Selecting representative working cycles from large measurement data sets

Conference Paper · March 2016

DOI: 10.13140/RG.2.1.3905.7049

Reno Filla
Scania

Selecting representative working cycles from large measurement data sets

Ted Samuelsson[1], Reno Filla[1], Bobbie Frank[1,2], Lennart Skogh[1]

[1]Volvo Construction Equipment, Emerging Technologies, 63185 Eskilstuna, Sweden
[2]Lund University, Faculty of Engineering, 221 00 Lund, Sweden

Abstract: A tool has been developed that selects one representative cycle, or a set of cycles, from large measurement data sets based on a specified set of repetitive signals, possibly weighed in their importance. Three different computation methods have been developed and tested, all preserving physical correctness. These approaches are described in detail and compared in the paper.

1 Introduction

In an area where both energy efficiency and productivity are of high concern it is important to know how vehicles and mobile working machines are used in real world applications in order to carry out design optimisations. Reference cycles are commonly used to assess the impact of design changes on the performance of components, subsystems or complete vehicles and machines by means of computation. This is convenient and practical, though caution is advised. Out of necessity, any such reference cycle is a simplified representation of the real-life use of a system or component that is assumed to be most common. Choosing a reference cycle, representing the behaviour of a system, is a way of modelling that particular system [1,2]. From this follows that reference cycles are specific for the situation and task or question at hand, just like simulation models are. This means that there can never be one single reference cycle that will give the right answers in any context, just like there will never be one single simulation model that can be used to assess any type of system property: complete system performance, energy efficiency, component durability, noise and vibration, etc. all require specialised models. Models are determined by the context and the questions to be answered − and so are reference cycles.

Synthetic i.e. artificially constructed cycles are not a good choice even when based on real measurement results such as in [3], since averaging, interpolating or time and value scaling easily impair physical correctness. For example, scaling a measured cycle in the time dimension (i.e. making it slower or faster in order to meet a specific target duration) breaks physicality of dependent signals that are based on a time derivative or time integral of another signal – if not individual amplitude scaling is performed in addition in order to adjust for the errors introduced. However, scaling the amplitude of a signal (i.e. adjusting all values by a common factor) breaks physicality when this signal is a factor or ratio of other signals. This means that in principal forces, powers, energies, consumptions, efficiencies, speeds, accelerations etc. must not be scaled, neither in time nor amplitude (value). That simple averaging of cycle values will not work is self-evident and requires no further discussion.

Elaborate ways of constructing cycles for testing of fuel consumption and exhaust emissions have been proposed [4,5] though the focus is more on representing a group of systems in average condition, like a fleet of cars in normal driving, rather than describing specific systems in specific conditions.

When, instead, using measured working cycles a conflict between validity and practicality arises: a measurement campaign often results in a large amount of cycles. For working machines and heavy vehicles these cycles are often repetitive and collectively represent the machine usage in the application under consideration. Due to practical reasons such as constraints on computational resources one often wants to capture the essence of the specific application with only one single cycle rather than a set of many. The challenge is to find the most representative working cycle within a given measured data set. The often employed quick and dirty way of selecting such cycles is by visual comparison and cannot be recommended.
Another variant of this scenario can be the task to find a reference cycle that based on a given set of measurements represents the most intense 10% operation, for example in order to account for the most demanding customers. Again, performing this task visually will probably not lead to an acceptable result in an acceptable time.

2 Application background

Our interest in computing reference cycles stems from working with the development of construction machines and their major subsystems like propulsion and hydraulics. In this paper a wheel loader in short loading cycles is used as an example.

Sometimes also dubbed V- or Y-cycle for its characteristic driving pattern (Fig. 1), a wheel loader's short loading cycle typically involves bucket loading of material like gravel, sand, or wood chips on an adjacent receiver within a rather tight time frame of 25-35 seconds, in extreme situations even faster than that [6].

Such a short loading cycle represents several interesting challenges for the design engineer: not only is the momentary power distribution between parallel subsystems hydraulics and propulsion to be balanced in order to minimise losses, but also a balance between the subsystems' capacities in terms of forces and speeds needs to be achieved. In [7] a typical traditional design process is described, consisting of static calculations of isolated subsystems and iteration of the results in calculation loops in order to find a good balance. In the then envisioned enhanced design process dynamically augmented (yet still static) calculations play a role as a bridge to dynamic simulation of the complete machine [7].

