DATA CHALLENGE 2020

NEW: We have added a FAQ to help candidates: **link**.

The PHM Society invites you to join other participants around the world from academia and industry in the PHME20 Data Challenge. An experimental **filtration system** is the subject of the competition and the challenge is to create high-performance prognostic models to estimate Remaining Useful Life (RUL).

Model adaptability and **sensitivity** are key performance criteria in this competition as prediction capability will be tested and assessed as a function varying experimental parameters and number run-to-failure data sets.

Application of **Data-Driven** as well as **Physics-based** modelling approaches are encouraged.

Winning teams will be invited to publish their work in the conference proceedings of the 5th European Conference of the PHM Society and a representative will be expected to make an oral presentation at the event.

Collaboration is encouraged and teams may comprise students and professionals from single or multiple organizations.

Competition Open February	February 14, 2020
Validation Data Posted Submission Portal Open	April 24, 2020
Competition Closed	May 15, 2020, 11:59:59 GMT
Winners Announced	May 29, 2020
Final Papers Due	June 7, 2020
PHM Conference Date	July 1-3 2020

Objective

Participants are challenged to develop models to provide estimates of **Remaining Useful Life (RUL)** of a filtration system. 'Run-to-Failure' data is provided with key condition monitoring parameters: flow rate, upstream pressure (before filter) and downstream pressure (after filter). Data is generated under controlled conditions, with key experimental variables of contamination particle size and concentration, i.e. solid-to-liquid ratio.

Model adaptability and sensitivity are key performance criteria in this competition as prediction capability will be assessed as a function of varying contamination particle size, concentration, and number of run-to-failure experiments. Training datasets will be made available immediately after the competition is launched and validation datasets will be provided prior to closing of the competition. Application of Data-Driven as well as Physics-based modelling approaches are encouraged.

The winning teams will be asked to prepare a full manuscript which will be featured in the PHM conference proceedings and a representative will be expected to make an oral presentation at the event. The prize will be awarded at the conference banquet.

For any questions about the competition, please **contact us**.

Teams

Collaboration is encouraged and teams may comprise students and professionals from single or multiple organisations. There is no requirement on team size. Register your team's entry by **filling the form**.

IMPORTANT: You can register to the Data Challenge (free of charge) without register to the conference and decide later to attend or not the PHM Europe 2020. The registration to the conference is not mandatory.

The winning teams will be selected and awarded contingent upon:

- Having at least one member of the team register and attend the PHM 2020 [RKM3] Conference.
- Submitting a peer-reviewed conference paper.
- Presenting the analysis results and technique employed at the conference.

The organizers of the competition reserve the right to modify these rules and disqualify any team for any efforts it deems inconsistent with fair and open practices.

System Description

An experimental rig to demonstrate filter clogging failure has been constructed and consists of the following major components: Pump, liquid tanks, tank stirrer, pulsation dampener, filter, pressure and flow rate sensors, data acquisition system connected to a computer.

The experiment rig is a circuit composed of a pump flowing a liquid from a tank to another through a filter. The circuit is instrumented with sensors able to monitor the **flow rate**, and the **liquid pressure before** and **after** the filter. The fluid injected in the system is a **suspension** composed by Polyetheretherketone (PEEK) particles and water with different concentrations. To eliminate possible pulsations in the flow the circuit includes a dampener. Figure 1 depicts the employed experimental rig.

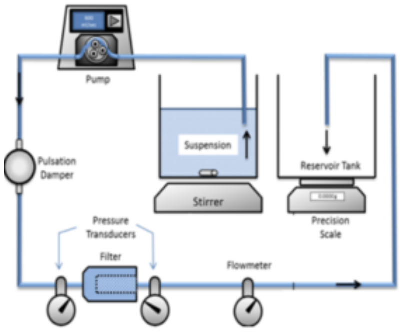

Figure 1: System of the experimental rig

The main components applied to construct this experimental rig are the following:

- **Pump**: Since the system will involve contaminants in the fluid, a peristaltic pump has been used as its mechanism is more tolerant to particles in the liquid. A Masterflex® SN-77921-70 (Drive: 07523-80, Two Heads: 77200-62, Tubing: L/S© 24) model peristaltic pump was installed in the system to maintain the flow of the prepared suspension. The pump is a positive displacement source, providing a flow rate ranging from 0.28 to 1700 ml/min (i.e. from 0.1 to 600 RPM).
- **Dampener**: The aim of using rigid tubing is to prevent the system from unwanted tube expansion due to pressure build up, which affects the actual pressure build up generated from filter clogging. A Masterflex® pulse dampener is installed on the downstream side of pump to eliminate any pulsation in flow.A majority of the system is furnished with a rigid polypropylene tubing, whereas the pump side is covered with a flexible Tygon® LFL pump tubing.
- **Particles**: The suspension is composed of Polyetheretherketone (PEEK) particles and water. PEEK particles have a density (1.3g/cm3) close to that of room temperature water and have significantly low water absorption level (0.1% / 24 hours, ASTM D570). Having a low water absorption level will prevent particles expanding when they mix with water. Subsequently, closer density with water allows particles to suspend longer in water.
- Flow Rate Sensor: A GMAG100 series electromagnetic flow meter (measurement range: 3 – 25,000 millilitres per minute) is installed in the system to keep track of the flow rate in the system.
- **Pressure Sensors:** Upstream and downstream Ashcroft® G2 pressure transducers (measurement range: 0 – 100 PSI) are installed in the system to capture the pressure drop (i.e. 'ΔP') across the filter, which is considered as the main indicator of clogging.
- **Filter**: the filter has a pore mesh size of 125μm as shown in Figure 2.

