Prognosis based on Varying Data Quality

Authors: Mauthe, Fabian; Bakir, El Melik; Scheerer, Tim; Zeiler, Peter
Institution: Hochschule Esslingen - University of Applied Sciences
Research Department: Reliability Engineering and Prognostics and Health Management
Contact information: ReliabilityAndPHM@hs-esslingen.de

Date: 16. June 2022

Description of the test bench:
This data set originates from a degradation process, which is representative for Prognostics and Health Management (PHM) applications. It is one of multiple data sets which are generated for various practice-relevant data situations which do not correspond to the ideal conditions of full data coverage. These data sets are uploaded to Kaggle by the user "Prognostics @ HSE" in a continuous process. The observed degradation process is the clogging of filters when separating of solid particles from gas. A test bench is used for this purpose, which performs automated life testing of filter media by loading them. A schematic of the test stand's basic operating principle can be seen in Figure 1. Due to various physical filter models described in the literature, it might be beneficial to support the actual data-driven models by integrating physical knowledge respectively models in the sense of theory-guided data science or informed machine learning (various names are common).

Filter medium:
The object to be tested is a filter mat made of randomly oriented non-woven fiber material. These filter mats are cut out of a flat material, which is fixed in the middle of the filtration chamber. The test material has an effective square flow area of $6131\ \mathrm{mm}^2$. Figure 3 shows a cut filter mat after a life cycle test, which is loaded with the test dust. The properties of the filter media are:

Mean fiber diameter	23 µm
Filter area	$6131\ \mathrm{mm}^2$
Filter thickness	20 mm
Filter packing density	0.014-0.0165
Clean filter pressure drop	25 Pa at flow of $1080\ \mathrm{m}^3/(\mathrm{h}\cdot\mathrm{m}^2)$

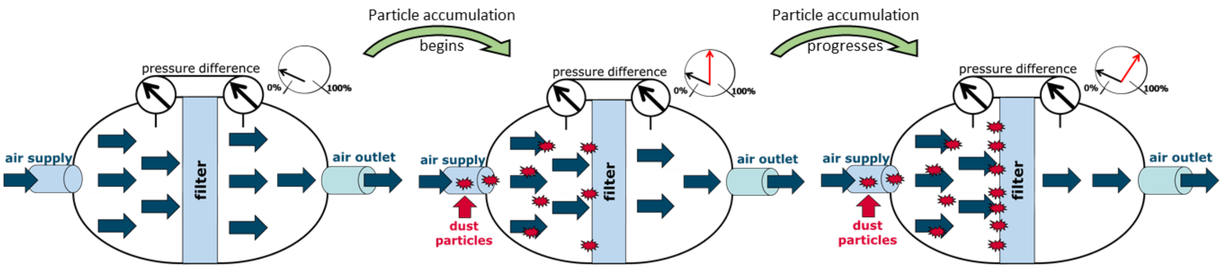

Figure 1 Basic operating principle of the test bench

Test particles and particle feeding:
For testing the filters life, standardized Arizona test dust (ISO 12103-1) in the particle sizes A3, is used for their loading. This dust has a defined distribution function with regard to particle size and chemical composition. The distribution of the particle sizes of the test dust batch used, is shown in Figure below and given in the file:

- Particel_size_distribution_ISO_12103_1_A3_Medium_PTI_ID_Data_Quality.mat

The dust particles are induced into the compressed air system of the test bench by a nozzle for powdery solids based on the Venturi principle. The test dust is contained in a cylinder. The dust is fed to the nozzle by an electrically driven spindle actuator as shown in Figure 3. Thus, the load profile can be adjusted by the velocity of the spindle actuator. By adjusting the velocity, the amount of particles added per time unit is set, independently of differential pressure and flow. In the data set, the current material feed is given in mm^3/s. Together with the bulk density of the dusts of

A3 Medium Test Dust $\quad 1.025\mathrm{g/cm}^3$,

the mass of the fed dust can be calculated.

