



Natural Resources
Canada

Ressources naturelles
Canada

Data Processing Methodology for Field Test Data

Zhenhuan Xu (CanmetMINING)
October 7, 2026

Data is an asset

Mining data provides information to the following areas:

Productivity

MEASURED AS

Cycle times, effective
time per cycle

Energy

MEASURED AS

Fuel and kWh per lap
and per tonne

Health and Safety

MEASURED AS

Vibration, temperature,
RH, pressure and noise
level

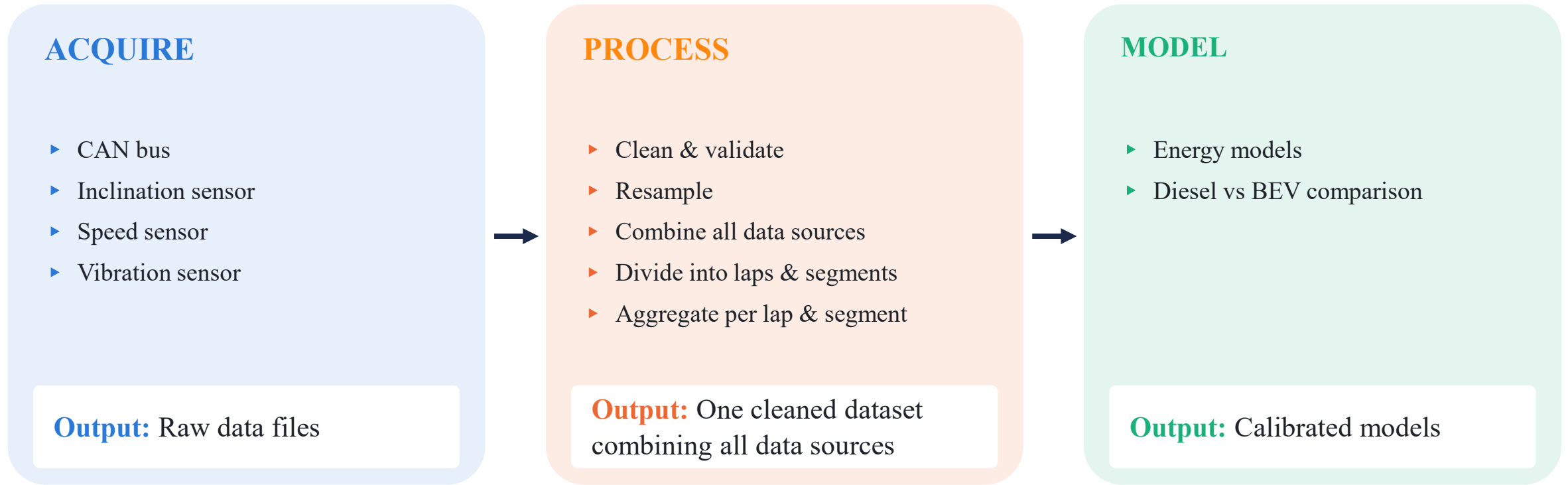
Design

MEASURED AS

Calibrated models based
on test data



Workflow: Acquire – Process – Model



“Garbage in, garbage out.” A model is only as good as the dataset it is built on.



Natural Resources
Canada

Ressources naturelles
Canada

General Steps for Data Processing

1

Extract

Extract the required data from the raw data file. This may include parsing the data, converting data types, and performing other necessary data preprocessing.

2

Clean

Remove duplicates, identify and handle missing values, and detect invalid values.

3

Resample

Sensors may record data from 10 Hz to a few thousand Hz, while 1 Hz is generally sufficient for energy calculations. Data may need to be resampled to a common frequency.

4

Join

Combine machine data with data collected by CanmetMINING data logger. Timestamps from different sensors may need to be aligned before joining.

5

Aggregate

Calculate statistical features such as average, RMS, minimum, maximum, and sum for energy, power, fuel consumption, speed, and vibration by lap and segment.



Example: Raw Data

Machine Telemetry Data

local_time	raw_data		
2025-02-12 17:58:42+00	{"BEVFaults": 0	"BoxVoltage": 25.147055061534047	"BEVWarnings": 0
2025-02-12 17:58:41+00	{"BEVFaults": 0	"BoxVoltage": 25.20038270391524	"BEVWarnings": 0
2025-02-12 17:58:40+00	{"BEVFaults": 0	"BoxVoltage": 25.20038270391524	"BEVWarnings": 0
2025-02-12 17:58:39+00	{"BEVFaults": 0	"BoxVoltage": 25.240378435701132	"BEVWarnings": 0
2025-02-12 17:58:38+00	{"BEVFaults": 0	"BoxVoltage": 25.173718882724643	"BEVWarnings": 0
2025-02-12 17:58:37+00	{"BEVFaults": 0	"BoxVoltage": 25.160386972129345	"BEVWarnings": 0
2025-02-12 17:58:36+00	{"BEVFaults": 0	"BoxVoltage": 25.160386972129345	"BEVWarnings": 0
2025-02-12 17:58:35+00	{"BEVFaults": 0	"BoxVoltage": 25.20038270391524	"BEVWarnings": 0
2025-02-12 17:58:34+00	{"BEVFaults": 0	"BoxVoltage": 25.160386972129345	"BEVWarnings": 0
2025-02-12 17:58:33+00	{"BEVFaults": 0	"BoxVoltage": 25.147055061534047	"BEVWarnings": 0
2025-02-12 17:58:32+00	{"BEVFaults": 0	"BoxVoltage": 25.160386972129345	"BEVWarnings": 0
2025-02-12 17:58:31+00	{"BEVFaults": 0	"BoxVoltage": 25.160386972129345	"BEVWarnings": 0