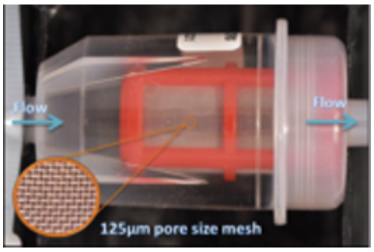

Figure 2. Filter under study

With this experimental rig, several experiments have been run with suspensions of different particle size and concentration. As concentration, for each particle size we perform 4 experiments with the concentration reported in Table 1.

Table 1: Suspension profile details

	Profile Number			
	1	2	3	4
Water (g) in the suspension tank	7968	7497	7079	6704
Particle (g)	32	32	32	32
Solid ratio	0.004 (0.40%)	0.00425 (0.43%)	0.0045 (0.45%)	0.00475 (0.48%)

To create the suspensions we add particles of three possible sizes: small, medium, large. Details about the particle size are reported in Table 2.

Table. 2: Particle size data

	Particle Size (µm)
Small	45-53
Medium	53-63
Large	63-75

Datasets and Challenge

The experimental dataset is composed of the **flow rate** (Flow_Rate) of the liquid, and the **pressure before** (Upstream_Pressure) and **after** (Downstream_Pressure) the filter, acquired at 10 Hz. To identify when the system fails, we compute the Pressure Drop (Upstream Pressure – Downstream Pressure), and we identify the filter as clogged whenever the pressure drop is higher than 20 psi.

The objectives of this data challenge are: (i) to predict when the filter is fully clogged creating models able to determine accurately the Remain Useful life of the filter, (ii) to assess the capability to create a model with performance that do not degrade when a smaller portion of the available dataset is used to create the model itself.

For these two sets of data are provided in different moments of the challenge.

Model training: 24 experiments describing the run to failure of the filter are used as training datasets. The dataset includes experiments with small and large particle sizes and different particle concentrations. For each particle size and concentration 4 experiments are reported. Table 3 describes the training dataset.

Model validation: 8 experiments describing the run to failure of the filter are used as validation datasets. The datasets include concentrations that were not available during the training phase. For each particle size and concentration 4 experiments are reported.

Table 3: Training set – 24 samples

Profile Number	Particle Size	Solid Ratio (%)	Sample Size
1 2 3	45-53 small	0.4 0.425 0.45	4 4 4

1 2 3	63-75 large	0.4 0.425 0.45	4 4 4

Performance evaluation

Participants are expected to generate 4 models to predict accurately the maximum useful life of a filter, and estimate the remaining useful life of the filter in different moments.

Each team has to build the following models:

1. Using **all the available experiments** (i.e., Model training + Model validation), to create the model
2. Using **75% of the available experiments** (i.e., Model training + Model validation), to create the model
3. Using **50% of the available experiments** (i.e., Model training + Model validation), to create the model
4. Using **25% of the available experiments** (i.e., Model training + Model validation), to create the model

The performance of **each model** will be evaluated based on an *error function*, evaluating the accuracy of the model with:

- the training set and validation set together
- a test set composed by experiment not available for training and validation. Testing on an independent dataset is performed to assess the generality of the deployed models.

Fig. 3 reports an example of the error function trend based on the amount of data used to create the model (x-axis), and the datasets used to evaluate the performance.

Figure 3: Example of performance trend

A *penalty score* is used to combine the accuracy of each model and the performance degradation introduced by reducing the amount of data used to build the model.

Details of the error function and the penalty score are provided on the « Penalty Score » tab. In addition, a spreadsheet with the calculations has been provided in the « Downloads » tab (for logged-in contestants).

Submission files

For the submission participants are required to submit a Jupyter Notebook describing the performed process, the 4 created models (e.g., pickle file), usable for the error function evaluation, and a short paper (maximum 4 pages) describing the applied methodology.

Penalty score

M_i = Model generated with i % experiments
TV = Training + Validation datasets
TE = Test dataset

For the Error Function we use the *Mean Absolute Error (MAE)*

Penalty Score:
$\text{Penalty(TV)} = \sum \text{MAE}(M_i(\text{TV}))\ i \in \{25,50,75,100\}$
$\text{Penalty(TE)} = \sum \text{MAE}(M_i(\text{TE}))\ i \in \{25,50,75,100\}$

Penalty Score = 1.5*Penalty(TE) + Penalty(TV)

Downloads

Remind that to register to the Data Challenge is free of charge.

After completing your **registration form**, the downloadable files will be available on your **profile page**.

phme20
TURIN (Italy) - 1/3 July 2020

SECRETARY PHME20

For any request, please contact:

+33 (0)2 47 27 33 30
secretary@phmeurope.org

KEY DATES

Abstracts submission EXTENDED deadline: **22 Feb 2020**
Paper and poster submissions due: **15 Apr 2020**
Paper review feedback: **15 May 2020**
Final paper and poster due: **31 May 2020**