004.dat	332	28/11/2013	10:57	365	31/12/2013	23:59	2	2013-**2014**
Added by PiM, 2013.10.04: Original File were edited to keep only data from 2013								
trykk2_06jun001.dat	360	25/12/2012	17:33	108	18/04/2013	19:13	2	**2012**-2013
trykk_20mar004.dat	332	28/11/2013	10:57	78	19/03/2014	10:05	2	2013-**2014**

This edit must be documented below the selection part as shown above.

6.C) Any other consequent editing must be indicated in the Note section, such as:

Note:
> **PiM 2016.05.03: Some data in** *trykk_29nov002.dat* overlap with data from *trykk_17sep001.dat* between **164 (04:19) - 204 (19:35)** and from *trykk_20mar002.dat* between **273 (21:59) - 274 (07:51).** These periods were deleted in *trykk_29nov002.dat*.

6) Time correction

It has happened that the clock of the datalogger shows a drift or is erroneous, which led a user to reset the clock. The *Note* section must contain observations of these errors/clock reset. Any corrections applied to the dataset have to be documented too.

7) Continuity in the dataset and Column headers

When the data logger has its load cell connections rewired or a new load cell is added, it is important to indicate the change in load cell as well as the old and new order of the column (with their headers).

From the analysis of continuity in the .dat files, we get the following order in the raw data:

For All Files: Code, DayOfYear, HH:MM, LC6 ,LC1e, LC4, LC2a ,**12_1, 12_2**, **LC7**, Battery (once a day)

8) Reordering, compiling and formatting for Hydra2

When editing of the files is done and files are saved in the **Raw** folder, a series of R programs will process and homogenise the load cell data. The three programs are summarised in the R file *CodeSummary_and_FinalTransfer.R* contained in **Raw**. **If the wiring/order of the load cell is identical, the only change to do is to change the year that needs to be processed** (line 44) and then run in R:

> source('CodeSummary_and_FinalTransfer.R')

Explanation of the three programs:
8.A) The code *Raw_editing.R* will homogenise the order of the load cell such as it follows:
New Order   Year - DoY - Hr:Mn - 6 - 1e - 4 - 2a - 97_2 - 97_1 - 7 - 2b - 01 - Battery
(DoY= Day of the Year, Hr:Mn = Hour Minute)

Each year or even each file is registered with the order identified from notes or the load cell records and transformed to fit the order above. All reordered files are then saved in a folder **OrderedR** for each year.

> **Add a comma at the end of the first line of each file** in the **Raw** folder. This is because the voltage is only recorded once a day, and R will not detect this column if it is not present on the first line causing an error message to stop this code.

8.B) *Raw_merging.R* compiles all the reordered files per year and saves the output in *Compiled_YEAR.dat* contained in a folder **ProcessingR** for each year.

8.C) The conversion of the compiled data into Hydra2 format is done using the R code *format4Hydra2.R*. It edits the date format, NaNs and extract each load cell separately so that the load cell data can be imported in Hydra II. The output is saved in a folder **ProcessingR/Hydra2** for each year.

8.D) *CodeSummary_and_FinalTransfer.R* also copies the final compiled outputs (**8.B**) to the **../Processing** directory (same level than the folder **Raw** containing all raw data) where further data analyses are performed.

9) Upload and edit and control data on the NVE server

The output files from step **8.C)** are then sent to Eva Klausen (eva@nve.no) that has the level of privileges to upload data into Hydra2.

Make sure that you have access to the *Hysopp* database in order to edit metadata and also privileges in order to control the data in *Hykon* (meaning you should have privileges of the group "hyk"). Contact Vidar Eide (vei@nve.no) if you lack any.

9.A) Hysopp is a database or SQL software that is used to edit and check metadata. First, go to the top panel and click on Stasjon>Stasjonsopplysninger. Enter station number of the load cells: vassdragnr: 151, Hovednr: 64, Punktnr: [1-11]. It should look like below for Engabreen LC1e (151.64.1).

You can edit the kommentær, location, time, observasjonssted and Seksjon sections. Editing of Status and Stasjonstyper have to be sent to Eva Klausen.

9.B) The uploaded data are visible in the *Arkiver* section highlighted in red above. The raw data are saved in the *Hytran* database that conserves all raw data and after being controlled it will be transferred to the *Hykval* database, as shown below.