1 Hz, JSON data format

Inclination Sensor Data

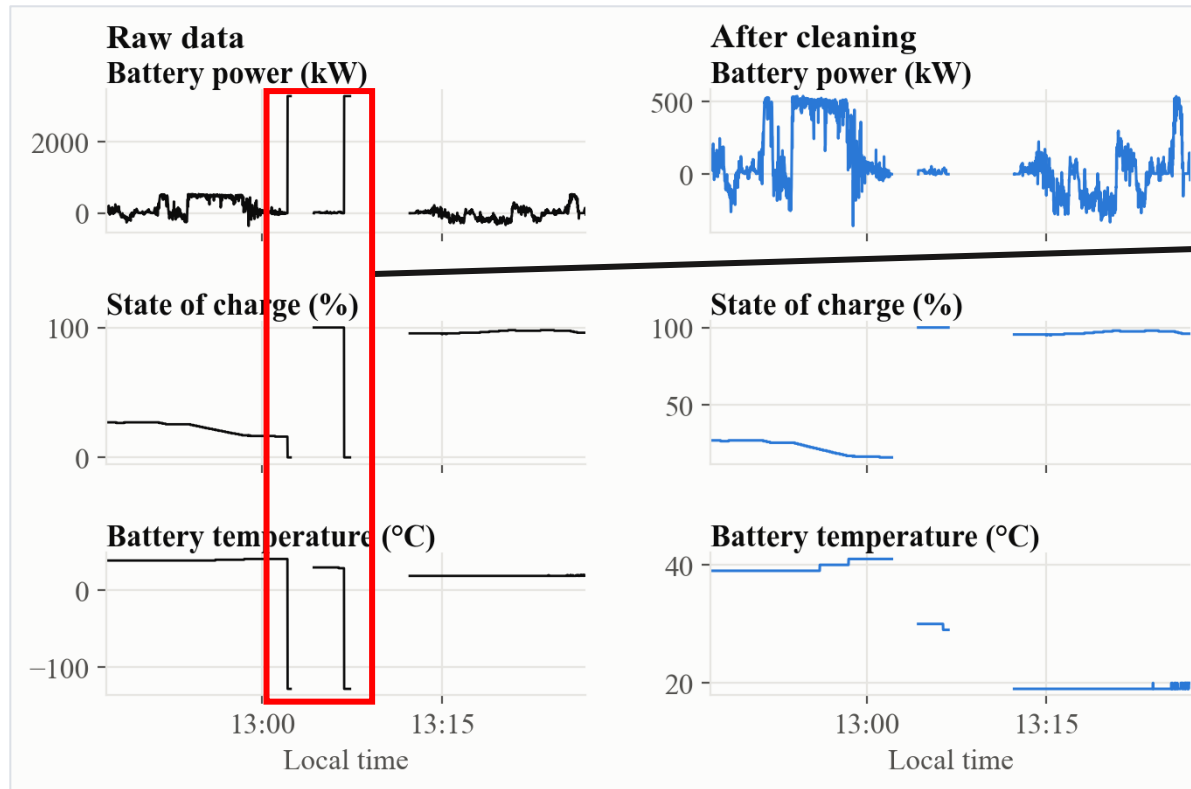
A	B	C	D	E	F	G	H	I
Timestamp	Accelerati	Accelerati	Angle_X_D	Angle_Y	Source File			
9:11:24	-0.055	-0.001	3.09	0.13	20250212074021_PCAN_USBBUS1_1.trc			
9:11:24	-0.054	-0.001	3.09	0.13	20250212074021_PCAN_USBBUS1_1.trc			
9:11:24	-0.054	0	3.09	0.13	20250212074021_PCAN_USBBUS1_1.trc			
9:11:24	-0.049	-0.001	3.09	0.13	20250212074021_PCAN_USBBUS1_1.trc			
9:11:24	-0.052	-0.001	3.09	0.13	20250212074021_PCAN_USBBUS1_1.trc			
9:11:24	-0.052	-0.002	3.09	0.13	20250212074021_PCAN_USBBUS1_1.trc			
9:11:24	-0.053	-0.002	3.09	0.13	20250212074021_PCAN_USBBUS1_1.trc			
9:11:24	-0.055	-0.001	3.09	0.13	20250212074021_PCAN_USBBUS1_1.trc			
9:11:24	-0.055	-0.003	3.09	0.13	20250212074021_PCAN_USBBUS1_1.trc			

100 Hz, simple numerical values

Multiple data sources, one test: for a single test, data from multiple sources needs to be combined into one dataset. Each data source must be processed and aligned before the datasets can be merged.



Example: Data Cleaning



CAN 'not available' error: power 3,276,800 W (largest signed 16-bit integer x 100), SOC 0 %, temperature -128 °C (smallest signed 8-bit integer).

Erroneous values are excluded from calculations to prevent them from distorting the results.

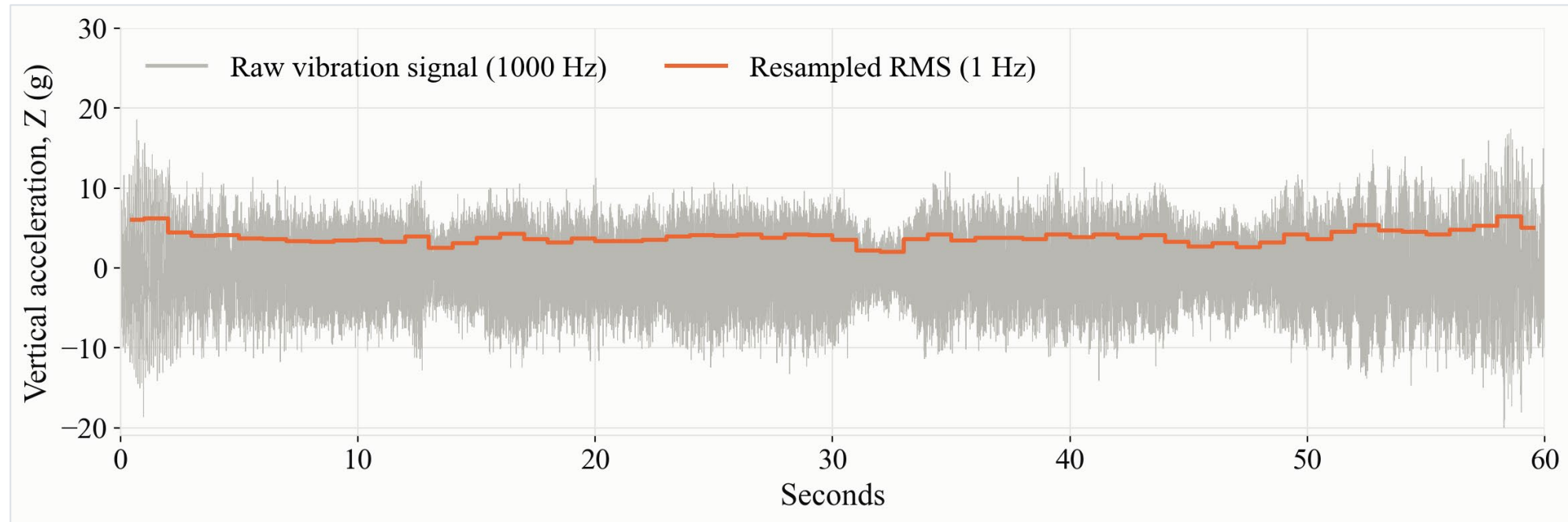
Sometimes numerical anomalies can indicate a change in operating conditions; in this case, they indicate a battery swap.



