Experiment Design for Waste Heat Recovery Modeling in Heavy Duty Trucks

Francesco Galuppo [*,**,***] **Pascal Dufour** [**] **Madiha Nadri** [**]
Thomas Reiche [*] **Vincent Lemort** [***]

[*] *Volvo Group, Powertrain Advanced Engineering, 99 route de Lyon, 69800, Saint Priest, France (e-mail: francesco.galuppo@volvo.com, thomas.reiche@volvo.com)*
[**] *Univ Lyon, Université Claude Bernard Lyon 1, CNRS, LAGEP UMR 5007, 43 boulevard du 11 novembre 1918, F-69100, Villeurbanne, France. (e-mail: pascal.dufour@univ-lyon1.fr, madiha.nadri-wolf@univ-lyon1.fr)*
[***] *Thermodynamics Laboratory, University of Liège, Campus du Sart Tilman, B49, B-4000 Liège, Belgium (e-mail: vincent.lemort@ulg.ac.be)*

Abstract: Transient working conditions of a waste heat recovery (WHR) system in a heavy-duty (HD) truck application require the control of internal variables of the thermodynamic Rankine cycle for the final large scale commercial integration of this technology. The intensive test demand and the large number of possible operating points of the HD truck engine suggest paying particular attention to the choice of the engine working points to use for model identification needed in further control design. This paper presents a methodology for the design of the engine working points to test in experiments in order to proceed to an open loop multi-linear and multi-structure model identification task from experimental data.

Keywords: Energy recovery systems; Powertrain modeling; Modeling for control

1. INTRODUCTION

A WHR system based on Rankine thermodynamic cycle has for main ambition to recover thermal energy from a heat source and then to convert it into mechanical energy. The organic Rankine cycle (ORC) is a particular Rankine thermodynamic cycle where the working fluid (WF) is organic. Organic fluids are characterized by high molecular mass, with lower boiling point and latent heat with respect to water (and other traditional fluids), which allows to maximize the quantity of heat recovered from low temperature heat sources like in HD trucks (Rijpkema et al., 2018). In an ORC the choice of the organic WF is an open question since no single optimal fluid exists: It impacts the ORC efficiency while fluid deterioration, environmental aspects or freezing must be accounted for (Stijepovic et al., 2012).
An ORC based WHR system for HD trucks can also be characterized by different architectures: the most investigated architectures envisage the recovery of thermal power from only exhaust gases or the simultaneous recovery of thermal energy from exhaust gases and EGR flow. The four major components (pump, exhaust gases heat exchanger, expander and condenser) in a ORC are shared by both these architectures, while the second architecture is characterized by one more major component, a heat exchanger for the recovery of thermal energy of Exhaust Gas Recirculation (EGR) flow. Research development of a new ORC configuration requires its optimization: 1) modifications of components to test, 2) vehicle integration, 3) for any architecture evaluation, experimental tests must be performed under controlled situations. Indeed, without a proper control strategy, this could result in unsafe working conditions for the WHR system components or in very low production rate. Moreover, to fit a realistic scenario, these tests are based a dynamic driving cycle of the HD truck, inducing large transient operations of the ORC.
On the other hand, due to the complexity of the phenomena involved in such an ORC (nonlinearities, couplings, phase change) the use of classical and simple controllers like a PID is prohibited. Hence a model based control development is required (Hernandez et al., 2016). Detailed modeling is often an important investment that assumes that the components, architecture and working fluid are chosen. Therefore, model based control is a major problem in the management of a WHR system based on an ORC. Many new studies (Seitz et al., 2018), (Zhao et al., 2018) are focusing on this important aspect that should allow the final ORC integration in serial production (Tona and Peralez, 2015).
Moreover, intensive experimental test demand is a critical issue when testing a WHR system. Indeed, in a dynamic driving cycle, the HD truck engine works in many operating points, which requires a potentially large number of experiments for modeling development. Therefore a particular focus is needed on the a priori design of these experiments to get few operating points that are meanwhile sufficiently representative of the global engine operation.