In each design phase the work of the wheel loader to be developed needs to be described. Rather than employing a set of static design points (i.e. neglecting the dynamics of machine operation and pretending there is only a limited number of steady-state operation points) engineers use reference cycles to preliminarily dimension components and subsystems. For highest relevance (and therefore best quality of the result) these reference cycles are distilled from real measurement campaigns and represent the typical duty and work task of the machine or system in question.

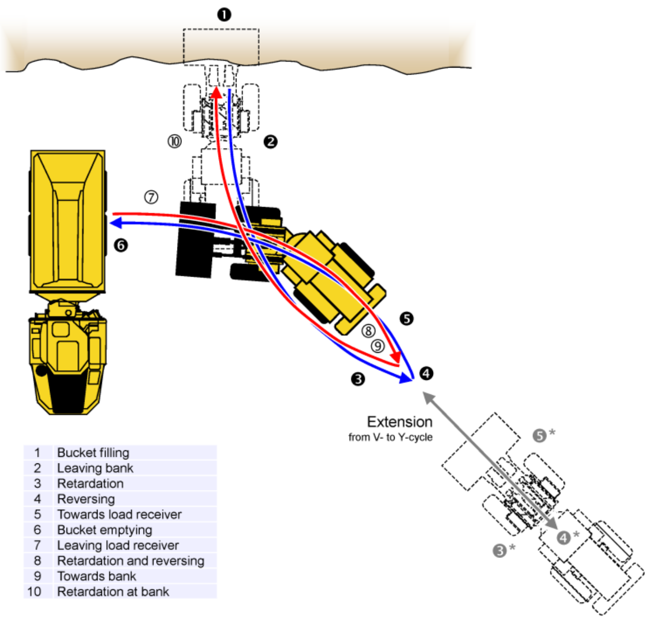

Fig. 1 Short loading cycle: phases and extension from V- to Y-cycle [6].

Another aspect is that today machine manufacturers work closely with suppliers of components and systems during the product's whole life cycle. For example, a second or third party might take on responsibility for conceptual and detailed design of the engine, complete driveline or hydraulics. Often, for competitive reasons an OEM might not want to give these parties too deep an insight into the detailed design rules for the complete machine. To minimise the confidential data transferred the system in question could be isolated and the inputs and outputs could be described by means of reference cycles. For example, for dimensioning the steering system of a wheel loader the required system output can be described as typical profiles of steering cylinder load and speed in a short loading cycle (Fig. 1).

3 Methods

The previous section establishes that engineers frequently encounter the problem of having to extract one or a limited number of representative cycles from a large dataset of measurements. This task can be broken down into two subtasks:

1. Identifying individual cycles in continuous measurements
2. Determining one or a set of reference cycles based on specified signals

3.1 Identifying individual cycles in continuous measurements

The most common way to gather measurement data is likely to operate a machine (in our case: wheel loader in short loading cycles) continuously for a certain time period, which results in a large data set where the individual cycles need to be identified and separated, preferably automatically.

A characteristic and repeating pattern needs to be found based on a single or several signals that allow identifying cycle start and finish, possibly even start and finish of specific phases within each cycle, such as the bucket filling phase (phase 1 in Fig. 1). Work with respect to this has been published in [8-10]. In Fig. 2 the lift angle of the wheel loader's loading unit and the current transmission gear is used to identify individual cycles.

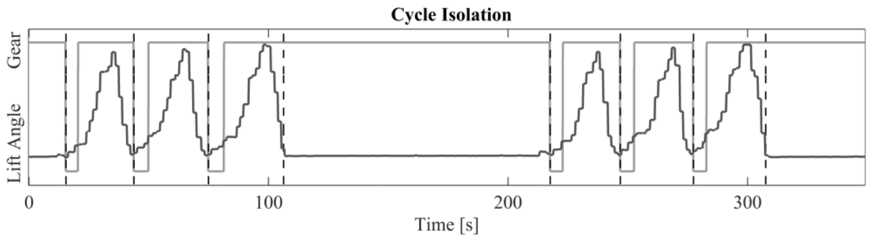

Fig. 2 Identification of individual cycles using active gear and lift angle.