Natural Resources
Canada

Ressources naturelles
Canada

Example: Resampling



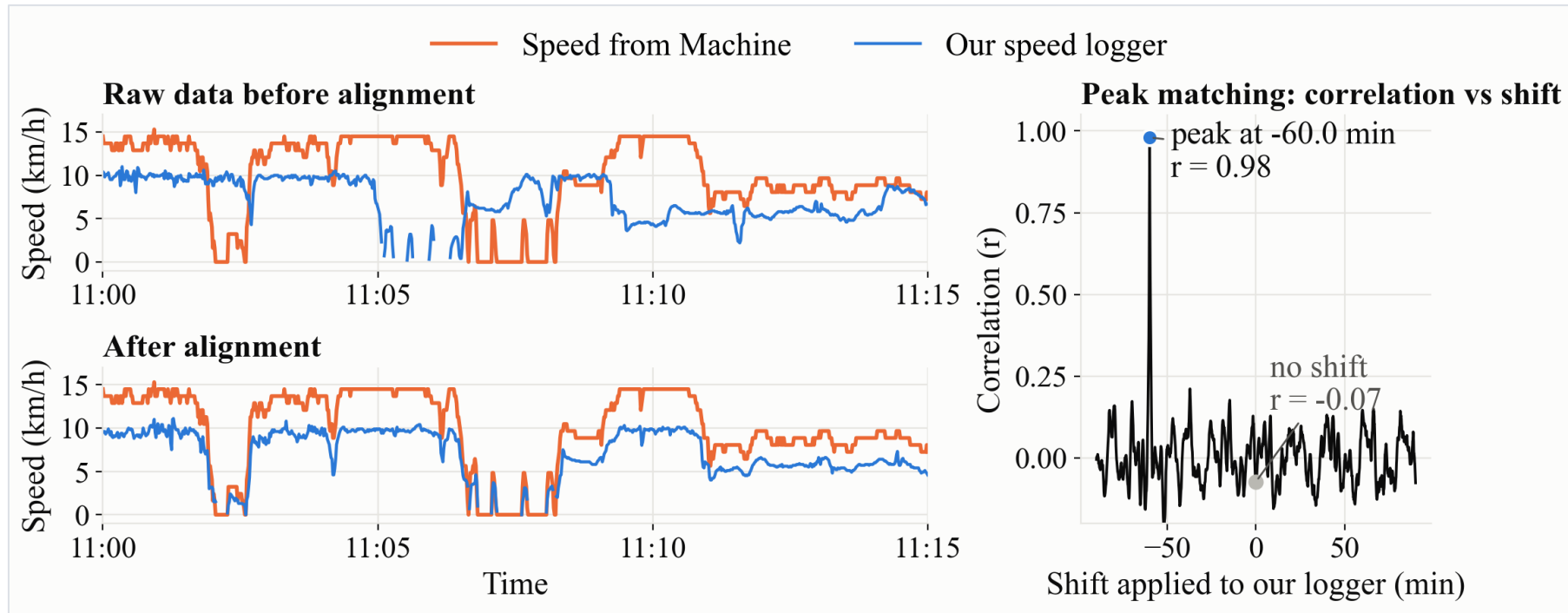
During resampling, data such as speed, inclination, and machine parameters (e.g., voltage and current) are typically averaged, while vibration data are represented using RMS values.



