Recovery of Hidden Information from Retrofit Data-Logging

Lukas Fabrykowski and Friedrich Legerer

Motivation

Data loggers usually take measurements every 10 seconds. This means that even in their simplest version – where they record only the timestamp, temperature, and pressure – they generate more than a thousand data points per hour. (See fig. 1) Despite the plethora of available information in most installations only a hard upper bound is checked on the vehicle itself, while a basic distribution analysis can be performed off-site. (See fig. 2 for an exemplary output of such an analysis.) Such an approach does indeed provide some insight into the prevalence and distribution of dangerously high or low readings. It does not, however, leverage the fact, that apart from the raw numbers there is more information hidden between the measurements.

Data quality

In order to draw implied conclusions from the readings, we have first to ensure the soundness of the underlying data. The first verification which is applied to the data and which can be executed on the fly are – similarly to the current use – lower and upper bounds for the values. A temperature measurement of the exhausts of a combustion engine below 50 degrees or above 1,000 degrees are physically infeasible and thus more than likely an indicator of a defective detector or an even more serious fault. The same goes for the pressure, where values above 600 mbar or long streaks of 0 mbar measurements should not be expected and therefore appropriately flagged. (The exact values should be tailored in cooperation with the manufacturers to the specific make of the car and filter as well as any other material factors. We have used reference values provided by the VERT-Association.)

After flagging the improbable values we proceed to divide the time series into operation cycles, which we define as consecutive measurements with a gap of no more than two minutes. This allows us to consider even such use cases like a public bus with its numerous stops from its departure until its return to the base as one operation cycle. Using those cycles we determine the ratio of suspect values to the regular ones. If it exceeds a certain threshold – we assumed 10% – the entire cycle is deemed untrustworthy and discarded; otherwise only the offending measurements are omitted.

Having removed observations on grounds of their extreme values we now turn to their variability. It is, after all, highly unlikely that either the temperature or pressure of the exhausts of a combustion engine will be (nearly) constant for any perceivable period. We do not, however, resort to the data's variance – primarily due to its lack of robustness, which means that a single outlier can arbitrarily influence the final value. Instead we turn to a central expression of information theory: entropy. Entropy, as defined by Claude Shannon, is quite similar to the analogous term in Thermodynamics, where it was introduced by Rudolf Clausius and later elaborated on by Ludwig Boltzmann: Considering a distribution $\mathbb{P}$ of measurements over the set of all possible measurements Z it is given by

$$H(\mathbb{P}) = -\sum_{x \in Z} \mathbb{P}(x) \ln \mathbb{P}(x)$$

Entropy by itself has the drawback that its value depends on the number of observations. In order to avoid this pitfall, we normalise with the maximal attainable entropy, which can be computed as $\ln \min(n, m)$, where n is the length of the observation and m the maximal possible number of distinct observations. The expression,

$$E(\mathbb{P}) = -\frac{\sum_{x \in Z} \mathbb{P}(x) \ln \mathbb{P}(x)}{\ln \min(|Z|, m)}$$

is the *bounded efficiency* of the distribution of measurements $\mathbb{P}$. For the purpose of this project we accept only operational cycles with efficiency greater than 0.3. (See fig. 3 for an exemplary evaluation of the efficiency.)

Data loggers record thousands of data points each day, yet they are usually only used to check against fixed thresholds – with the appropriate computations we can detect faults in the instruments of the logger and even attempt to forecast the level of ash deposits.

