# A Method to Estimate the Remaining Useful Life of a Filter Using a Hybrid Approach Based on Kernel Regression and Simple Statistics

Roman Łomowski[1], Szczepan Hummel[2]

[1,2]*Syncron Poland Sp. z o.o., Warsaw, 00-105, Poland*
*roman.lomowski@syncron.com*
*szczepan.hummel@syncron.com*

## ABSTRACT

This paper describes the method used by the Uptime team for the estimation of the remaining useful life of a filter during the 2020 PHM Conference Data Challenge. The proposed method is a hybrid of two methods: (1) based on median lifetime of a filter for a particular contamination profile and (2) kernel regression for a sensor-based prediction after a certain threshold of differential pressure is reached. The threshold value was chosen based on visual assessment followed by grid search for fine tuning. Median lifetime of a filter for unseen contamination profiles was estimated using interpolation. Choosing the right interpolation method was a challenge as training data contained samples with only two values of contamination particle size. Interpolation was chosen based on other publicly available information about the relationship between contamination profile and filter lifetime. The results (ranked 1st with the total penalty score of 49.67) showed that an observation made based on one dataset can be useful for solving similar problems in the case of limited data availability. This suggests that there is a potential for using transfer learning in PHM applications.

## 1. PROBLEM DESCRIPTION

Contaminant filtration is a process needed in multiple applications as contaminants in liquids can be devastating to many types of equipment. One of the common reasons why filtration systems require maintenance is filter clogging. This year's PHM Data Challenge problem was to create a prognostic model that estimates a filter's Remaining Useful Life (RUL) defined as the time until upstream pressure (before filter) is greater than downstream pressure (after filter) by at least 20 psi. The model could use information about: the size of contaminant particles, their concentration in the liquid (specifically water) and sensor readings. Data consist of run-to-failure histories gathered under controlled conditions where contaminant concentration and particle size are fixed and known for each sample.



### 1.1. The Experimental Rig

Run-to-failure data were gathered using an experimental rig constructed as proposed by Skaf, Eker, and Jennions (2017) and described on the official competition web page (PHM Data Challenge, 2020). The circuit is presented on Figure 1.

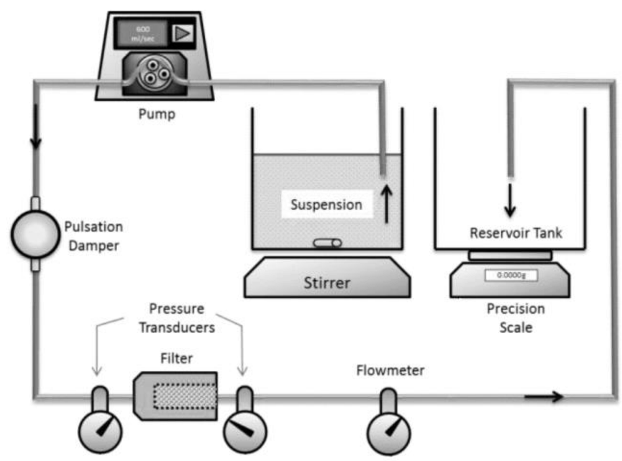


Figure 1. The experimental rig

The main components of the circuit are:

- source suspension tank – contains prepared suspension,
- pump – enforces flow through the filter,
- pulse damper – reduces pulsation in the flow,
- upstream pressure sensor,
- filter,
- downstream pressure sensor,
- flow rate sensor,
- reservoir for filtrated liquid.

### 1.2. Data

#### 1.2.1. Training Data

Training data available to participants contain 32 samples. Each sample is a data set coming from an experiment run with a particular contamination profile. A contamination profile is described with three parameters:

- solid ratio (%) of a contaminant within fluid,
- min particle size (μm),
- max particle size (μm).

Experimental data contain sensor readings (upstream pressure, downstream pressure, and flow rate) gathered at frequency of 10 Hz.

The samples in the training data are distributed among contamination profiles as shown in Table 1.

Table 1. Distribution of training samples.

| Particle size (μm) | Solid ratio (%) | Number of samples |
|---|---|---|
| 45-53 | 0.4 | 4 |
| small | 0.425 | 4 |
| | 0.45 | 4 |
| | 0.475 | 4 |
| 63-75 | 0.4 | 4 |
| large | 0.425 | 4 |
| | 0.45 | 4 |
| | 0.475 | 4 |

#### 1.2.2. Test Data

Test data hidden to participants contain 16 samples distributed among contamination profiles as shown in Table 2.