9.C) The data are controlled in the software Hykon. Data can be listed using the search engine and Stasjonsnummer. The type of data (Hytran or Hykval) is chosen in the top-right corner. To control the Hytran data, select one or several Hytran files (for that tick "Velg flere perioder" on the left) and click *Plotte*.
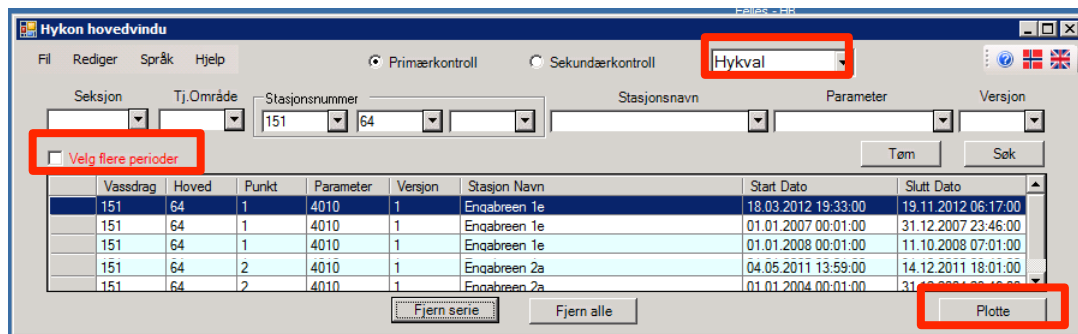

In this window, you can remove points, insert missing values & correct time/data. First remove too low values using the scissor tool on the top panel and entering in "Mindre enn" a value 100Hz lower than the frequency zero of the load cell. This will remove a large part of erroneous data.

To see better errors, you can activate the option in Vis>Linje gjennom missing. Manually zoom in, select data that you want to erase and in the table view replace the selected values by missing (right-click and then missing option).

A common problem is to add missing values when data follow each other although they have a gap. Right-click the last point before the gap and choose insert. Add one hour to the time, enter "missing" in the value box, click the green right arrow and finally sett in.

When the data are controlled, save (red box) the data in Hykval(Default).
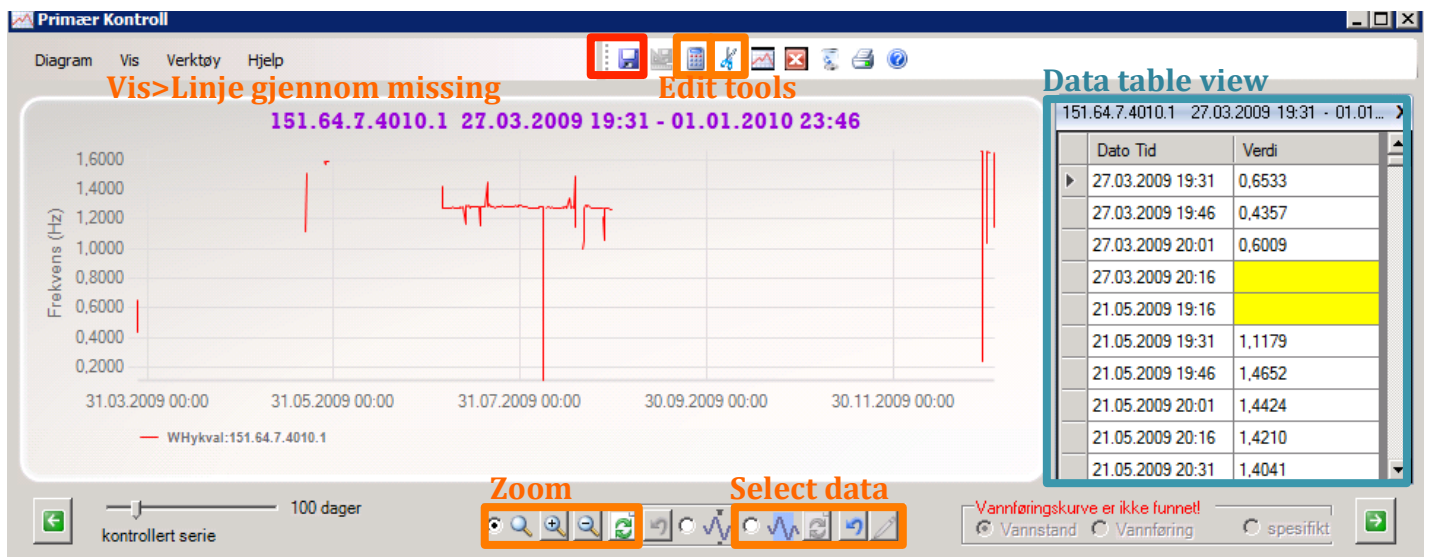
9.D) Verify that the controlled data are good using the Hydra2 software *finut*. If you spot any problem, go back to Hykon.