Natural Resources
Canada

Ressources naturelles
Canada

Example: Timestamp Alignment



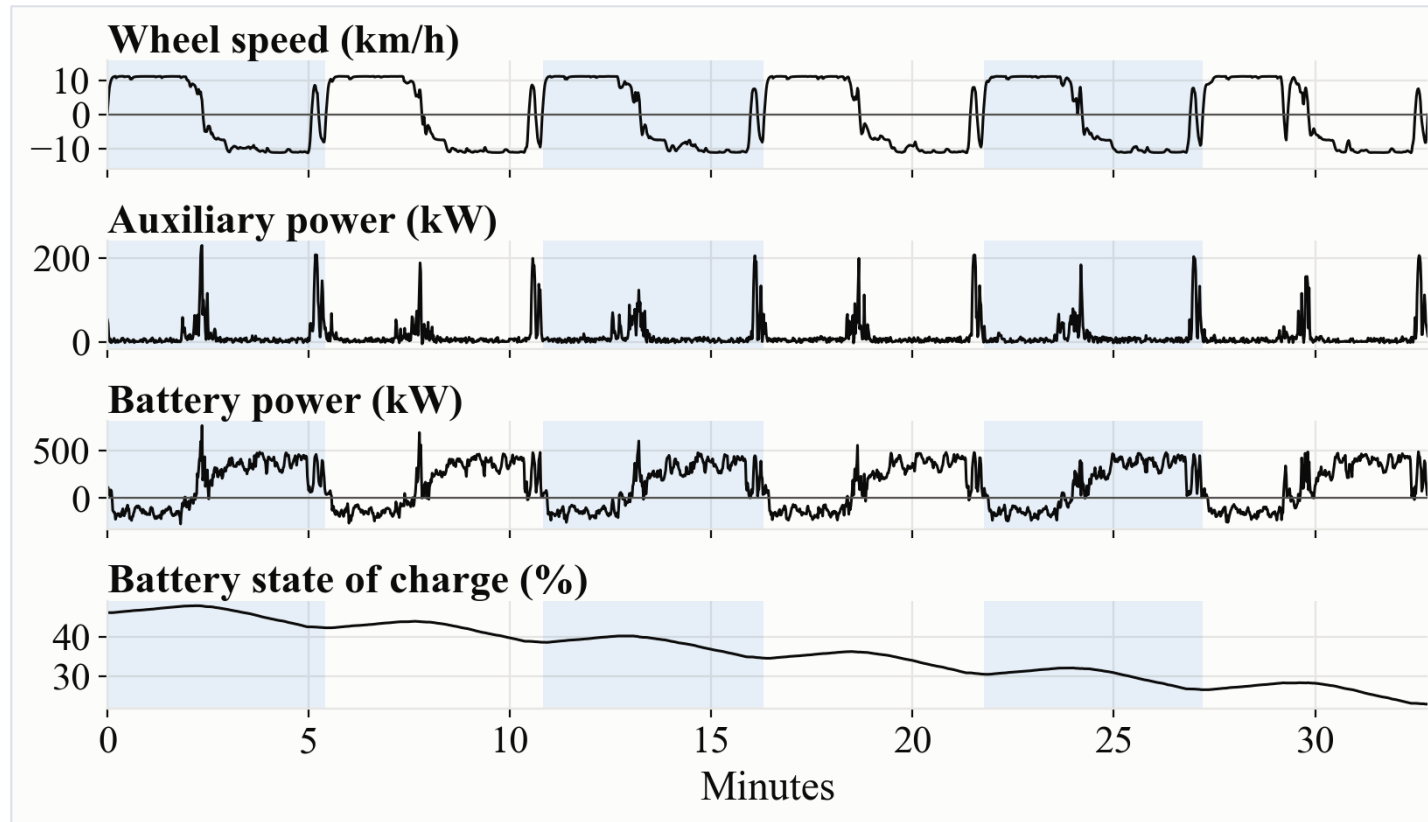
Compare the same variable across devices to identify and correct time offsets between them. Event-based alignment can also be used when no common variable is available across devices.



Natural Resources
Canada

Ressources naturelles
Canada

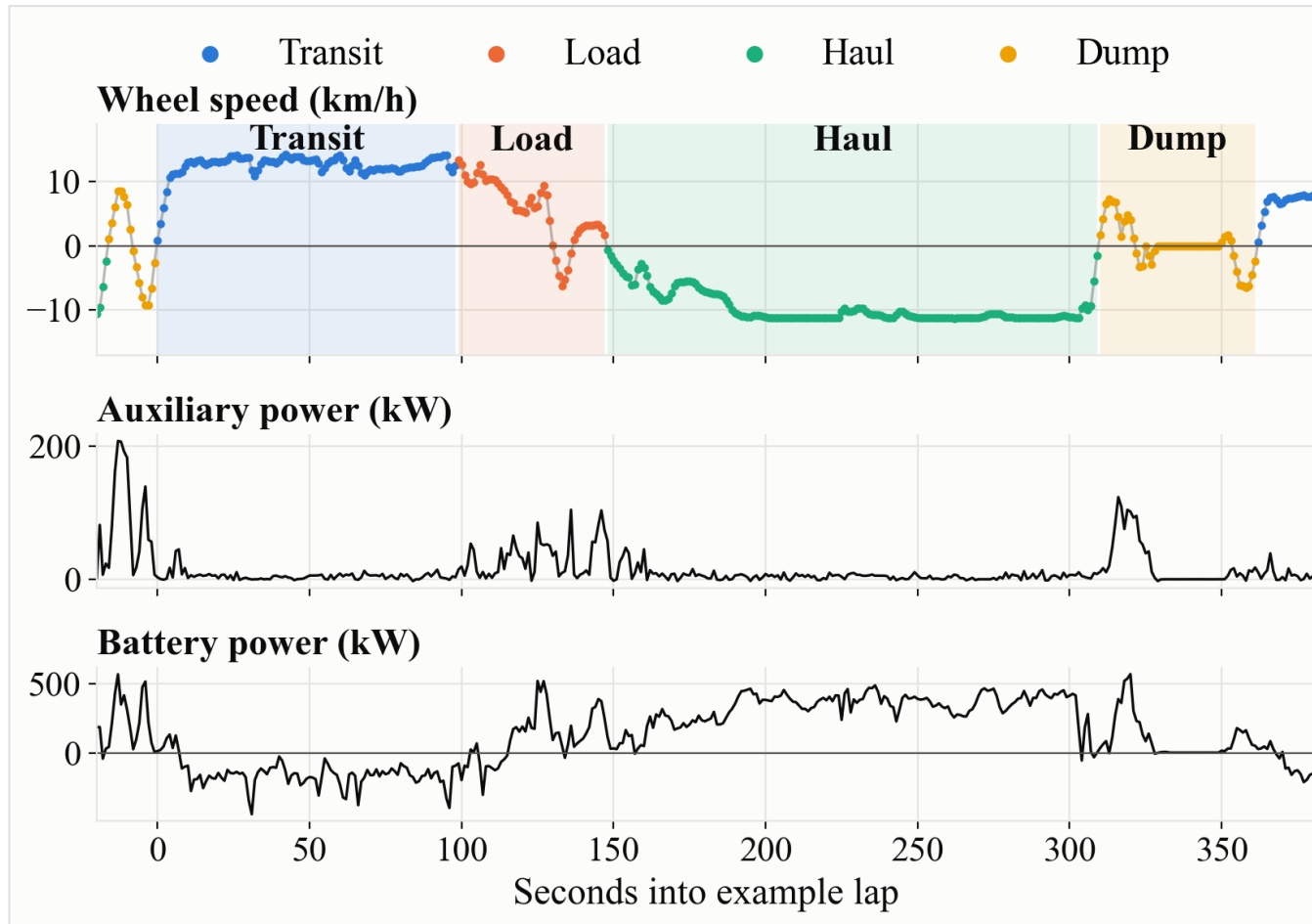
Example: Repeatable Cycle



- ▶ Speed, auxiliary power, battery power and state of charge all carry the same pattern.
- ▶ Rule-based classification can be applied since the pattern is clear.