Table 2. Distribution of test samples.

| Particle size (μm) | Solid ratio (%) | Number of samples |
|---|---|---|
| 53-63 | 0.4 | 4 |
| medium | 0.425 | 4 |
| | 0.45 | 4 |
| | 0.475 | 4 |

### 1.3. Performance Evaluation

Model is evaluated based on predictions generated every 10 seconds after the experiment begins. The Mean Absolute Error (MAE) is used for the error function according to the competition rules. To be more specific: let $PRUL_m(s,t)$ be the predicted RUL of the filter for a sample $s$ at time $t$ by a model $m$, let $RUL(s,t)$ be the real (ground truth) remaining useful life of the filter for a sample $s$ at time $t$, let $T_s = \{10n : n \in N,\ 10n < RUL(s,0)\}$ be the set of time instants for a sample $s$ on which the model is evaluated. Then the error function $E$ on a set $S$ of samples is defined as:

$$E(\mathrm{m},\mathrm{S}) = \frac{1}{|\mathrm{S}|}\sum_{\mathrm{s}\in\mathrm{S}}\sum_{\mathrm{t}\in\mathrm{T_s}}\frac{|PRUL_m(\mathrm{s},\mathrm{t}) - RUL(\mathrm{s},\mathrm{t})|}{|\mathrm{T_s}|} \qquad (1)$$

Let *Train* be the set of all samples from training data. Let *Test* be the set of all samples from test data. The penalty score $P$ of the model m is defined as:

$$P(m) = E(m, Train) + \frac{3}{2}E(m, Test) \qquad (2)$$

Participants were supposed to provide 4 models:

- $m_{100}$ – using all the available samples to train the model,
- $m_{75}$ – using 75% of the available samples to train the model,
- $m_{50}$ – using 50% of the available samples to train the model,
- $m_{25}$ – using 25% of the available samples to train the model.

Solutions are being scored based on *Total Penalty Score* calculated as:

$$\begin{aligned} &Total\ Penalty\ Score = \\ &= P(m_{100}) + P(m_{75}) + P(m_{50}) + P(m_{25}) \end{aligned} \qquad (3)$$

### 2. THE SOLUTION

Subsection 2.1 describes our model. Section 2.2 describes the data analysis that gave us the intuition needed to construct the model. Section 2.3 describes how we choose subsets of training data for training of models $m_{75}$, $m_{50}$ and $m_{25}$. Section 2.4.1 describes analysis that led us to the conclusion that obtaining and using any additional information would probably have greater impact on *Total Penalty Score* than further experimentation with better models. Section 2.4.2 describes what we tried to do in order to obtain additional information and what we managed to find. Section 2.5 presents how the findings were used to come up with the appropriate weights used by the model.

### 2.1. The Model

For a sample $s$, we define $FlowStart(s)$ as the first time instant when flow rate was at least 100 ml/min for this sample.

Let useful life of a sample $s$ be defined as $UL(s) = RUL(s, 0)$. Let useful life since flow start for a sample $s$ be defined as

$$\widehat{UL}(s) = UL(s) - FlowStart(s)$$

Let $Differential\,Pressure_t(s)$ be a difference between upstream pressure and downstream pressure for a sample $s$ and a time instant $t$.

Let $SR(s)$ be the solid ratio of a sample $s$. Let $TrainSR = \{0.4\%, 0.425\%, 0.45\%, 0.475\%\}$ be the set of all solid ratios occurring in the training data set.

Let middle particle size $MPS(s)$ be the arithmetic average of maximum and minimum particle size of a sample $s$. Let $TrainMPS = \{49, 69\}$ be the set of all middle particle sizes occurring in training data set.

For $r \in TrainSR$ and $p \in TrainMPS$, let:

$$UL_{r,p} = median\left(\left\{UL(s)\colon \begin{matrix} s \in Train, \\ SR(s) = r, \\ MPS(s) = p \end{matrix}\right\}\right),$$

$$\widehat{UL}_{r,p} = median\left(\left\{\widehat{UL}(s)\colon \begin{matrix} s \in Train, \\ SR(s) = r, \\ MPS(s) = p \end{matrix}\right\}\right).$$

Let $wavg(\vec{X}, \vec{w})$ be the weighted average of elements of a vector $\vec{X}$ with weights $\vec{w}$. If the order is clear, we will use sets instead of vectors in this notation.