Identification of specific phases within each cycle has been proven useful to improve the performance of the various methods for finding representative working cycles. In this paper we chose to divide the short loading cycle into four phases, corresponding to the four legs of the cycle (Fig. 1).

3.2 Determining one or a set of reference cycles based on specified signals

Three different methods have been developed to determine one or a set of reference cycles out of a group of cycles, based on specified signals, each utilizing a different algorithm to select a real, i.e. measured cycle out of the collective, thus preserving physical correctness:

1. Mean Cycle Method: constructs an intermediate artificial cycle by the use of means and then selects the measured cycle closest
2. Synthetic Mean Method: constructs an intermediate artificial cycle representing the peak frequency and amplitude of the data, then selects the measured cycle closest
3. Least Error Method: compares all cycles with each other and then chooses the measured cycle with the smallest deviation to the collective with respect to amplitude and length

All three identification methods (each proposed and advocated for by a different co-author of this paper, though improved collaboratively) have in common that the final selection is based on a calculated error value. To improve the selection process several signals within the measurement data set can be specified as significant (if desired even weighted), for example to find the cycle that is most representative in terms of engine power and traction force applied to the ground. For scalars that only give a single value per cycle (like cycle time, productivity, fuel consumption and efficiency) the individual cycle's value is instead compared with the median value for the set and the error is calculated from the difference.

3.3 Mean Cycle Method

The Mean Cycle Method is based on the work in [3]. The improvement in our version is that the constructed artificial mean cycle is only an intermediate step, used to select one of the measured cycles out of the data set. Therefore no problems with broken physicality exist.

The principle approach is to first time-scale all cycles to the same length (the median duration) and then to create an "average" cycle by taking the arithmetic mean at each time sample for all the cycles. As pointed out previously, there is a high likelihood that the physicality in this "average" cycle is corrupt, but it is only used as an intermediate step. To reduce excessive smoothing for cycles of significant different durations the cycles are divided into phases (the four legs of the short loading cycle) and time-scaling is done per phase.

The method includes the following steps:

1. For each phase:
    a. Calculate median phase length. *The median is used to avoid high impact from outliers.*
    b. Time-scale corresponding phase in all cycles to median length.
    c. Calculate mean signal. *Calculating the mean amplitude value in each time step, using all cycles.*
    d. Calculate error of amplitude and time for each cycle. *The amplitude error is the accumulated error in each time step, while the time error is the phase's time deviation from the median duration.*
    e. Remove outliers. *Removing cycles with a total error larger than the specified quantile.*
    f. Recalculate mean cycle without outliers.
    g. Recalculate errors to the new mean cycle.
2. Calculate each cycle's total error. *The sum of errors from each phase.*
3. Select most representative cycle. *Select the cycle with the smallest total amplitude and time error.*

This method is fairly straight forward and computationally inexpensive. Fig. 3 shows the result of the method performed on a collective of 100 working cycles. The selected cycle resembles the artificial cycle rather well.

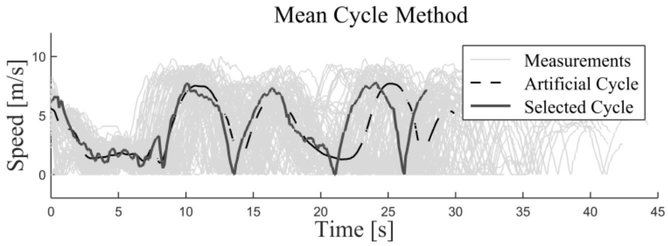

Fig. 3 Mean Cycle Method applied (specified signal is vehicle speed).