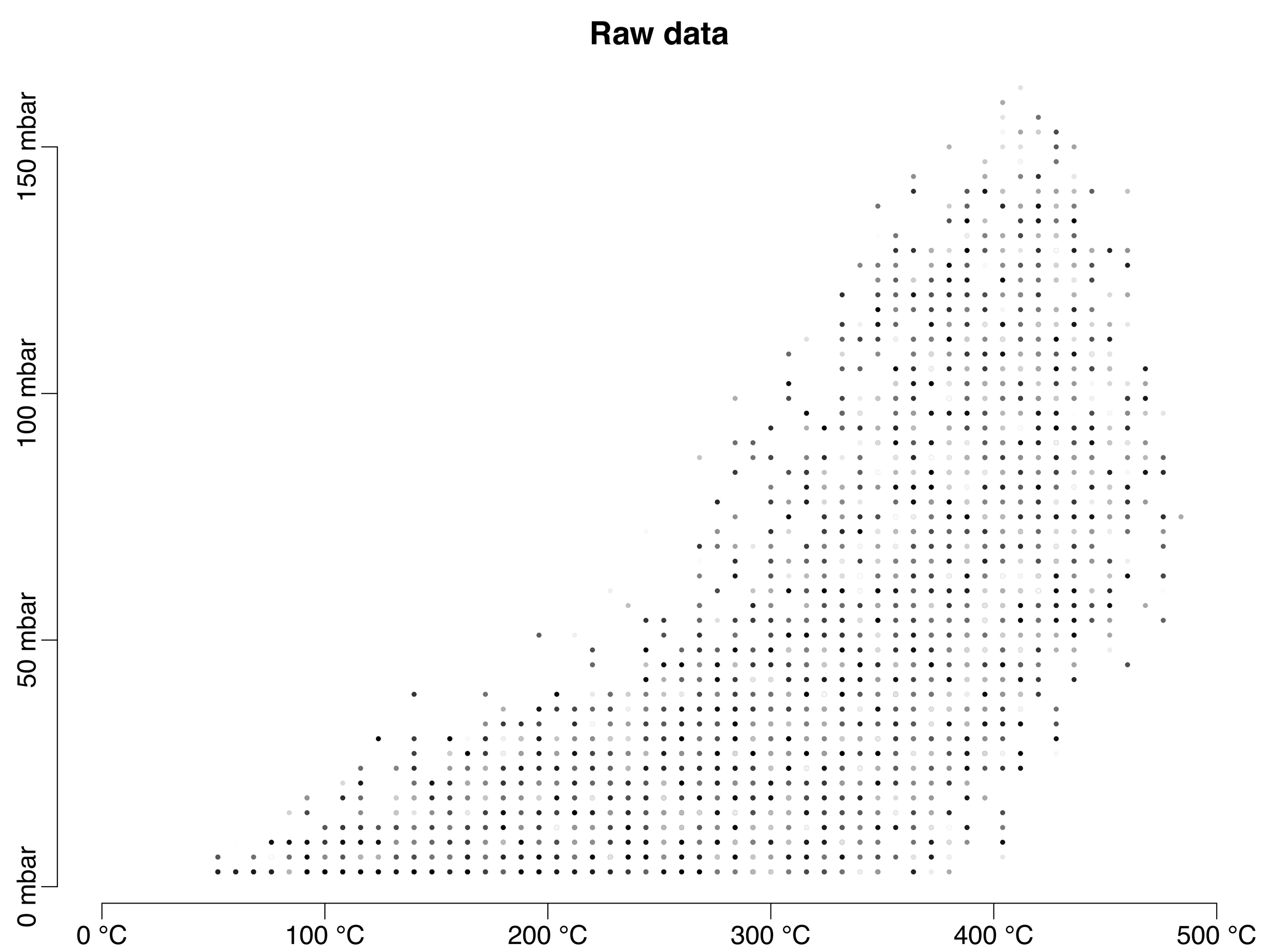

Fig. 1 – Raw data from a logger. Lighter points are older

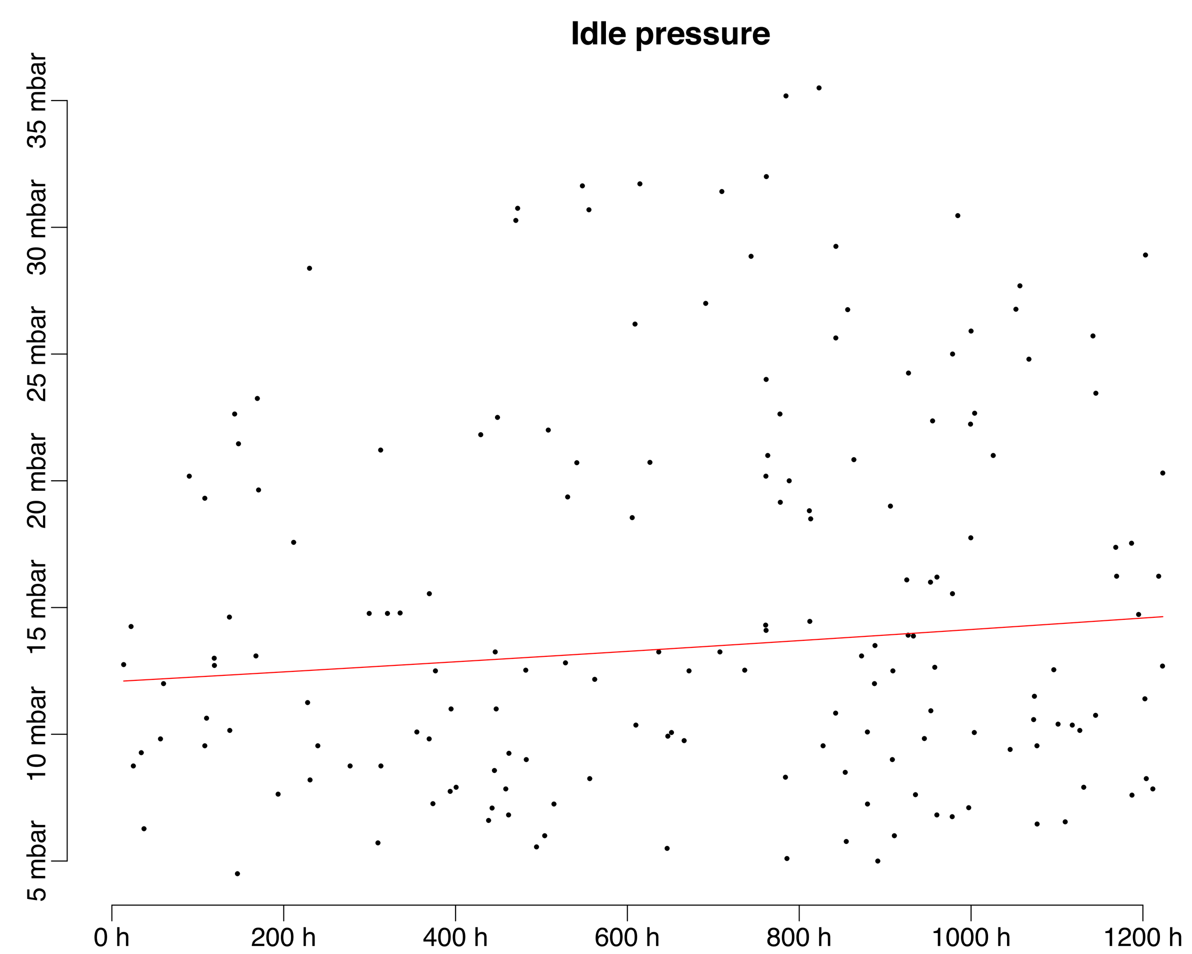

Fig. 5 – Detected idle pressures and trend line

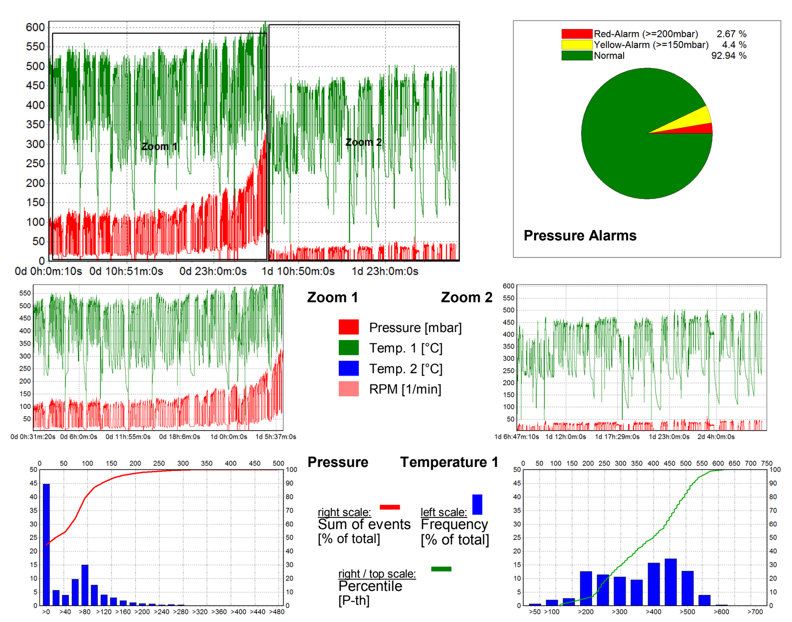

Fig. 2 – Output from a current analysis-tool

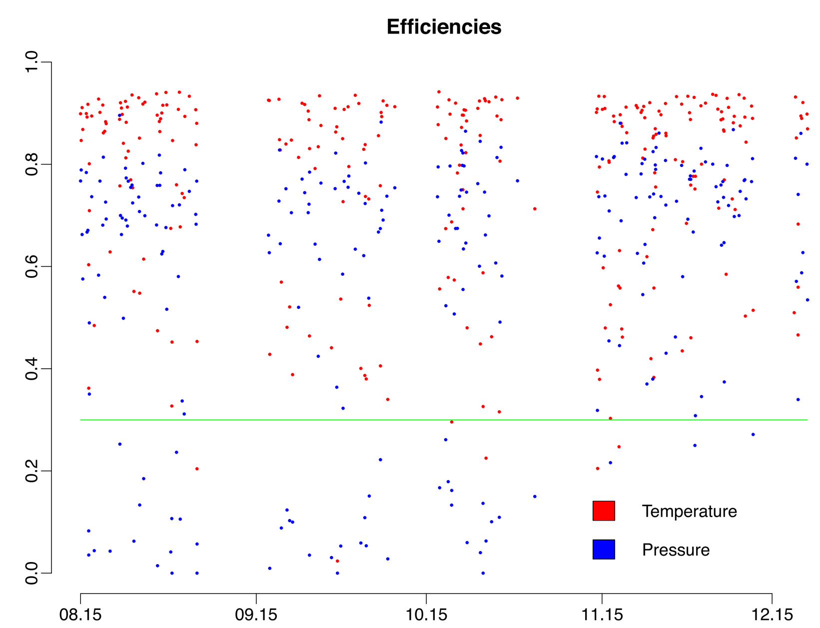

Fig. 3 – Efficiencies of the measurements

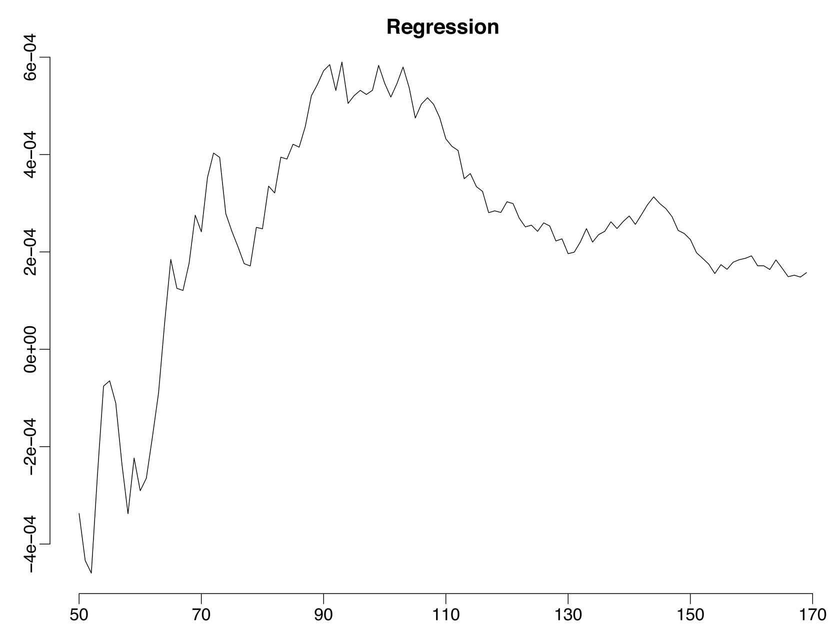

Fig. 4 – The slope of the regression vs. number of observations

Determining ash levels

With only those measurements left, which are considered plausible, we can try to infer the ash levels of a filter. Obviously it will be impossible to determine the actual levels having only measurements of temperature and pressure at our disposal, but we have a good proxy: when the ash levels increase in