⋆ The French ministry of higher education and research is acknowledged for the financial support of this CIFRE PhD thesis 2016/1205.

For example in (Grelet et al., 2015a), the choice of the 14 operating points is not clear (in terms of number and characteristics).

This paper is focusing on developments where ORC architecture, WF and new major components are still evaluated, and yet model based control is needed for the performance evaluation: it would be costly and difficult to develop an accurate first principle model for each configuration (Grelet et al., 2015b). Then, we advocate to use a quick linear model based control approach for each configuration based on first experimental engine test with a specified driving cycle. The idea is to be able to have a first step of analysis based on a simple controller in order to be able to choose the final set of architecture, WF and components that will be later studied in more details in the second step in a detailed first principle modeling approach used for more advanced control design.

For any chosen ORC configuration, the methodology presented here aims first to reduce the size of new ORC experiments to perform in the view of linear multi-model development. From these new experiments, the experimental data analysis allows to choose several simple linear model structures from a library. Several identification algorithms can then be used to estimate the model parameters (the identification method itself is not the core of this paper). Then, the methodology aims also to guide the choice of model structure to be used for the control design; depending on the experimental results, the analysis will allow to choose the model structure to use in the linear controller design: 1) it might be a single linear model (usually, a First Order Plus Time Delay model (FOPTD) is used), 2) or a multi model approach based on one linear model structure (Aufderheide and Bequette, 2003; Grelet et al., 2015a) (which is the classical multi model approach, usually based on FOPTD), 3) or a multi model approach based on multiple linear model structures (up to the knowledge of the authors, such an approach is not present in the literature). Then, once the best controlled configuration is found from this analysis, keep the linear controller as such (if the performances are sufficient) or invest on a more detailed model development of this selected configuration in order to design an advanced model based controller (possibly non linear like in (Peralez et al., 2017)) that should improve the controlled WHR performances.

The paper is organized as follows: Section 2 presents the motivation and problem statement. Section 3 presents the novel procedure to select the working points used for further multi-linear model identification. In Section 4 the methodology is applied to a case study. The available test bench deals with an energy recovering from hot exhaust gases in a 13Lt Euro 6 HD truck engine with ethanol as WF.

2. MOTIVATION AND PROBLEM STATEMENT

This paper focuses on a development phase where the architecture, the WF and the main components are to be tested experimentally one by one. Yet for each case, a controller must be used. In an ORC, the WF thermodynamic conditions at the inlet of the expander are important and should be carefully controlled. Indeed, a wet expansion with presence of droplets of liquid could damage the expander machine. To avoid this, it is important to have a tight control of the inlet expander temperature to prevent liquid fluid from entering the expander machine. This is usually done by controlling the superheat $SH_{out,ev}$ to a chosen set-point. $SH_{out,ev}$ is defined as the difference of temperature between the actual temperature $\theta_{out,ev}$ of the WF at the inlet of the expander and the saturation temperature $\theta_{sat}(p_{out,ev})$ of the WF at this location at the actual pressure $p_{out,ev}$. A positive value of superheat indicates that the WF is in vapor state. Therefore the superheat value should be always maintained over zero when the expander is operating (Xu et al., 2017). This key variable is very important because it also significantly affects the efficiency of the expander. This behavior has to be controlled by the manipulated variable (process input); the pump rotational speed of the ORC system that provides the whole flow rate of WF in the system. Due to the nonlinearities, couplings and WF phase changes, this behavior is complex to be controlled (Tona and Peralez, 2015; Grelet et al., 2015a). It also depends on an external signal entering the ORC: the driving cycle, which induces a dynamic load to the engine and, by consequence, a dynamic heat recovery.