By $\mathrm{MA_h(x_t)}$ we denote the moving average of a time series $x$ with the sliding window of size $h$.

Window size $h$ and threshold $\theta$ are hyper-parameters optimized through experimentation. Calculation of weights $\vec{\mathrm{w}}$ is described in Section 2.5. Vector $\vec{\mathrm{w}}$ contains 2 weights: one for *MPS* of 49 and one for *MPS* of 69.

For a time instant $t$ and a sample $s$ the model provides prediction $PRUL_m(s, t)$ as follows.

1. If $t < FlowStart(s)$ and $MPS(s) \in TrainMPS$:
   $$PRUL_m(s, t) = UL_{SR(s),MPS(s)} - t$$
2. If $t < FlowStart(s)$ and $MPS(s) \notin TrainMPS$:
   $$PRUL_m(s, t) = wavg(\{UL_{SR(s),p}\colon p \in Train\mathrm{MPS}\}, \vec{w}) - t$$
3. If $t \geq FlowStart(s)$, $MPS(s) \in TrainMPS$ and $MA_h(Differential\,Pressure_t(s)) \leq \theta$:
   $$PRUL_m(s, t) = \widehat{UL}_{SR(s),MPS(s)} - \hat{t},$$
   where $\hat{t} = t - FlowStart(s)$.
4. If $t \geq FlowStart(s)$, $MPS(s) \notin TrainMPS$ and $MA_h(Differential\,Pressure_t(s)) \leq \theta$:
   $$PRUL_m(s, t) = \mathrm{wavg}(\{\widehat{\mathrm{UL}}_{\mathrm{SR(s),p}}\colon \mathrm{p} \in Train\mathrm{MPS}\}, \vec{\mathrm{w}}) - \hat{\mathrm{t}},$$
   where $\hat{t} = t - FlowStart(s)$
5. If $MA_h(Differential\,Pressure_t(s)) > \theta$ the model uses kernel regression with 3d Gaussian kernel based on features: $MA_h\big(Differential\,Pressure_t(s)\big)$, $SR(s)$ and $MPS(s)$ to estimate *RUL*.

Kernel Regression bandwidth parameters for the 3 input variables are optimized through cross-validation on training data using *statsmodels* Python package. The optimization is done independently for models $m_{100}$, $m_{75}$, $m_{50}$ and $m_{25}$ on the respective data sets, as specified in Section 1.3.

Hyper-parameters were optimized globally through multiple experiments and the result of the optimization follows:

- $h$ = 20
- $\theta$ = 3.5

### 2.2. Analysis

The purpose of this section is to show what sequence of observations led to the creation of the model described in Section 2.1.

Figure 2 shows how differential pressure evolves in time for different experiments. Recall that *RUL* is calculated as number of seconds until differential pressure of 20 psi is first reached.

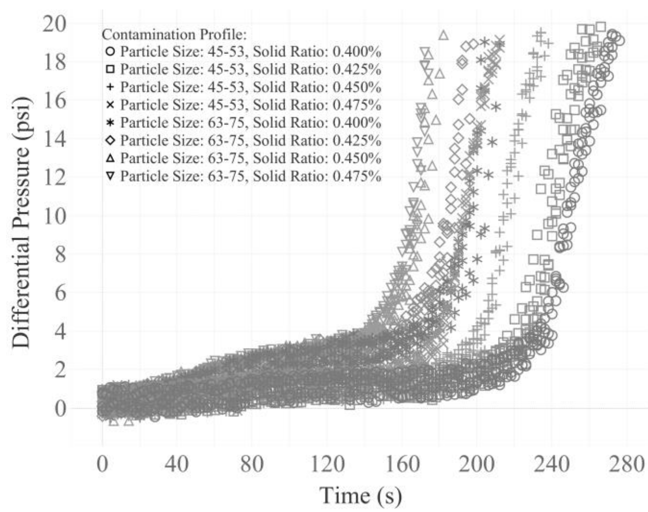


Figure 2. Differential pressure evolution per profile.