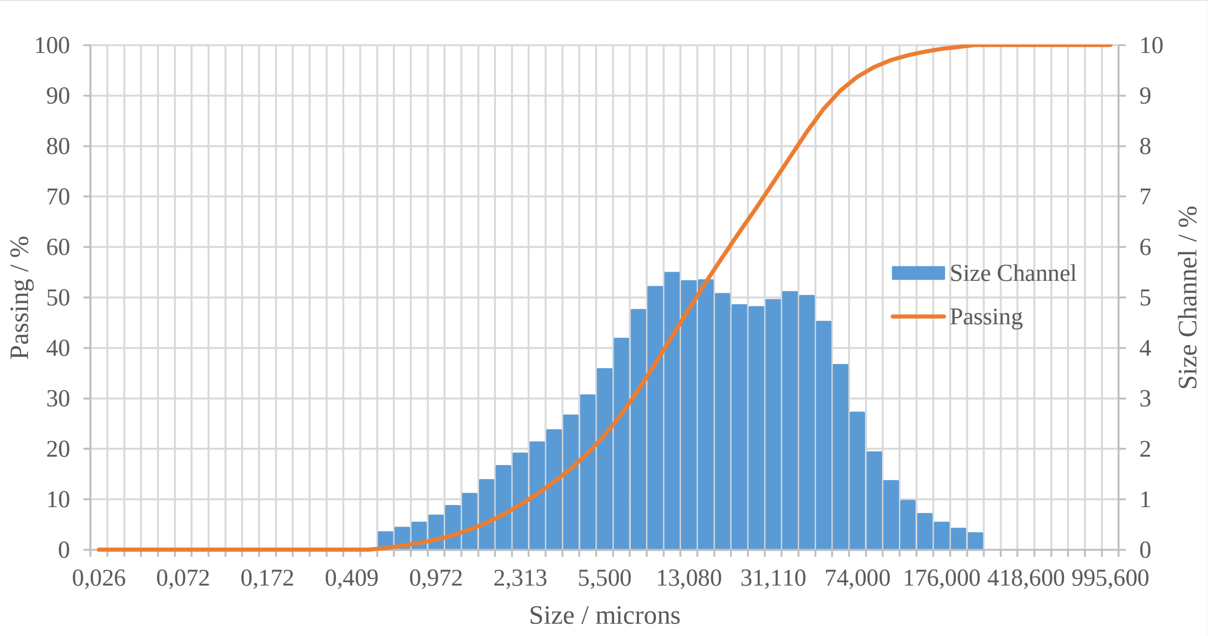

Figure 2: Particle Size Distribution: ISO 12103-1, A3 Medium

Sensors:
In this test bench, the flow rate, particle feed and the differential pressure across the filter are recorded during a lifetime test. Here, the flow sensor is located upstream of the particle feed, which leads to a falsification of the flow. The impact of this falsification is determined experimentally and a compensation function is modelled to correct it. The given measured values in the data are the corrected values. Thereby, the flow sensor indicates the flow rate for atmospheric pressure. The differential pressure sensor has a measuring range of 0 to 2500 Pa.
Since air flows along the sensor connections in the filter chamber, the sensor connections were designed to minimize interference. A typical course of a measurement of the differential pressure from the filter application can be seen in Figure 4.

Operating modes:
In this test bench, during the life test, there is the possibility of selecting between two basic operating modes via the 5/3-way valve and the 3/2-way valve, which can be seen in Figure 5. Thus, during a test, it is possible to switch between flow and pressure control as needed. The respective operating mode is documented in the data of each life test. In the case of the "Prognosis based on Varying Data Quality" data set, all life tests are carried out under flow control.

Procedure for a measurement:
In order to reduce the external influences, each measurement has the same start position of the spindle and the same stop criterion. Therefore, the spindle with the cylinder containing the test dust is located at the lower limit switch. Also, the cylinder is filled with the same amount of dust. Before starting the measurement, the setting parameters such as feed rate profile of the spindle and the flow rate are defined in the user interface.