Otherwise, for a particular ORC configuration, the role of the engine and the driving cycle is crucial: The thermal power entering the ORC depends on the temperature and mass flow of the exhaust gases, and eventually on the EGR flow which is provided by the engine. The operating points of the engine are influenced by the driving cycle (highway conditions encountered by the truck, driving of the driver, slopes, ...). As a consequence, the available thermal power Q (W) is extremely variable during the driving cycle and represents a important disturbance for the system, and hence the SH control. Therefore the driving cycle and the engine type must be accounted for the control oriented modeling.

In order to consider all the points in which the engine truck operates during the driving cycle, it is useful to represent them in an engine map, where the couple of values, engine speed, Ω (rpm), and engine torque, T (Nm), unequivocally identifies an operating point. Depending on the chosen driving cycle, the engine map usually contains a few tens of thousand points. The main issues of the experimental study of an ORC are the short time available for the test and its related cost. Therefore, it is not possible to handle and test all the operating points in the engine map in an experimental campaign for a ORC based WHR system. By consequence, the engine map, which contains all the operating points of the engine for a specific driving cycle, has to be analyzed by taking into account a sufficient reduced number of points.

Two questions will be answered in this paper: how to design a reduced set of operating points in the engine map to test experimentally while keeping a large spectrum of different behaviors of the ORC ? Then based on these experimental data how to design a multi linear model used for further control design ?

3. METHODOLOGY

This section presents a methodology to establish the set of representative operating points of the engine map over the driving cycle that will be used in the experimental campaign. This result will be used to design and validate the linear multi-model structure.

3.1 Engine Map and Driving Cycle Discretization

Let T_s be the constant sampling time (in s) of the data contained in the driving cycle (sampling time of the provided data is of the order of $0.1s$). Since the duration of a classical driving cycle for long HD truck is usually between 1 and 5 hours, the number of time samples $N_s \in \mathbb{N}$ in a driving cycle is of order of 10^5. The engine map characterizes each point during the road cycle by the following triplet: engine torque T, engine speed Ω and available thermal power Q.

In order to decrease the number of samples N_s needed for the model based control analysis, a first discretization of the engine map by clustering data based on a discretization step ΔT for the torque and $\Delta\Omega$ for speed is performed. One cluster is one operating point with one underlying local dynamic model for the relation between the considered input(s) and output(s). The couple of values determining the clusters in terms of T and Ω are considered at the center of the cluster. The ratio of the time cycle spent in a cluster compared to the total driving cycle duration is the frequency of occurrence f_o (%). By removing the clusters that are usually empty in the engine map (many of the possible combinations of T and Ω do not take place in the engine operation), it is possible to reduce the number of clusters to N_c, as the number of clusters to be analyzed. Based on the engine map and the discretization step tuning, the initial number N_s of samples is therefore decrease to N_c, usually with a ratio of 100. Hence, the order of $N_c = 10^3$ clusters must still be handled; but this is still high and a new criteria has to be defined in order to reduce it.

3.2 Reduction by Frequency of Occurrence and Power Analysis

Let us now define the engine map matrix $M_{EM} \in \mathbb{R}^{N_c \times 5}$ with columns as follows: cluster index (sorted by ascending cluster numbering), engine torque T, engine speed Ω, heat recovery available Q and frequency of occurrence f_o. The proposed procedure for N_c reduction consists of the following steps, that are also summarized in the Algorithm1:

Step 1: sorting M_{EM}:

- by descending frequency of occurrence f_o (i.e. the first (the best) cluster is the most common cluster during the cycle). This leads to build the matrix $M_{EM}^{f_o} \in \mathbb{R}^{N_c \times 5}$.
- in parallel, by descending values of heat recovery available Q to design the matrix $M_{EM}^{Q} \in \mathbb{R}^{N_c \times 5}$ (i.e. the first (the best) cluster is the one characterized by the highest content of thermal power).

Step 2: a combination of the obtained matrices: let us now set $N_c^t \in \mathbb{N}$ ($N_c^t < N_c$), as the reduced target number of clusters we want to use for further model based control development. $M_{EM}^{f_o}$ and M_{EM}^{Q} are analyzed together (by finding the N_c^t operating points in common) to build $M_{EM}^{f_o \cap Q} \in \mathbb{R}^{N_c^t \times 5}$.