**Observation 1.** For a given contamination profile the difference in useful life in between samples is not that big, see Table 3.

Table 3. Variation of Useful Life (UL) per profile.

| Solid Ratio | Middle Particle Size | Shortest *UL* | Longest *UL* |
|---|---|---|---|
| 0.4 | 49 | 273.1 | 277.5 |
| | 69 | 203.5 | 213.4 |
| 0.425 | 49 | 256.1 | 266.2 |
| | 69 | 192.8 | 198.8 |
| 0.45 | 49 | 234.2 | 238.1 |
| | 69 | 175.2 | 182.1 |
| 0.475 | 49 | 210.8 | 212.5 |
| | 69 | 172.6 | 176 |

Let us now see what additional information about *RUL* we can get from sensor readings. Figure 3 shows *RUL* – the target variable – versus downstream pressure and flow rate.

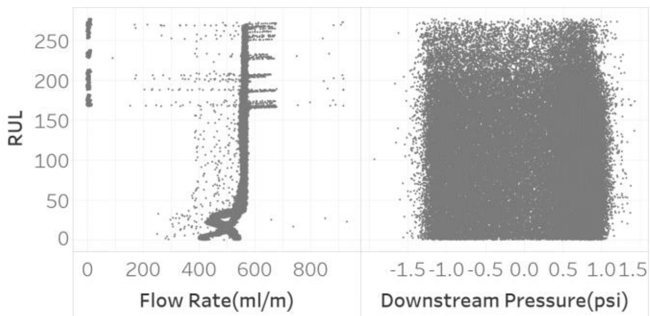


Figure 3. RUL vs flow rate and downstream pressure.

**Observation 2.** There is almost no functional dependency of *RUL* from downstream pressure or flow rate.

Downstream pressure looks like a noise when drawn versus *RUL*. Flow rate is almost constant, with very wide range of *RUL* values. This is probably because the pump has enough power to maintain constant flow. Near clogging some flow drop patterns occur, but it might not be easy to understand at which stage of the pattern we actually are, given that the patterns are significantly shifted in between particle sizes, and the test data are going to contain an unknown particle size.

Moreover, the drop in the flow rate can be caused by increasing pressure, as the pulse dampener contains gas. When liquid pressure starts to rise quickly, the gas will give way to liquid and therefore some part of the flow through the filter will be reduced. In the context of this note, flow rate may be a measure derived from pressure measures, which makes it even less useful for a model.

Let us now look at upstream pressure and differential pressure and how they relate to *RUL*. Figure 4 shows only subset of data, otherwise it would be unreadable.

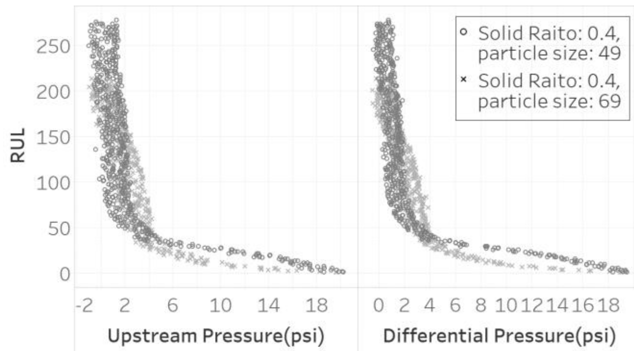


Figure 4. RUL vs upstream and differential pressure shown for subset of data for two profiles.

**Observation 3.** Differential pressure seems to bring very similar information as upstream pressure but is less noisy.

In the model, we use a smoothed version of differential pressure – namely its moving average – to reduce noise even further.

**Observation 4.** For a given contamination profile, if differential pressure is above 4 psi it allows for accurate prediction of RUL. The relation is not linear – at least for some contamination profiles.

Figure 5 shows the flow rate near the beginning of each of the experiments.