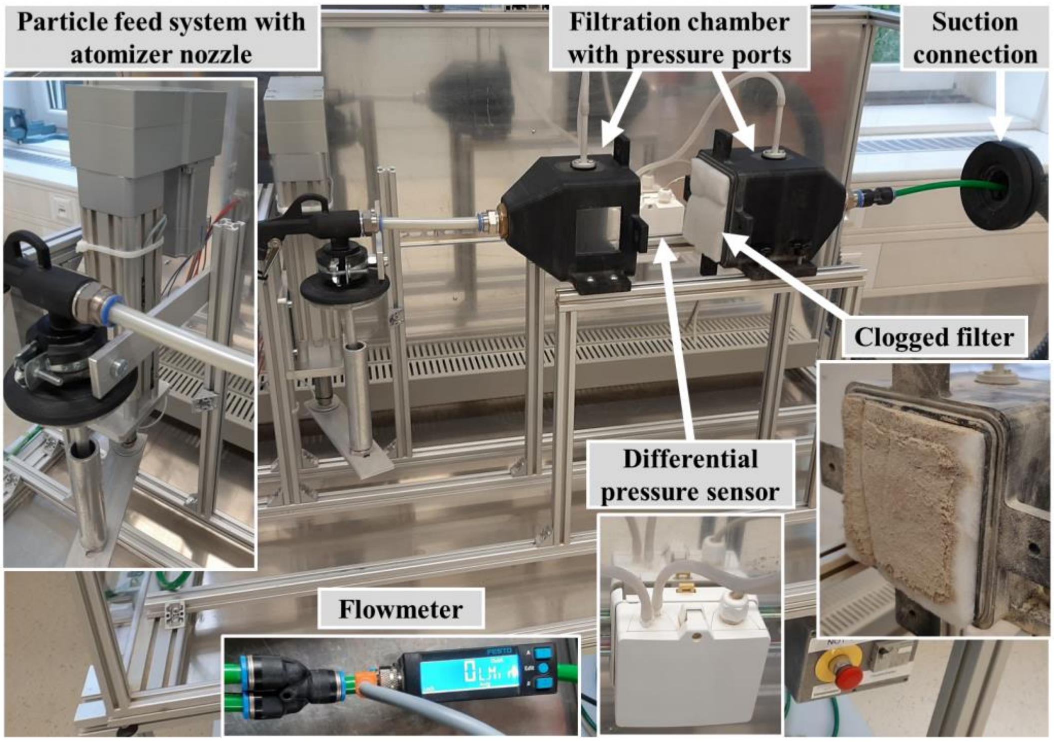

Figure 3: Filtration test bench with its particle feed system and filtration chamber

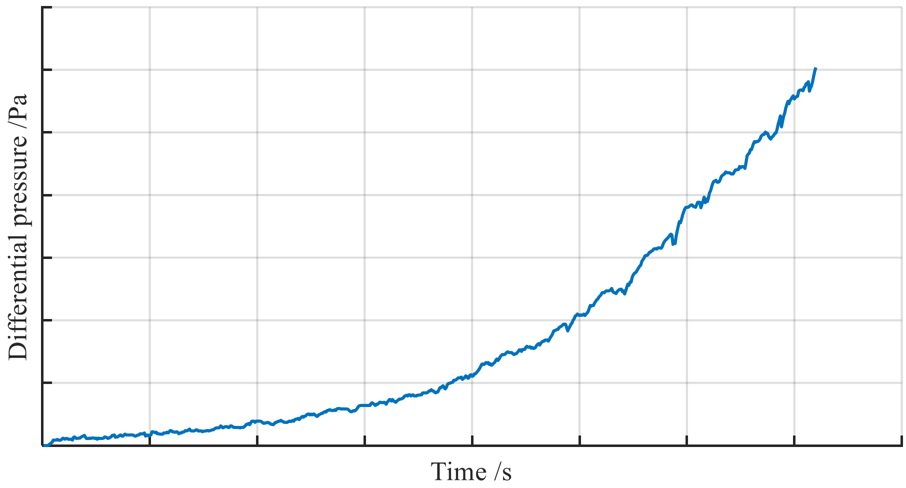

Figure 4: Typical course of a filtration test run

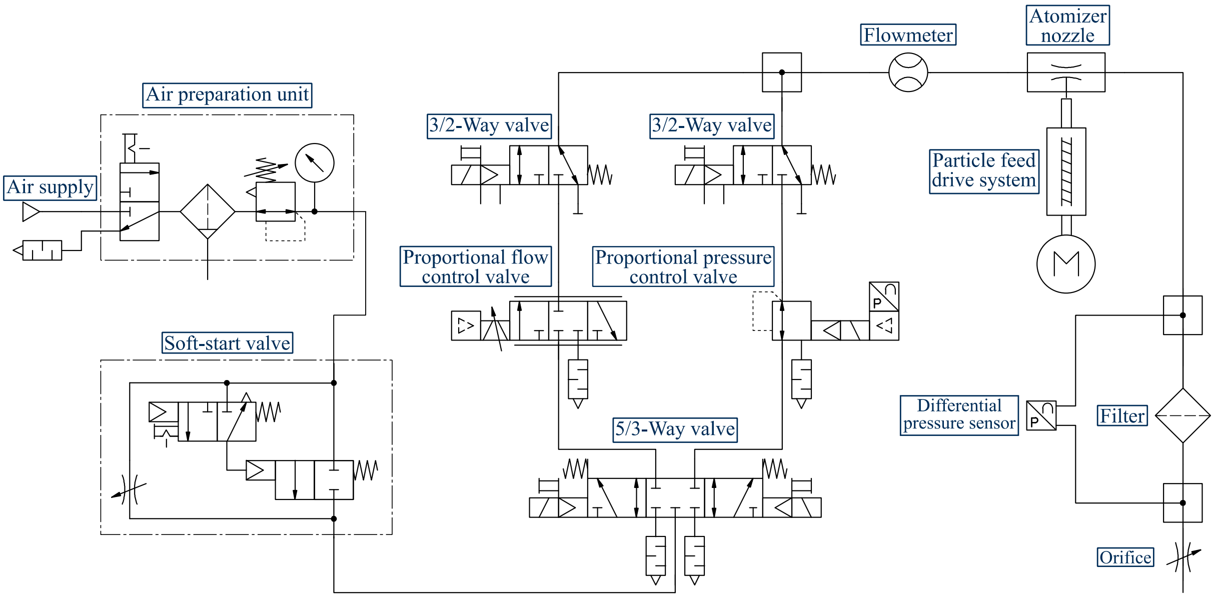

Figure 5: Pneumatic layout of the test bench

Given data situation:

Data quality is significantly affected by the sensors, sensor location and peripherals used. This yields a different bias and different noise of the data. If technical systems are in the market with different sensors and their positioning, for example at the costumer's request, this results in varying data qualities. This typically occurs when cost-benefit analyses are carried out in the course of product development to identify the most suitable components for different customers from a techno-economic perspective. Developing a diagnostic or prognostic application those variations have to be taken into account. Therefore, to obtain accurate results, models employed must recognize different measurement qualities in order to properly address them.