Step 3.: the first column of $M_{EM}^{f_o \cap Q}$ contains the N_c^t indexes of the operating points that have to be tested in experiment.

Algorithm 1 : Choosing ORC working points

$set\ M_{EM},\ M_{EM}^{f_o},\ M_{EM}^{Q}$
$set\ N_c^t$
$N_c^i \leftarrow N_c^t$
$test \leftarrow true$
$while\ test\ is\ true$
  $M_{EM}^{f_o \cap Q} \leftarrow M_{EM}^{f_o}(lines\ 1\ to\ N_c^i) \cap M_{EM}^{Q}(lines\ 1\ to\ N_c^i)$
  $if\ number\ of\ lines(M_{EM}^{f_o \cap Q})\ is\ N_c^t$ then
    $test \leftarrow false$
  $else$
    $N_c^i \leftarrow N_c^i + 1$
  $end\ if$
$end\ while$

These N_c^t operating points, each of them described by a torque T and an engine speed Ω, lead therefore to the experiments to run. Also, based on the analysis of $M_{EM}^{f_o \cap Q}$, the sum of the N_c^t occurrences shows the percentage P_{DC} (%) of the driving cycle which is indeed accounted for in the analysis.

3.3 Model Structure and Identification

The N_c^t experiments that have been designed in the previous subsection are now performed on the real ORC setup. The manipulated process variables are now tuned to a constant value (one vector of values for each experiment): the drawback is that they are adjusted by trial and error in open loop control to get similar specified controlled variables at steady state for all experiments. From the particular steady state obtained in each experiment, input signals must now be designed to get the data used for identification procedure. Since a linear model control structure is also sought, it is advocated to use step inputs. Indeed, the shape of the temporal response is, in this case, easier to interpret to obtain the linear structures of the model. Moreover, the use of an excitation signal like pseudo random binary sequence (PRBS) allows to get richer data and to avoid the process outputs to have less uncontrolled dynamic behaviors.

4. CASE STUDY: 13 LT EURO 6 ENGINE

4.1 Experimental Test Bench Description

The case study at Volvo (Saint-Priest, France) deals with a WHR system where exhaust gas thermal power is recovered in order to produce mechanical power. In the test cell, a 13 Lt Euro 6 engine and cooling water system test bench are available; the engine provides the hot exhaust gas which enters the evaporator, while the water system simulates the behavior of a cooling system of HD truck. The WHR system includes components, sensors and valves