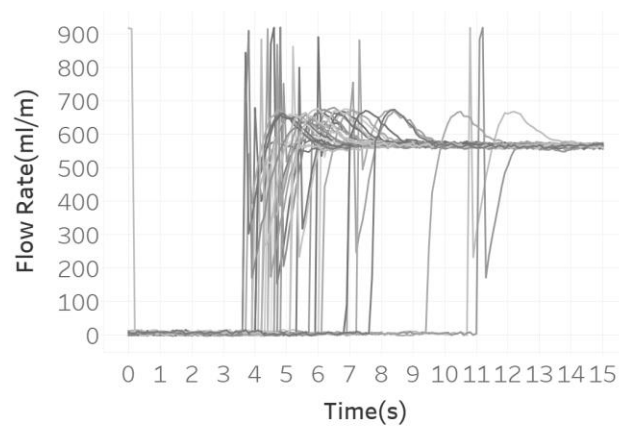


Figure 5. Different flow start times for different samples.

**Observation 5.** In majority of the experiments (samples), the liquid did not flow through the filter from the very beginning (see Figure 5). The delay of flow start varies from 3.5 s to even 11 s.

This is probably because of the different levels of liquid in the test rig at the start of each experiment. To mitigate the effect of delayed flow start, we have introduced useful life since flow start apart from simple useful life into the model.

### 2.3. Data Set Split

We have split the training data into 4 sets of samples. Each set contained exactly one sample with a given contamination profile. For example, among samples with particle size of 45-53 µm and solid ratio of 0.4%, the sample with the lowest number (Sample 1) is taken to Set 1, the next one (Sample 2) to Set 2, etc.

Model $m_{25}$ uses Set 1 during training; model $m_{50}$ uses Set 1 and Set 2, etc.

### 2.4. Choosing Next Steps

#### 2.4.1. Pseudo Profiling

A popular technic for software developers when they need to improve performance of an application is to perform profiling - measure how much each part of the application contributes to the overall execution time - and then focus on improving the most significant parts. Therefore, we decided to perform analysis of similar purpose to identify which actions would be most beneficial for further improvements of models' *Total Penalty Score*.

We performed this analysis after constructing a solution with the average penalty score on training data set for models $m_{25}$ , $m_{50}$ , $m_{75}$ and $m_{100}$ of about 1.9 s so improving it further could not bring more than 1.9 s of improvement.

We tried to estimate how big of an impact the correct choice of interpolation method for unknown particle size would have on the average penalty score of our models. We were considering interpolations based on assumption that useful life is linearly dependent on Middle Particle Size (MPS), $MPS^2$, $MPS^3$, $MPS^{-1}$, $MPS^{-2}$ or $MPS^{-3}$. For most of the cases we found some argument supporting it. We also checked what would be predicted useful life for the samples with unknown MPS if we assume each of those possibilities. Table 4 summarizes those values.

Table 4. Interpolated useful life for MPS=58 depending on different interpolation assumption for different solid ratios.

| | Solid ratio | | | | Average |
|---|---|---|---|---|---|
| | 0.475 | 0.45 | 0.425 | 0.4 | |
| $MPS^3$ | 197.7 | 214.3 | 237.5 | 249.6 | 224.8 |
| $MPS^2$ | 196.2 | 212.0 | 234.7 | 246.8 | 222.4 |
| $MPS$ | 194.6 | 209.5 | 231.9 | 244.0 | 220.0 |
| $MPS^{-1}$ | 191.3 | 204.6 | 226.1 | 238.3 | 215.1 |
| $MPS^{-2}$ | 189.7 | 202.2 | 223.3 | 235.4 | 212.7 |
| $MPS^{-3}$ | 188.2 | 199.8 | 220.5 | 232.6 | 210.3 |

As one can see if one of the approaches is correct then using a different one would cause predictions based on the interpolated value miss by about 2.4 s. We use the interpolated values as long as $MA_h(Differential\ Pressure_t) \leq \theta$; which was the case for about 80% of all cases. Since error on the test set is multiplied by 1.5, using wrong interpolation would increase our average penalty score by 2.4 s * 80% * 1.5 = 2.88 s. Therefore, choosing the right interpolation variant has greater effect on the final model performance than trying to reduce the error we can measure on the available data even to 0. We were considering models that would check whether differential pressure buildup speed for unknown middle particle size is more similar to speed observed for smaller or bigger particles and make prediction based on assumption that useful life will be similar to the same type of particles. However, we did not have data to validate such a hypothesis and therefore we resigned from this approach.

#### 2.4.2. The Need for Information

This analysis led us to the decision that we should focus all our efforts on obtaining more information about how contaminant particle size affects the clogging process. That would help us choose the right variant or build a model that would not implicitly guess the relationship between useful life and middle particle size.