Example: Rule-based Segment Classification



Example rules used to identify the four segments:

- ▶ **Transit:** Starts with speed zero crossing. Ends at a speed plateau followed by a sustained slowdown. Small auxiliary power.
- ▶ **Load:** Starts after Transit. Ends the second speed passes down through zero into reverse. Increased auxiliary power.
- ▶ **Haul:** Starts after Load. Ends at the next zero crossing with no sustained negative speed after it. Small auxiliary power.
- ▶ **Dump:** Starts after Haul. Ends where the next Transit begins. Increased auxiliary power.



Data Processing Streamlined

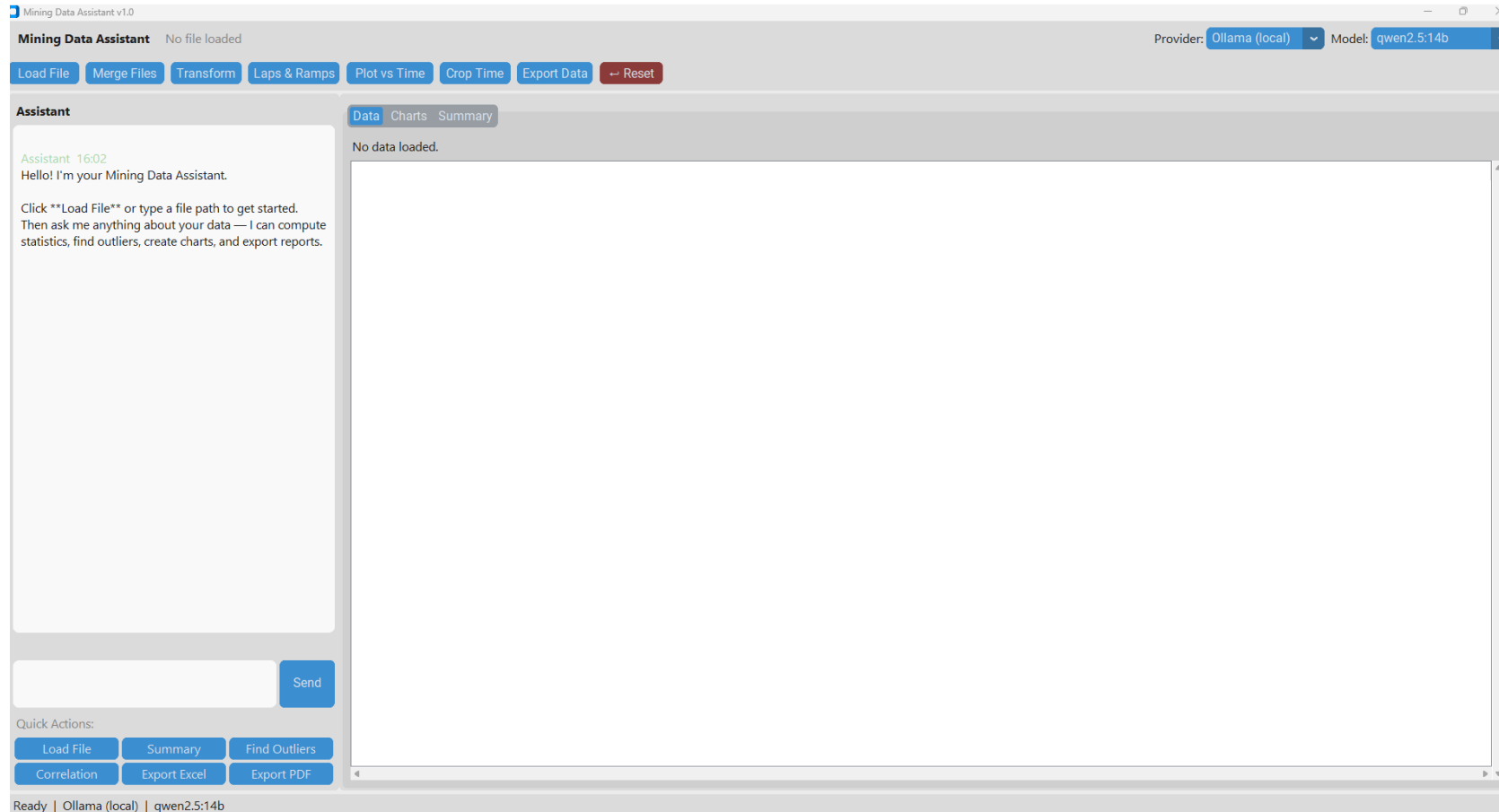


Can we streamline data processing?

- ▶ We use Python for data processing. At this stage, a new script must be developed or an existing script modified for each test, since the data format, sampling frequency, variables of interest, and time-alignment requirements can vary significantly between tests.
- ▶ However, the overall data-processing workflow is similar across tests, making it well suited for a more automated and streamlined approach.



Data Processing Assistant



► **Idea:** Stop writing a new script for each test. Instead, build a user-interface tool that takes user inputs, reads the data, automatically processes the data and exports the processed data to Excel or .CSV.

► **AI assistant (in development):** An optional AI assistant that can answer questions about the dataset, perform tasks and explain how the tool works.