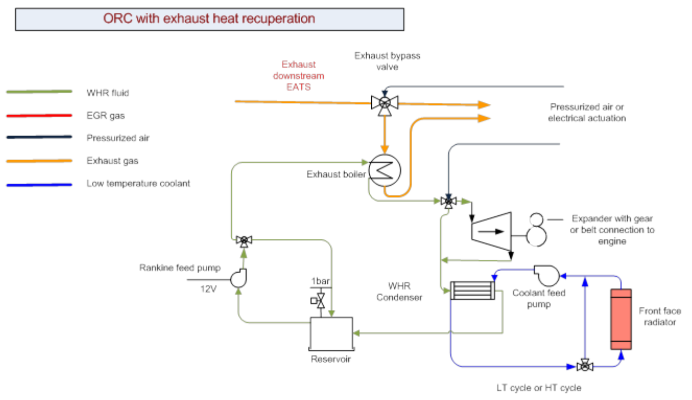

Fig. 1. Experimental test bench ORC architecture

(see Fig. 1). The pump provides the total flow rate of WF, which receives thermal power from the exhaust gas in the evaporator. For this experimental campaign, whose main ambition is not producing mechanical power but identifying specific operating points for control oriented modeling, the WF circulates into the by-pass path in parallel with the expander machine (which maintains the role of imposing the high pressure in the circuit via its nozzle). Here the WF pressure decreases into the by-pass valve and the WF is cooled down and condensed in the condenser, which is cooled by the cooling water system. It is important to notice that the circulation of the fluid into the by-pass path does not influence the thermodynamic variables of the WF at the inlet and outlet of the evaporator. During the experimental campaign, due to the planning of new component receipts, two different configurations of the test bench were performed; the key difference is in their dealing with the expander machine (and its nozzle). Both expanders belong to the family of volumetric machines, but the first machine (tests 1 to 11) admits lower pressure levels with respect to the second machine (tests 12 to 37).

4.2 Engine Map and Driving Cycle Analysis

The methodology proposed in this paper is applied considering a French highway driving cycle Lyon-Chambéry-Grenoble (LCG) of almost 3 hours, characterized by $N_s = 96000$ time samples (sampling time $T_s = 0.1\ s$). The discretization ($\Delta T = 50Nm$, $\Delta\Omega = 50rpm$) of the normalized engine map (Fig. 2) has been tuned according to the engine map, taking into account small variations in terms of thermal power (Fig. 3) of two operating points located in the same cluster. After discretization the number of non empty clusters N_c, to each of them corresponds an operating point, is 1620.

4.3 Frequency of Occurrence and Power Analysis

Following Algorithm 1 the N_c operating points are sorted by frequency of occurrence. In the same time those operating points are sorted by available thermal power. Hence the two matrices $M_{EM}^{f_o}$ and M_{EM}^{Q} are obtained. Starting from the chosen number of operating points N_c^t that is allowed, it is now possible to determine the percentage P_{DC} of the total operating points in the driving cycle that are represented (see Table 1). For this case study $N_c^t = 37$

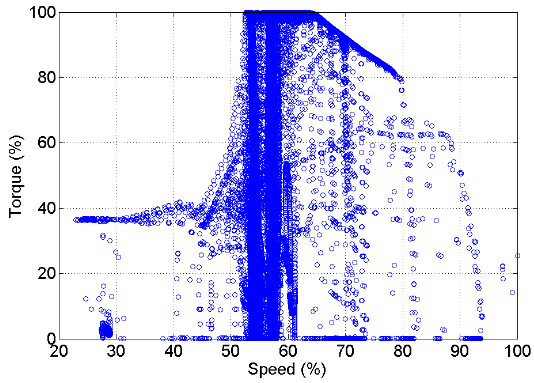

Fig. 2. Engine map: occurrence in the LCG driving cycle

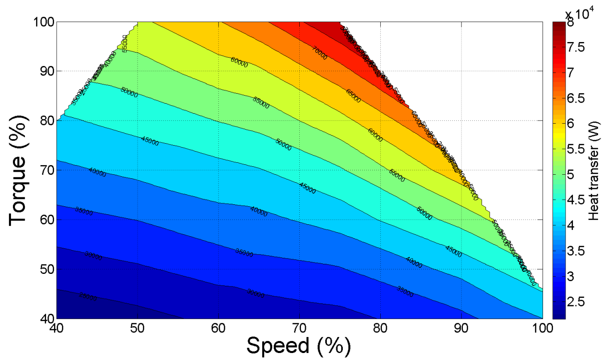

Fig. 3. Engine map: Thermal power in the LCG driving cycle

is chosen, according to the constraints related to the experimental campaign already mentioned in the introduction. This number of operating points represents roughly 56% of the total points in the driving cycle. Therefore a good compromise is found since only $\frac{N_c^t}{N_c} = 2.3\%$ of the possible operating points are enough to cover more than 50% of the driving cycle. Hence, this tuning of N_c^t is a good compromise between the cost and time of experiments, the availability of the cell at that time on the one hand and the modeling effort needed on the other hand. Once the number of points to test is defined, $M_{EM}^{f_o \cap Q}$ unequivocally determines the points to test, in terms of engine speed and engine torque (Fig. 4) where three main regions are identified: i) the lower left corner of the map, showing a point which is characterized by high occurrence in the driving cycle and weak thermal power content with respect to the other points; ii) the lower central one, which corresponds to the cruise region in the engine map, where high frequency of occurrence is still dominant; iii) the upper central region, which is characterized by high thermal power content operating points.