3.4 Synthetic Mean Method

The Synthetic Mean Method tries to improve over the Mean Cycle Method by constructing an artificial cycle without calculating arithmetic mean values – in order to avoid the smoothing effect present otherwise, even when considering cycle phases. The ambition has been to resemble the measured cycles in a better way by first calculating all cycle peaks' amplitude and position and then averaging individually. This method's steps are:

1. Find all significant peaks for each cycle. *Significant peaks are defined as the largest peak inside a specified time interval.*
2. Calculate most frequent number of peaks found.
3. Collect all cycles with number of peaks corresponding to 2.
4. For each cycle: between each peak find the valley's position and amplitude. *If more than one valley can be found, use the valley with the lowest value.*
5. Check if cycle boundaries are peaks or valleys. *Check derivative at start and end of cycle.*
6. Calculate the amplitude in the middle between peaks and valleys.
7. Calculate the mean position and amplitude for each point (peak, valley and midpoint).
8. Interpolate between the points. *Create the artificial cycle by the use of cubic interpolation between the points.*
9. Calculate all cycles' error from the artificial cycle. *Combine the errors for amplitude, time and number of peaks. Calculation of amplitude error by time-scaling the synthetic cycle to the duration of the current measured cycle.*
10. Select most representative cycle. *Select cycle with the lowest total error.*

Since this method only uses signals matching the most common number of peaks the amount of data used for constructing the artificial signal will be lower than for the Mean Cycle Method. The actual number of measured cycles contributing to the synthetic mean is dependent on the specified signals' properties. A large range in number of signal peaks identified leads to fewer cycles being used in constructing the synthetic mean. Reducing the aggressiveness of the peak finding algorithm or dividing the cycles into phases can help ensuring the synthetic mean's representativeness.

Obviously, this algorithm cannot handle monotonous signals like accumulated fuel consumption. Figure 4 shows the result of application to the test collective. The selected cycle resembles the artificial cycle rather well, the latter being less smoothed than what is the case when using the Mean Cycle Method.

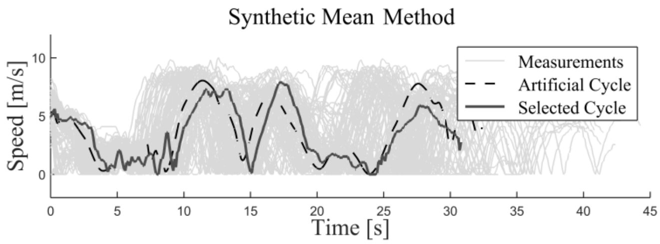

Fig. 4 Synthetic Mean Method applied (specified signal is vehicle speed).

3.5 Least Error Method

The Least Error Method differs from the other methods in that no artificial cycle is constructed as reference. Instead, each cycle is compared to all others and the cycle with the smallest amplitude and time error is selected as the most representative.

This method's steps are:

1. For each cycle:
    a. Scale all other cycles to the current cycle's length.
    b. Calculate the amplitude error between the current cycle and all the others, individually.
    c. Calculate the cycle's length error with respect to the median cycle time.
2. Remove outliers. *Cycles with total error larger than specified quantile.*
3. Repeat step 1.
4. Select most representative cycle. *Select cycle with the smallest total error.*

The method is computationally more expensive than the others, since the number of calculations is quadratic proportional to the number of cycles – compared to the linear relationship of the Mean Cycle Method and the Synthetic Mean Method. However, computation of the Least Error Method can easily be parallelised, decreasing the required calculation time.

The selected cycle from this method can be viewed in Fig. 5. Dividing the individual cycles into phases will further increase the quality of the result, however also further increase computation time.

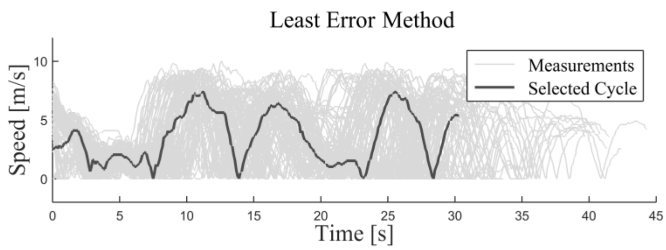

Fig. 5 Least Error Method applied (specified signal is vehicle speed).

4 Results and discussion

All three methods work well in the test suite employed by us, consisting of 100 working cycles generated by different machine operators in a study reported in [11].

When using the methods with only one signal specified as significant, the output varies. Fig. 6 shows the cycles identified as representative top 5 when vehicle speed is specified as the only significant signal.