Natural Resources
Canada

Ressources naturelles
Canada

Data Processing Assistant – Merge

Merge Files

Step 1 of 4 — Assign Files to Categories

Machine Data + Add Files

- machine data.xlsx

Sensor Data + Add Files

- speed_1.csv
- speed_2.csv
- speed_3.csv
- speed_4.csv
- inclination.csv

Vibration Data + Add Files

- vibration_1.csv
- vibration_2.csv

Next: Select Time Columns →

Merge Files

Step 2 of 4 — Select Time Columns

Select the time column for each file (files in the same category may use different column names):

Machine Data

machine data.xlsx timestamp

Sensor Data

speed_1.csv Time (hh:mm:ss)

speed_2.csv Time (hh:mm:ss)

speed_3.csv Time (hh:mm:ss)

speed_4.csv Time (hh:mm:ss)

inclination.csv Timestamp

Vibration Data

vibration_1.csv Time

vibration_2.csv Time

← Back Merge & Resample to 1 s →

► Import all data into three categories. Vibration data must be imported under the correct category because it is resampled using RMS rather than an average.

► The time column is important because it is used for timestamp alignment across different data sources.



Data Processing Assistant – Merge

Merge Files

Step 3 of 4 — Cross-Category Time Synchronisation

Resampled results:

Machine: 9,001 rows 2026-03-19 12:30 → 2026-03-19 15:00
Sensor: 15,060 rows 2026-03-19 11:23 → 2026-03-19 15:34
Vibration: 8,278 rows 2026-03-19 16:34 → 2026-03-19 18:52

Data frequency (records/second):

[Sensor] inclination.csv

View Chart

Do all categories share the same timezone / timestamps?

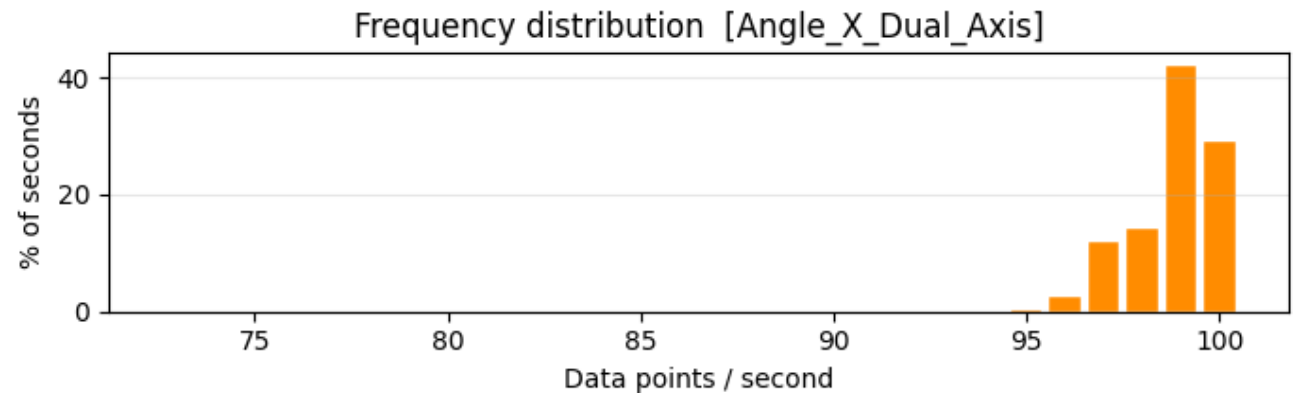
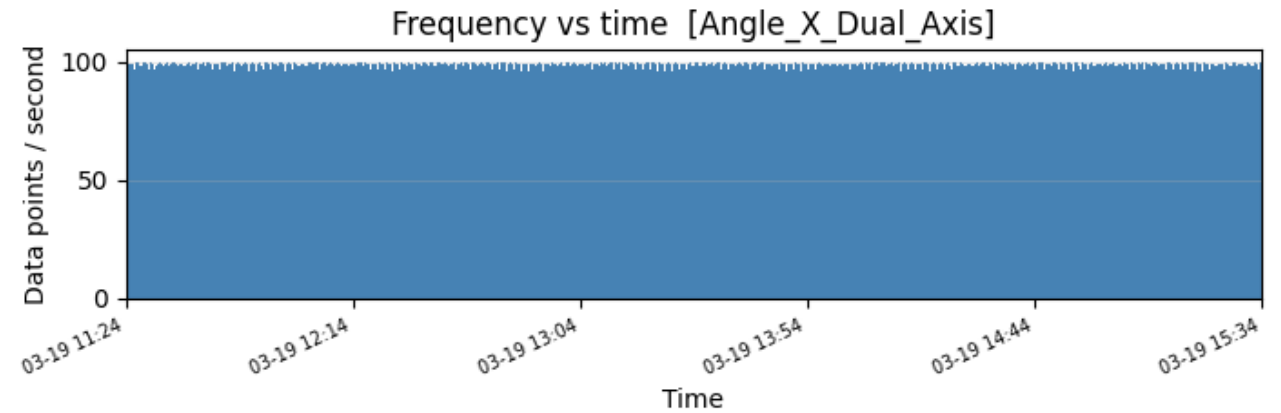
☒ Yes — merge directly on shared time values
☐ No — I need to adjust time offsets between categories

Reference time (this category's time column is used in the final output):

☒ Machine ☐ Sensor ☐ Vibration

☐ Crop output to a specific time window

[← Back](#) [Merge All & Load →](#)



Data Processing Assistant – Merge

Merge Files

Step 4 of 4 — Select Columns to Keep

Merged dataset: 23,338 rows × 199 columns — tick the columns you want to keep.