Table 1. Influence of the tuning of the number of experiments over the ratio of operating points covered

N_c^t $(-)$	30	37	55	120	150	1620
P_{DC} (%)	50	56	70	90	95	100

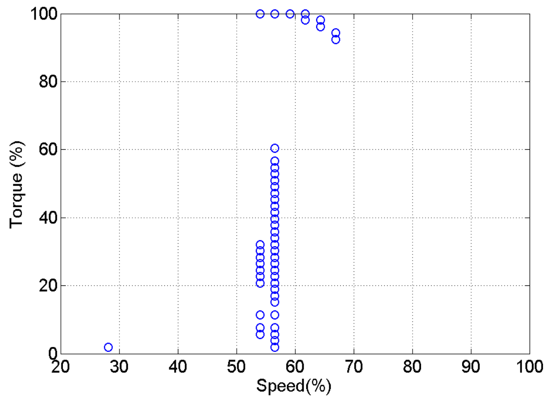

Fig. 4. Engine map: Final points to test in the LCG driving cycle

4.4 Experimental Step Response Analysis

The N_c^t experimental tests are performed to reach a particular steady state (since the engine operating point does not change during one experiment). For each operating point, once the steady state is reached, a step input on the pump speed is performed for the identification task. The step input induces a variation of the output $SH_{out,ev}$ (e.g. Fig. 5 for the experiment 35) which is, as previously mentioned, a function of temperature and pressure at the outlet of the evaporator. A positive (negative) variation of pump speed (and by consequence of the mass flow rate of working fluid) implies a negative (positive) variation of $SH_{out,ev}$. In fact, an increase (decrease) of the mass flow rate, at constant operating point (constant thermal power to the ORC), causes a temperature reduction (raise) and the a raise (reduction) of pressure. The combination of the two variable variations implies a change in the superheat, which is faster immediately after the step and slower when pressure and temperature stabilize.

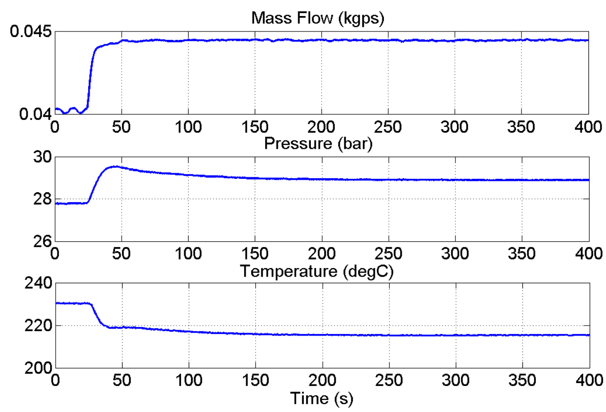

Fig. 5. Experiment 35: Step analysis

4.5 Modeling and Identification

The case study represents a Single Input Single Output system (SISO) since the output variable, $SH_{out,ev}$, is directly influenced by the total mass flow rate of working fluid, which is the input. Based on the shape of the 37 experiment step response (see Fig. 6 for experiments 22, 24, 35) three model structures are employed:

(1) First order (FO) model:

$$F(s) = \frac{G}{1 + \tau s}$$

(2) FOPTD model:

$$F(s) = \frac{G.e^{-Ls}}{1 + \tau s}$$

(3) Mix of a FO model and a FOPTD model (FO-FOPTD):

$$F(s) = \frac{G_1}{1 + \tau_1 s} + \frac{G_2.e^{-Ls}}{1 + \tau_2 s}$$

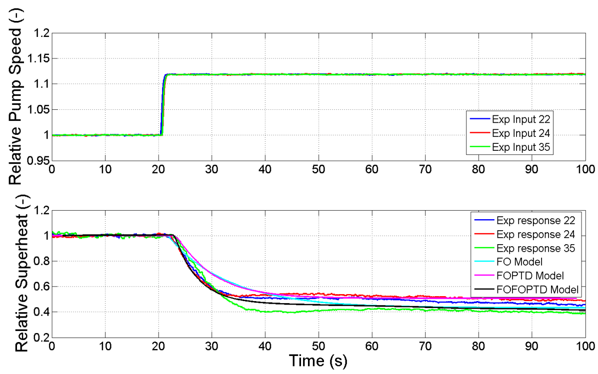

Fig. 6. Experiments 22, 24, 35: normalized experimental inputs and responses, model response with the best model. Experiment 22 gives the normalized values.

Where G_i, τ_i, L represent respectively a static gain, a time constant and the time delay. For each experiment, the identification procedure selects one by one the three model structures and adjusts the model parameters to reach the best possible approximation (in terms of root mean square error). A local optimization algorithm is performed with Matlab (fmincon). The modeling errors are computed for all experiments and the three models (see Fig. 7): These results tell us that the modeling task is always achieved with a final error between 2 % and 9% for the three applied model structures, but not a single model can be considered as the best for all experiments.

Focusing on the FOPTD models, the analysis of the identification task (see Fig. 8 for all experiments) shows the strong non-linearities of the model (see Table 2) since the variation of the parameters covered for all experiments is large: In fact, the static gain G changes in a ratio of one to six, while the time constant T and the lag L change in a ratio of one to five and one to two, respectively. Moreover the gain and the lag look quite similar for both expanders, while the time constant related to the second expander is higher: the main physical difference between the two expanders is the evaporating pressure that they can impose in the system at the same operating point. This can lead to different behavior of the response because SH is a function of the evaporating pressure. Since we do not want to have a strong investment in nonlinear modeling, these different results imply that a multi-linear model, with multiple model structures, has to be used in the future control design.

Moreover if we consider only the best linear model for each experiment, the result of the identification task is particularly interesting: for each experiment, the best model accuracy is always characterized by an error between 2%

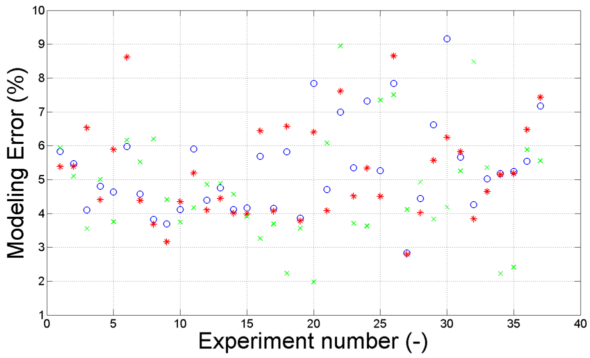

Fig. 7. Modeling error (o: FO model, *: FOPTD model, x: FOFOPTD model)

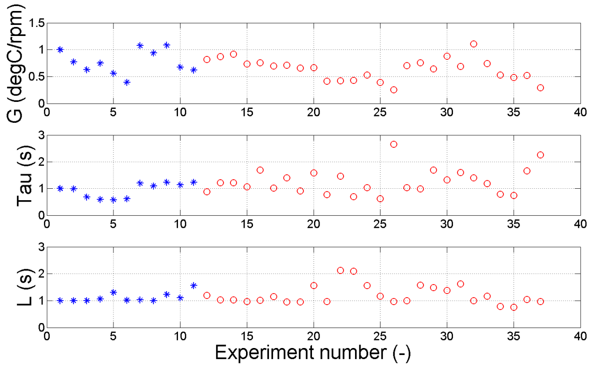

Fig. 8. Normalized parameters for the FOPTD model for all experiments (* is for expender 1, o is for expender 2). The first experiment gives the reference values

and 6% (see Fig. 7). It is then possible to notice (see Table 3) that the mixed model FOFOPTD is more likely to be the best model than the other two classes of linear model: Hence it is outperforming the FOPTD model, which is usually considered for such developments in the literature. Therefore, our proposed methodology leads to an usual conclusion for further linear control design: is is better to consider 3 model structures (FO, FOPTD, FOFOPTD) rather than only the FOPTD structure.