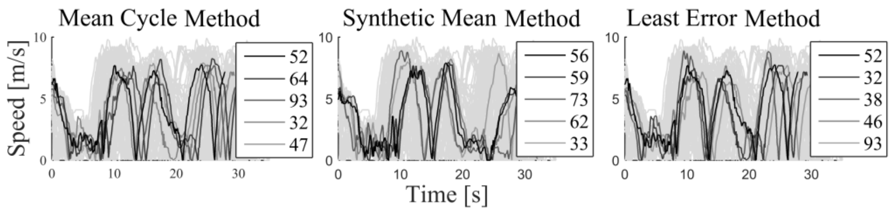

Fig. 6 Vehicle speed profiles of the top 5 representative cycles per method when only vehicle speed is specified as significant. Legend shows the cycle index# (first choice on top).

The deviation in the selected cycles between the three methods is less pronounced when using a signal with fewer peaks, like lift angle (see Fig. 2), since the artificial signal constructed by the Mean Cycle Method and Synthetic Mean Method will become more similar.

Adding additional significant signals, thus including them in the process to determine the representative cycle, leads to a more similar output from the methods. This can be seen in Fig. 7, where in addition to vehicle speed also vehicle direction, current transmission gear, lift angle and tilt angle were selected as signals of equal significance. Furthermore, the scalars cycle duration, bucket load and fuel consumed were specified as significant signals. Those signals were used to calculate fuel efficiency (in t/L, i.e. material loaded per unit fuel) and productivity (in t/h, i.e. material loaded per time unit).

When comparing the vehicle speed plots in Fig. 6 and Fig. 7 it is apparent that the selection of the representative cycle(s) is less impacted by a single signal specified as significant when additional, equally weighted, significant signals are specified.

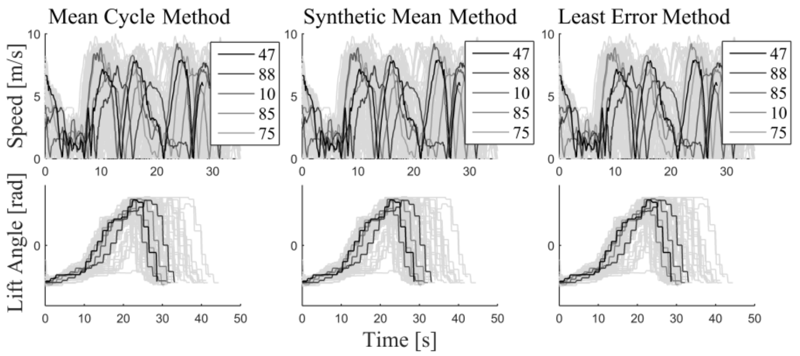

Fig. 7 Vehicle speed and lift angle profiles of the top 5 representative cycles per method when ten signals are specified as significant (see text).

All three methods selected cycle #47 (visualised in Fig. 8) as most representative.

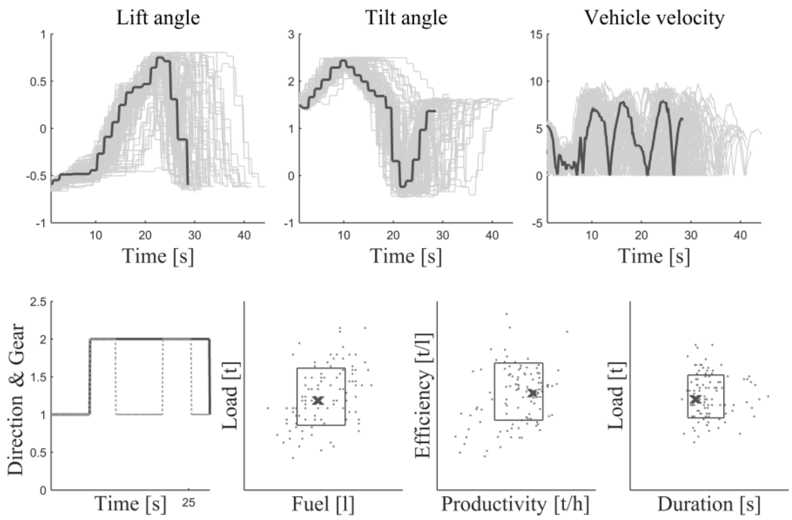

Fig. 8 Cycle #47 as determined as most representative by all three methods. Rectangle in (x,y)-plots represents one standard deviation from the mean.