Select Visible **Deselect Visible**

- ☒ radar_bypass_mode (float64)
- ☒ radar_proximity_sensor_distance (float64)
- ☒ modbusBridgeState (float64)
- ☒ record_disk_monitor_free_space_gb (float64)
- ☒ Frequency (Hz) (float64)
- ☒ Speed (km/h) (float64)
- ☒ Acceleration_X (float64)
- ☒ Acceleration_Y (float64)
- ☒ Angle_X_Dual_Axis (float64)
- ☒ Angle_Y (float64)
- ☒ RMS_ "X (40g)" (float64)
- ☒ RMS_ "Y (40g)" (float64)
- ☒ RMS_ "Z (40g)" (float64)

← Back **Load →**

Mining Data Assistant merged_data.csv | 23,338 rows × 12 cols

Load File **Merge Files** **Transform** **Laps & Ramps** **Plot vs Time** **Crop Time** **Export Data** **↶ Reset**

Assistant

Data **Charts** **Summary**

Showing 500 of 23,338 rows | 12 columns

Date	Time	batterie	battery_battery_rx_output	battery_battery_rx_pack	Speed (km/h)	Acceleration_X (g)	Acceleration_Y (g)	Angle_X_Dual_Axis (deg)	Angle_Y (deg)	RMS_X (40g)	RMS_Y (40g)	RMS_Z (40g)
2026-03-19	11:23:21				0.0							
2026-03-19	11:23:22				0.0							
2026-03-19	11:23:23				0.0							
2026-03-19	11:23:24				0.0							
2026-03-19	11:23:25				0.0							
2026-03-19	11:23:26				0.0							
2026-03-19	11:23:27				0.0							
2026-03-19	11:23:28				0.0							
2026-03-19	11:23:29				0.0							
2026-03-19	11:23:30				0.0							
2026-03-19	11:23:31				0.0							

- ▶ Users can select the columns they want to keep.
- ▶ Data processing progress is displayed in Assistant.



Data Processing Assistant – Segment and Aggregate

Laps & Ramps

Define Segments Aggregate

Data range: 2026-03-19 11:23:21 → 2026-03-19 18:52:02
Formats accepted: YYYY-MM-DD HH:MM:SS | HH:MM:SS | H:MM AM/PM

LAPS

Import from Image/PDF... Import from Excel... + Add Lap

Name	Start	End	
Lap 1	2026-03-19 12:39:40	2026-03-19 12:46:39	×
Lap 2	2026-03-19 12:46:54	2026-03-19 12:50:49	×
Lap 3	2026-03-19 12:51:01	2026-03-19 12:52:43	×
Lap 4	2026-03-19 14:01:17	2026-03-19 14:07:34	×
Lap 5	2026-03-19 14:07:53	2026-03-19 14:10:29	×

RAMPS

Ramps are sub-segments within a lap. 'Lap' field is optional.

Name	In Lap	Start	End	
1 to 2	Lap 1	2026-03-19 12:39:40	2026-03-19 12:44:10	×
2 to 3	Lap 1	2026-03-19 12:44:11	2026-03-19 12:44:35	×
3 to 4	Lap 1	2026-03-19 12:44:36	2026-03-19 12:44:54	×

+ Add Ramp

Close Save Segments

Laps & Ramps

Define Segments Aggregate

Aggregate by: ☒ Lap ☐ Ramp ☐ Both ☐ Whole dataset

☐ Apply row filter — only include rows where:
battery_battery_rx_output_current < 0

Check the functions you want for each column ('Run Aggregation' also auto-saves segment definitions):

Column	mean	max	min	std	sum	count	first	last	range
battery_battery_rx_output_current	☐	☐	☐	☐	☐	☐	☐	☐	☐
battery_battery_rx_output_voltage	☐	☐	☐	☐	☐	☐	☐	☐	☐
battery_battery_rx_pack_soc	☐	☐	☐	☐	☐	☐	☐	☐	☐
Speed (km/h)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Angle_X_Dual_Axis	☐	☐	☐	☐	☐	☐	☐	☐	☐
Angle_Y	☐	☐	☐	☐	☐	☐	☐	☐	☐
RMS_ "X (40g)"	☐	☐	☐	☐	☐	☐	☐	☐	☐
RMS_ "Y (40g)"	☐	☐	☐	☐	☐	☐	☐	☐	☐

Max. missing data per segment: 100 — if more than this % of values are NaN, show NaN for that cell (100 = no fil

Run Aggregation

Close Save Segments

- ▶ Users can define laps and segments by importing an Excel file or defining them manually (PDF option still under development).
- ▶ During aggregation, several options are available to help preserve data integrity.



Natural Resources
Canada

Ressources naturelles
Canada

Data Processing Assistant – Segment and Aggregate

lap	segment	segment_start	segment_end	segment_duration _seconds	battery_battery_r x_output_current _mean	battery_battery_rx _pack_soc_first	battery_battery_rx_ pack_soc_last
Lap 1	A						
Lap 1	B						
Lap 2	A						
Lap 2	B						
Lap 3	A						
Lap 3	B						
Lap 4	A						
Lap 4	B						
Lap 5	A						
Lap 5	B						

- Once Run Aggregation is clicked, an Excel file containing the selected statistical features is automatically generated.