We identified following possible sources of such information:

- physics,
- building our own test rig and running experiments for another filter,
- other publicly available data showing clogging process of liquid filter.

**Physics**

The clogging process seems to be complex and have multiple stages. We decided that it would be risky to assume we can calculate what will be happening for the unseen middle particle size without any option to validate our results against data.

**Building Our Own Rig**

We did not have enough time to do so.

**Other Datasets**

We started to search for publicly available datasets. Such a dataset could be used: (i) for inspiration, (ii) to validate physics-deduced thesis, (iii) for transfer learning – popular approach in image processing and (iv) for choosing an interpolation method.

Unfortunately, we did not find such a dataset. Instead we found a paper (Skaf, Eker, & Jennions, 2017) describing similar experiments with a figure showing differential pressure buildup that contained different particle sizes, including the range that we needed (see Figure 6). While we found that the data from these experiments did not match exactly with our use case, they were gathered using experiments of a similar nature. We have used information contained in the figure to choose the best interpolation method.

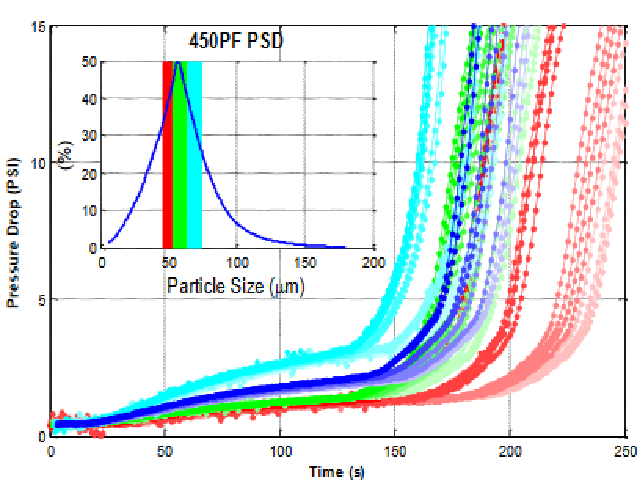


Figure 6. Differential pressure evolution in time for different particle sizes (Skaf, Eker, & Jennions, 2017).

### 2.5. Weights

In this section we describe how the weights $\vec{w}$ used to average useful life in between different particle sizes to obtain the estimate for an unknown particle size, as described in Section 2.1, were selected.

Cyan and red lines on Figure 6 correspond to particle sizes 63-75 µm and 45-53 µm, respectively. These are the sizes that were included in the training data set. Green lines correspond to particle size of 53-63 µm. Note that the figure presents differential pressure only up to 15 psi. For each interpolation method mentioned in Section 2.4.1, we compared the time to reach a differential pressure value of 15 psi estimated by this method for particle size of 53-63 µm with the time interval in which green lines hit the top of the chart. It turned out that the method that assumes that useful life is proportional to $MPS^{-3}$ best fits the chart.

Recall that when we calculate $UL$ for $MPS = 58$, vector $\vec{w}$ contains two values: for $MPS = 49$ and for $MPS = 69$. The weights are:

$$\vec{w} = \left(\frac{69^{-3} - 58^{-3}}{69^{-3} - 49^{-3}}, \frac{58^{-3} - 49^{-3}}{69^{-3} - 49^{-3}}\right) \approx (0.381, 0.619)$$

Such that e.g.:

$$UL_{r,58} \approx 0.381 \cdot UL_{r,49} + 0.619 \cdot UL_{r,69}$$

## 3. MODEL EVALUATION

During training and evaluation of the model we used some portion of data for training and all the available data for evaluation, as final score would be calculated on all the data. During such an evaluation models $m_{100}$, $m_{75}$, $m_{50}$ and $m_{25}$ achieved scores shown in Table 5.

Table 5. Scores of the models.

| Solid Ratio | MAE |
|---|---|
| $m_{100}$ | 1.303 |
| $m_{75}$ | 1.305 |
| $m_{50}$ | 1.493 |
| $m_{25}$ | 1.846 |
| Total | 5.947 |

We know that our Final Total Score was 49.67, so, based on Eq. 2., we can calculate that the sum of errors of our four models on Test set was 29.149, which is about five times worse than the results on the published data. It is not a surprise. Our approach was heavily dependent on low variation in useful life among samples of the same contamination profile. It was very helpful in obtaining low score on samples that have contamination profile that was represented in training set of a particular model but rendered the model vulnerable to inaccuracy of useful life interpolation.