Which method to choose depends on the time budget and the intended application of the representative working cycle. The Least Error Method is, if not parallelised, significant slower than the other methods when dealing with large measurement sets, while the Synthetic Mean Method may consume some time for initial tuning of the peak finding algorithm. The latter might be preferred for analysis applications were it is important that the selected cycle representative in terms of peakiness, for example in simulations giving input to durability analysis.

All methods employ time scaling during the selection process in order to be able to compare cycles and cycle phases. As has been discussed in the beginning of this paper, all scaling is highly likely to break physicality and should therefore be used as sparingly as possible. Table 1 shows how much time scaling the different methods employed. However, it needs to be pointed out again that the result, i.e. the cycle determined to be most representative is not impacted at all, since any scaling is only be performed as an intermediate step. Also, each method penalises cycle length error thus intrinsically strives to minimise scaling.

Table 1. Time scaling of top 10 selected cycles and whole set (100 cycles) for each method.

	Mean		Median		SD	
Method	Top 10	All	Top 10	All	Top 10	All
Mean Cycle	12%	19%	11%	16%	5%	14%
Synthetic Mean	8%	13%	4%	10%	4%	10%
Least Error	6%	7%	7%	5%	8%	13%

Note that for the Least Error Method the scaling shown is the average scaling needed for each cycle. In contrast to the Mean Cycle Method and Synthetic Mean Method where time scaling is only performed once per cycle under consideration (when compared to the artificial mean) when using the Least Error Method each cycle is scaled multiple times (once for each comparison with a peer) .

Fig. 9 shows a comparison of the artificial signal that the Mean Cycle Method and the Synthetic Mean Method construct for the test data set of 100 working cycles when vehicle speed is specified as the significant signal. As mentioned previously, the shape of artificial signal constructed by the Synthetic Mean Method is less smooth compared to the Mean Cycle Method.

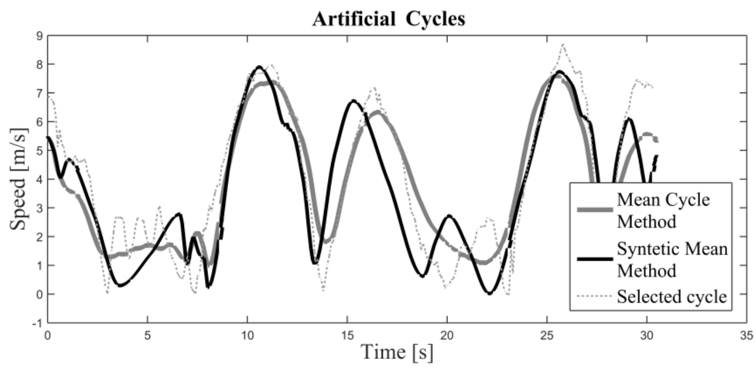

Fig. 9 Comparison of constructed artificial signals (specified signal is vehicle speed).

Until now it has not been explicitly mentioned that one interesting aspect of the tool developed is that it can be used to select a cycle with a specified target. For example, if a representative cycle with respect to engine power is selected, there would be about 50% of the cycles with lower and 50% with higher engine power generated. By changing the quantile value (default is 0.5 for the median) the tool can select cycles representing another duty, for example 90% engine power.

When combining this with the use of multiple signals the tool can be used to select, for example a cycle with high productivity, low fuel consumption and overall normal operator inputs on accelerator pedal and hydraulic levers (Fig. 10). Compared to the selection shown in Fig. 8 the values for productivity and efficiency are significantly higher.