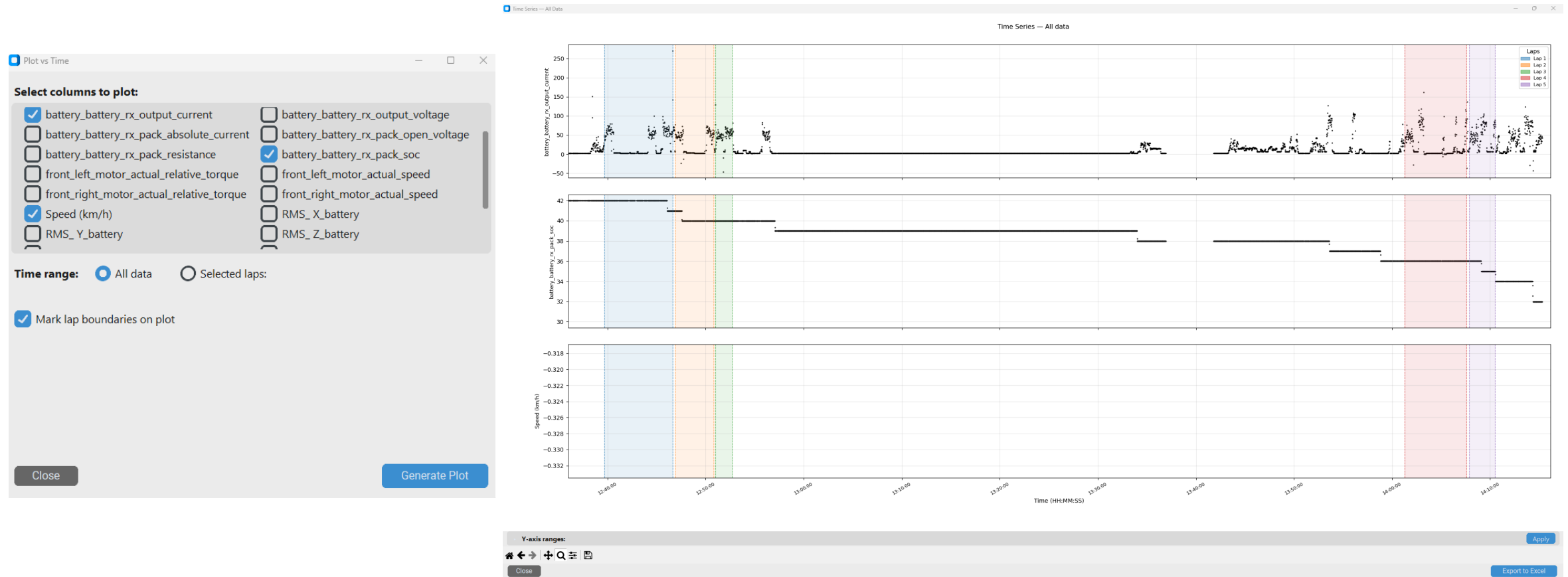
Note: values are blurred here due to sensitive nature of the data which was collected from equipment in an operating mine.



Natural Resources
Canada

Ressources naturelles
Canada

Data Processing Assistant - Plot



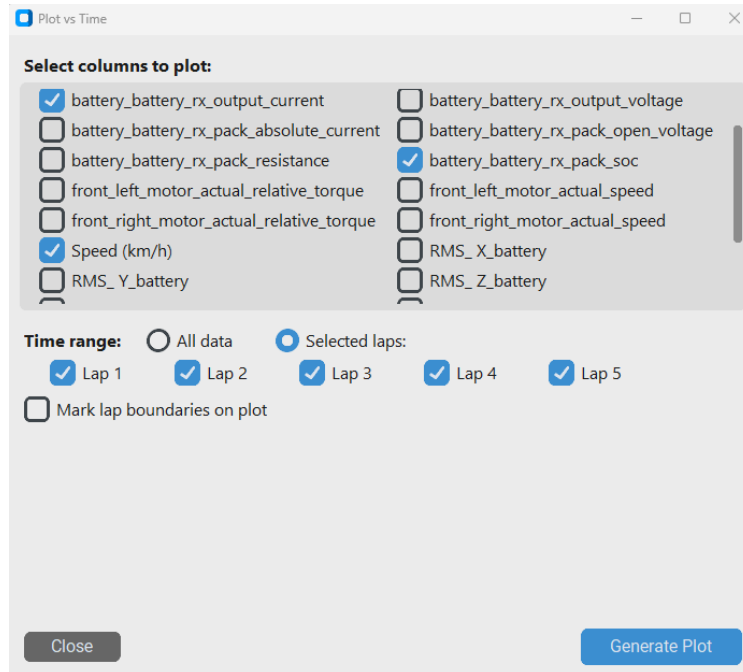
- Plots for the whole dataset can be generated and exported for selected columns, with lap boundaries clearly marked.



Natural Resources
Canada

Ressources naturelles
Canada

Data Processing Assistant - Plot



- Plots can also be generated and exported for each individual lap, with each segment automatically labeled.



Natural Resources
Canada

Ressources naturelles
Canada

From Data Processing to Research Publications

CanmetMINING battery electric vehicle field test series: Rokion R400 utility vehicle — surface:

<https://doi.org/10.1080/19236026.2022.2109924>

CanmetMINING diesel and BEV field test series: MacLean Engineering diesel and battery electric cassette truck:

<https://doi.org/10.1080/19236026.2023.2266351>

New Afton Mine diesel and battery electric load-haul-dump vehicle field test: Heat and dust contribution study:

<https://doi.org/10.1080/19236026.2024.2313952>

Energy consumption ratio and heat output comparison of battery electric vehicles and diesel-powered machines: Results of field trials in European Union and Canadian underground mines:

<https://doi.org/10.1080/19236026.2022.2109924>

CanmetMINING battery electric vehicle field test series: Relay utility vehicle:

<https://doi.org/10.1080/19236026.2023.2258716>



Natural Resources
Canada

Ressources naturelles
Canada

Conclusion

Why data processing matters

- ▶ Data processing is the bridge between data acquisition and modelling.
- ▶ Clean, well-structured datasets are what make modelling results accurate — garbage in, garbage out.
- ▶ By processing field-test data, we turn raw measurements into research findings that can be shared through technical reports and publications.
- ▶ As data-driven technologies and AI become more popular, good-quality data only becomes more valuable.

Processing by lap and segment

- ▶ Processing data by lap provides an overview of how each variable of interest changes across repeated cycles during the test.
- ▶ Processing data by segment maps the data to different machine actions and identifies the characteristics associated with each action.
- ▶ Lap- and segment-based processing also provide a way to validate data integrity by comparing the results with notes recorded during the tests.

Advantage of streamlining

- ▶ The streamlined/automated tool enables users without coding experience to process data.
- ▶ Results from different tests are more consistent in terms of processing steps, output formats, and graphing styles.
- ▶ It reduces data-processing time, allowing more time to be spent on tasks that require critical thinking.





Natural Resources
Canada

Ressources naturelles
Canada

Q & A

THANK YOU!

Contact: zhenhuan.xu@nrcan-rncan.gc.ca