## 4. FURTHER WORK AND DISCUSSION

In practical applications we usually require more accuracy from a model when a failure is closer than when it is far away. This is reflected by relative measures, like e.g. Relative Accuracy proposed by Saxena, Celaya, Saha, Saha, and Goebel (2010). It would be interesting to evaluate our model versus such measures. The importance of appropriate weights in statistical approach used by our model at early stage of an experiment (before threshold differential pressure is reached) would be smaller, but we would get a more restrictive evaluation of the kernel regression part that is used in crucial period when the clogging is going to occur soon. Unfortunately, such an evaluation cannot be done if the test data set is not published.

## NOMENCLATURE

| | |
|---|---|
| $PRUL_m(\mathrm{s,t})$ | Predicted Remaining Useful Life of a sample *s* at a time instant *t* by a model *m* |
| $RUL(\mathrm{s,t})$ | Remaining Useful Life (ground truth) of a sample *s* at a time instant *t* |
| $E(\mathrm{m,S})$ | Error function of a model *m* measured on a set of samples *S* |
| *Train* | Set of all samples from training data |
| *Test* | Set of all samples from test data |
| $m_p$ | Model built using *p*% of the available samples for training |
| *UL(s)* | Useful Life of a sample s since experiment start until clogging |
| $UL_{r,p}$ | Median Useful Life since experiment start among samples from *Train* with solid ratio of r and middle particle size of p |

| | |
|---|---|
| $\widehat{UL}(s)$ | Useful Life since flow start until clogging for a sample *s* |
| $\widehat{UL}_{r,p}$ | Median Useful Life since flow start among samples from *Train* with solid ratio of *r* and middle particle size of *p* |
| $\mathrm{MA_h(x_t)}$ | Moving average of a time series *x* with sliding window of size *h* |
| *FlowStart*(*s*) | The first time instant when the flow rate was at least 100 ml/min for a sample *s* |
| *SR(s)* | Solid Ratio of a sample *s* |
| *TrainSR* | Set of all Solid Ratios occurring in the *Train* dataset |
| *MPS(s)* | Middle Particle Size of a sample *s*. It is average of min and max particle size |
| *TrainMPS* | Set of all Middle Particle Sizes occurring in the *Train* dataset |
| $wavg(\vec{X},\vec{w})$ | Weighted average of elements of a vector $\vec{X}$ with weights $\vec{w}$ |
| h | Window size of a moving average |
| $\theta$ | Differential pressure threshold above which we are using kernel regression for predictions |

## REFERENCES


Skaf, Z., Eker, O. F., & Jennions, I. K. (2017). System component degradation: filter clogging in a UAV fuel system. *Asia Pacific Conference of the Prognostics and Health Management Society* (334-342), July 12-15, Jeju, Korea.

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for Offline Evaluation of Prognostic Performance. *International Journal of Prognostics and Health Management*, vol. 1 (1).

Seabold, S., Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference, June 28 - July 3, Austin, Texas*, US.

PHM Data Challenge (2020). [Online], http://phmeurope.org/2020/data-challenge-2020


## BIOGRAPHIES

**Roman Łomowski** Graduated from University of Warsaw with MSc in Computer Science and BSc in Mathematics. He has more than 10 years of experience in software development and more than 3 years of experience with machine learning. He works in Syncron as Senior Data Scientist. During his education he achieved: silver medal on International Mathematical Olympiad twice (Washington 2001, Glasgow 2002), First and Second Prize at International Mathematics Competition for University Students (Skopje 2004, Cluj-Napoca 2003 respectively).

**Szczepan Hummel** received a PhD degree in Computer Science and M.Sc. degree in Mathematics from University of Warsaw, Poland. He holds a position of a Principal Data Scientist at Syncron. He gained experience in building predictive models at Bunge, where he worked mainly with agricultural and trade data. His previous research interests were automata theory, formal languages, and descriptive set theory. Currently he concentrates on building failure prediction and anomaly detection models based on sensor readings data using machine learning and statistics.