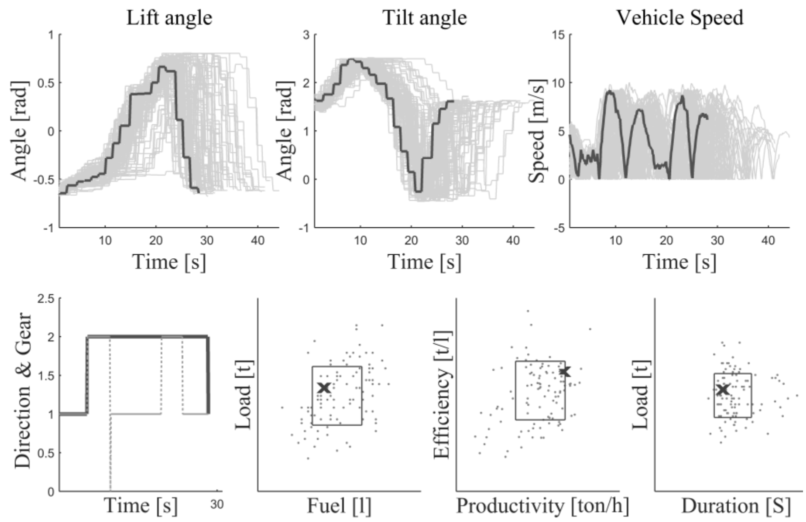

Fig. 10 Cycle representing high productivity, low fuel consumption and normal operator input. Rectangle in (x,y)-plots represents one standard deviation from the mean.

5 Conclusion

Having to select a representative cycle out of a large data set of measurements is a task that engineers are faced with on a frequent basis. It is possible to accomplish this without having to resort to time-wise expensive and error-prone visual comparisons or simple and flawed calculations of arithmetic means.

In this paper we presented three different methods for automatically selecting a single or a set of representative cycles out of a collective – with the advantage of the result not being impaired by broken physicality. The output can be used in various applications, such as identification of how (and how effective) an operator uses a machine (possibly in comparison to his/her peers), to determine/verify specific use cases, or as representative input to further analysis in a product development process.

6 References

1. Filla R, Ericsson A, and Palmberg J-O: Dynamic Simulation of Construction Machinery: Towards an Operator Model. Proceedings of the International Fluid Power Exhibition IFPE 2005 Technical Conference, Las Vegas (NV), USA, pp 429-438, March 16-18, 2005.
2. Filla R: Representative Testing of Emissions and Fuel Consumption of Working Machines in Reality and Simulation. SAE Technical Paper 2012-01-1946, 2012.
3. Patil O, Jagoda A, and Yuan QH: Real World Drive Cycle Construction for Off-Highway Vehicles. Proceedings of the International Fluid Power Exhibition IFPE 2014 Technical Conference, Las Vegas (NV), USA, March 4-8, 2014.
4. Dai Z, Niemeier D, and Eisinger D: Driving Cycles: A New Cycle-Building Method That Better Represents Real-World Emissions. University of California, UC Davis-Caltrans Air Quality Project, Task Order No. 66, USA, September 2008.
5. Dukulis I and Pirs V: Development of driving cycles for dynamometer control software corresponding to peculiarities of Latvia. Proceedings of the International Scientific Conference on Research for Rural Development, vol. 15, pp 95-102, April 2009.
6. Filla R: Simulating Operability of Wheel Loaders: Operator Models and Quantification of Control Effort. Proceedings from the 2nd Commercial Vehicle Technology Symposium CVT 2012, Kaiserslautern, Germany, March 14-15, 2012.
7. Filla R and Palmberg J-O: Using Dynamic Simulation in the Development of Construction Machinery. Proceedings of The Eighth Scandinavian International Conference on Fluid Power SICFP 2003, Tampere, Finland, vol. 1, pp 651-667, May 7-9, 2003.
8. Filla R: Study of a Method for Assessing Operability of Working Machines in Physical and Virtual Testing. International Journal of Vehicle Systems Modelling and Testing, vol. 7, no. 3, pp. 209-234, 2012.
9. Ohlsson-Öhman K: Identifying operator usage of wheel loaders utilizing pattern recognition techniques. Master thesis, ISRN LiTH-ISY-EX--12/4591--SE, Division of Automatic Control, Department of Electrical Engineering, Linköping University, Linköping, Sweden, 2012.
10. Nilsson T, Nyberg P, Sundström C, et al.: Robust Driving Pattern Detection and Identification with a Wheel Loader Application. International Journal of Vehicle Systems Modelling and Testing, vol. 9, no. 1, pp. 56-76, 2014.
11. Frank B, Skogh L, and Alaküla M: On wheel loader fuel efficiency difference due to operator behaviour distribution. Proceedings from the 2nd Commercial Vehicle Technology Symposium CVT 2012, Kaiserslautern, Germany, March 14-15, 2012.