a DPF, so does the pressure during idle operation. Therefore we have to extract the times, when the machine was idling and examine the pressure during those periods. To do so we consider the temperature. During idling the engine does not perform any work and can cool down, which also results in cooler exhausts – this can be observed as an exponential drop in the readings. This suggests the following procedure to determine the back-pressure during idling:

- Find all decreasing segments in the temperature readings
- Try to fit an exponential decline
- If there is no fit, discard the segment, otherwise compute the average pressure in this time frame

Fig. 5 shows such idle back-pressures in relation to operating hours.

Finally a trend line can be fitted to the idle pressures. Due to the nature of the ash sediments an exponential regression should be used. (Fig. 4 shows how such a regression stabilises with increasing number of observations.) With a sufficient number of empirical values a threshold for the projected idle pressure might be given, which could correspond to a filter that has to be cleaned from ash.

Acknowledgements

The authors are indebted to Dr.h.c. Andreas Mayer for encouragement, mutual discussions in the course of work as well as providing data together with Mr. Paul Nöthiger – inventor of the data logger.

This project has been financially supported by the VERT-Association as part of the "Datalogger Statistics" programme, which we gratefully acknowledge.

Contact

Lukas Fabrykowski, lukas@fabrykowski.com
Friedrich Legerer, VERT-Association, ztdr.legerer@aon.at