All data in this data set were measured on the same filter test bench with the same sensors. However, to represent the described data situation, the measurement data were subsequently manipulated. Each run-to-failure trajectory included is subject to a signal-to-noise ratio (SNR). The SNR refers to a ratio between the existing measured value and an interfering signal. Further, the SNR varies between the various trajectories. In addition to noise, usage of different sensors or their positioning can cause an offset in the measured data. For this reason, a certain bias, constant for one trajectory, was applied to each run-to-failure trajectory. In order to represent multiple configurations, there are four levels of bias in the entire data set (indicated by a-d), whose magnitude is not stated and which were randomly assigned to trajectories.

In the present data set, the life cycle of filters or rather their condition data, represented by the differential pressure, is considered. Failure of the filter occurs when the differential pressure across the filter exceeds 600 Pa. The time until a filter failure occurs depends especially on the amount of dust supplied per time, which varies within a run-to-failure cycle. The previously explained data characteristics are addressed by means of corresponding training and test data.

Training data contains for each data point the differential pressure, flow rate, dust feed and the current time as well as remaining useful life (RUL). The differential pressure measurements are obscured by additional noise in accordance with the random SNR and shifted by one of the four bias values. Test data are right-censored and also contains differential pressure, flow rate, dust feed, and current time. The differential pressure measurements are likewise subject to noise and bias. The RUL-information is given for the last data points of the censored trajectories in the test data.

Task:

The aim is to predict the RUL of the censored filter test cycles given in the test data. In order to predict the RUL, training and test data are given, each consisting of 35 and 20 run-to-failure cycles. The test data contains randomly right-censored run-to-failure trajectories and the respective RUL for the prediction task. The main challenge is to develop a RUL prediction as accurately as possible despite varying data qualities of the trajectories.

Due to the detailed description of the setup and the various physical filter models described in literature, it is possible to support the actual data-driven models by integrating physical knowledge respectively models in the sense of theory-guided data science or informed machine learning (various names are common).

Data structure:
There are two data sets, the test data set and the training data set. These contain the individual measurement series of the life cycle tests. The individual life cycle tests have the same structural design. Therefore, they contain all necessary measurement data and setting parameters. The structure of a life cycle test is shown below:

– Train_Data	(table)
– Data_No	(double)
– Measured_Data	(table)
– Differential_pressure	(double)
– Flow_rate	(double)
– Time	(double)
– Dust_feed	(double)
– RUL	(double)
– Sampling	(double)
– Bias_type	(char)
– Test_Data	(table)
– Data_No	(double)
– Measured_Data	(table)
– Differential_pressure	(double)
– Flow_rate	(double)
– Time	(double)
– Dust_feed	(double)
– Sampling	(double)
– RUL	(double)
– Bias_type	(char)

The Data.mat file contains all the training and test data, with the previously mentioned structure. The individual measurement series are contained in the training and test data, with the recorded data and the necessary setting parameters on the PLC. “Data_No” indicates to what run-to-failure cycle the data belongs. In the table “Measured_Data”, the measured differential pressure, flow rate, time, and dust feed are listed at each point in time. “Dust_feed” corresponds approximately to the loading of the filter with dust per time (mm^3/s). In the training data, the RUL is also specified at each point in time. For the right-censored test data, the RUL is specified at the last point in time. “Sampling” indicates the frequency at which the measurement data is recorded (Hz). In addition, the known bias type is specified for all run-to-failure trajectories.

Acknowledgement:
Thanks go to Simon Hagmeyer (Research Assistant) and Marcel Braig (Research Assistant) for their contribution to the creation of the data set.

Data set creator:
Hochschule Esslingen - University of Applied Sciences
Robert-Bosch-Straße 1
73037 Göppingen
Germany
Research Area Reliability Engineering and Prognostics and Health Management

Data set citation:
Hagmeyer, S., Mauthe, F., & Zeiler, P. (2021). Creation of Publicly Available Data Sets for Prognostics and Diagnostics Addressing Data Scenarios Relevant to Industrial Applications. *International Journal of Prognostics and Health Management*, Volume 12, Issue 2, DOI: 10.36001/ijphm.2021.v12i2.3087