Table 2. Statistic analysis of the normalized FOPTD parameters for the two expanders ($\bar{\cdot}$: mean value, $\cdot_\sigma$: standard deviation)

Experiments	$\bar{G}$	G_σ	$\bar{T}$	T_σ	$\bar{L}$	L_σ
1-11	0.77	0.23	0.94	0.27	1.12	0.36
12-37	0.64	0.20	1.26	0.48	1.21	0.18

Table 3. Recurrence of each model structure as the best among the 37 experimental runs

	FO	FOPTD	FOFOPTD
Recurrence	3	14	20

5. CONCLUSION

This paper provided a methodology for choosing the experiments to be performed in an ORC to get models used for further WHR control design. Future work deals with control design, like multi-model scheduling controllers or model predictive controllers, which have to be adapted to handle more than one model structure.

REFERENCES

Aufderheide, B. and Bequette, B. (2003). Extension of dynamic matrix control to multiple models. *Computers and Chemical Engineering*, 27, 1079–096.

Grelet, V., Dufour, P., Nadri, M., Reiche, T., and Lemort, V. (2015a). Explicit multi model predictive control of a waste heat rankine based system for heavy duty trucks. *IEEE 54th Annual Conference on Decision and Control (CDC)*, 25(3), 179–184.

Grelet, V., Dufour, P., Nadri, M., Reiche, T., and Lemort, V. (2015b). Modeling and control of waste heat recovery rankine based systems for heavy-duty trucks. *IFAC International Symposium on Advanced Control of Chemical Processes (ADCHEM)*, 569–574.

Hernandez, A., Ruiz, F., Desideri, A., Ionescu, C., Quoilin, S., Lemort, V., and Keyser, R.D. (2016). Nonlinear identification and control of organic rankine cycle systems using sparse polynomial models. *IEEE Conference on Control Applications*, 129, 19–22.

Peralez, J., Nadri, M., Dufour, P., Tona, P., and Sciarretta, A. (2017). Organic rankine cycle for vehicles: Control design and experimental results. *IEEE Transaction on Cntrol System Technology*, 952–965.

Rijpkema, J., Andersson, S., and Munch, K. (2018). Thermodynamic cycle and working fluid selection for waste heat recovery in a heavy duty diesel engine. *SAE International in United States*, Technical Paper 2018-01-1371.

Seitz, D., Gehring, O., Bunz, C., Brunschier, M., and Sawodny, O. (2018). Model-based control of exhaust heat recovery in a heavy-duty vehicle. *Control Engineering Practice*, 70, 15–28.

Stijepovic, M., Linke, P., Papadopoulos, A., and Grujic, A. (2012). On the role of working fluid properties in organic rankine cycle performance. *Applied Thermal Engineering*, 36, 406413.

Tona, P. and Peralez, J. (2015). Control of organic rankine cycle systems on board heavy-duty vehicles: a survey. *IFAC-PapersOnLine*, 48(15), 419–426.

Xu, Yebi, A., Onori, S., Liu, Z.F.X., Shutty, J., Anschel, P., and Hoffman, M. (2017). Transient power optimization of an organic rankine cycle waste heat recovery system for heavy duty diesel engine applications. *SAE International Journal of Alternative Powertrains*, 6(1), 25–33.

Zhao, R., Zhang, H., Song, S., Tian, Y., Yang, Y., and Y.Liu (2018). Integrated simulation and control strategy of the diesel engine-organic rankine cycle (orc) combined system. *Energy Conversion and Management*, 156, 639–654.