**Methods**

Improving the performance of a ship requires a detailed analysis of the energy demands to optimize the dimensioning and operation of the technologies and processes to install on-board. It requires the analysis of the hourly, daily and seasonal variations of the heat/power consumption, operating (e.g. ship speed) and external conditions (e.g. air temperature). However, gathering and processing those data may be time-consuming because of the high number of points to assess. It is therefore desirable to reduce this large quantity of information while keeping the relevant details on the relations between the variables of interest.

One way is to represent these yearly profiles in a limited number of representative (typical) periods, avoiding the repetition of similar data sets. Clustering consists of grouping such sets in a single group (cluster) so that the items in the same group are more like each other than to those belonging to other groups. Several algorithms are introduced in the literature to partition these datasets. The one selected in this work is the Lloyd’s (k-means) algorithm, which is a non-deterministic method computationally faster than conventional hierarchical tools:

1. N*k* points are chosen (randomly or not) as cluster centers;
2. each observation is assigned to the nearest cluster center (set partitioning);
3. the centroids (means) of each cluster are assigned as new cluster centers;
4. steps (2) and (3) are repeated until the convergence criterion (e.g. minimal decrease in squared error, maximum number of evaluations…) is satisfied.

The k-means algorithm does not necessarily result into an optimum clustering, as it depends on the cluster initialisation (starting point *v*) and number (N*k*). The approach applied by Fazlollahi et al. was used in the present work to select the optimum number of clusters (N*k*). The number of clusters (N*k*) should be as low as possible for data handling purposes, while preserving a high accuracy of the retrieved data. It builds on the calculation of three criteria:

1. the average intra-cluster distance to assess the density of each cluster, preferably small;
2. the average inter-cluster distance to assess the distance between each, preferably high;
3. the expected square error, which is a statistical measure suggested by Pham et al.. It evaluates the ratio of the observed to expected squared errors for the clusters - a low value indicates an improvement from defining *k-1* to *k* clusters.

In parallel, the cluster quality was systematically evaluated by calculating the following performance indicators:

1. the profile deviation (deviation between original and typical period profiles), σprofile;
2. the deviation from the load duration curve of the average values of each period, σCDC;
3. the relative error in load duration curve deviation ELDC;
4. the maximum duration load curve difference ΔMLDC;
5. the number of periods with relative errors Δprod higher than 7%.

This set of performance objectives is defined as a set of constraints with an upper limit that should be respected when minimising the number of clusters, applying the e-constraints algorithm. The dataset is further reduced by partitioning each typical day into a set of segments using a similar approach.

The aim of the present clustering is to identify typical periods that are appropriate for improving the thermodynamic performance of ship energy systems. It can be achieved through either reducing the external energy demands (better energy management), e.g. with storage systems or internal recovery, or through implementing new technologies that result in higher fuel-to-demand efficiencies (better energy conversion). Hence, the attributes that are the primary focus of this study are:

1. the total power consumption, including the mechanical and electrical loads;
2. the total heating demand, further divided into the low- and high-temperature needs, which is related to the external temperature;
3. the total exergy destruction on-board, which quantifies the thermodynamic performance of the ship energy systems.

A one-year data with 35,040 time steps (sampling of 15 minutes) was selected and divided into typical periods based on these three attributes.

The clustering approach proposed by Fazlollahi et al. [CIT] provides as output typical days having the same data frequency as the original time series (in this case, 15 minutes). From a computational perspective, however, this might not necessarily be the most efficient sampling. Operating conditions within a certain cluster (typical day) are often nearly constant for a longer time, suggesting that a variable time resolution would lead to a lower dimensionality of the data while retaining most of the original information.

To this purpose, in this paper we employ the Adaptive Piecewise Constant Approximation (APCA) approach proposed by XXX et al.

**Results**

A data clustering was applied at first on the *total power and heating demands* along the operation year, using 1000 starting points generated randomly for cluster initialisation. As presented in Section [**number**], the number of typical days depends on the values of the intra- and inter-cluster distances and on the expected squared error, obeying five constraints on the load duration profiles. A higher number of typical days would result in smaller data losses, but the calculation of the ESE indicator shows that this gain of information is negligible.

The average power demand is about 4700 kW with a standard deviation of 3000 kW and a 95% percentile of 8300 kW. These figures denote a high variability, but the number of optimum clusters is only 2, which shows that the power demand follows repeatable trends over time. The first typical day corresponds to more than 32,000 hours (low to high demands, up to 12,000 kW), and the second one to ~3000 hours (peak demands, above 12,000 kW). On the contrary, the total heating demand presents variations of smaller amplitude, but a higher number of clusters is necessary. Five typical days appear sufficient to represent the load duration curve and energy demand profiles. The cases with the greatest power consumption and different trends are already considered in these five clusters, and the addition of an extreme period is not necessary.

The large number of typical days (5) compared to those required (2) when clustering only the power demand (2) illustrates the lack of direct correlation between the power and heating requirements.

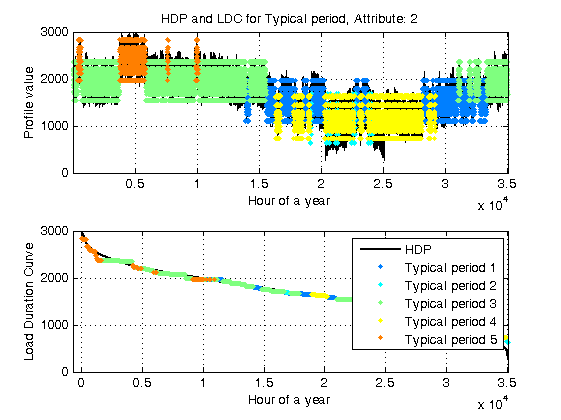
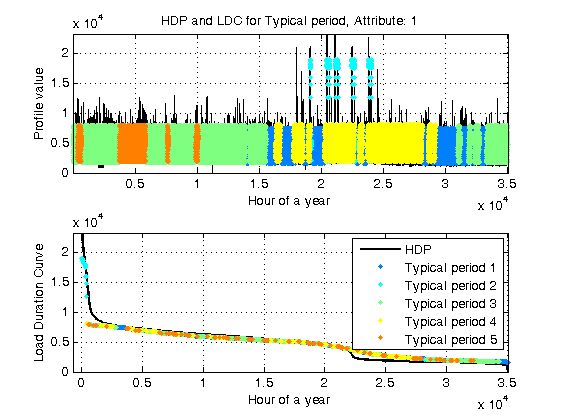


Figure [number]: Power (left) and heat (right) profiles and duration curves, represented with the 35,040 datapoints (black curve) and five typical days (other colours).

The quality of the clustering was assessed by calculating the performance indicators for both attributes (Table [number]). The suggested typical periods present low relative errors and are slightly better for characterisation of the heat demand profiles.

Table []: Performance parameters for the 5-cluster power and heat

|  |  |  |
| --- | --- | --- |
|  | Power consumption | Heat demand |
| σprofile | 0.65 | 0.10 |
| σCDC | 0.11 | 0.11 |
| ELDC | 0.23 | 0.07 |
| ΔMLDC | 0.18 | 0.06 |
| Δprod,0.07 | 155 | 168 |

The five typical days are further segmented and the closeness of the plots for both demands illustrates the quality of the segmentation. In most typical days, the power consumption varies in a 2000-8000 kW range, with an average consumption of about 4500-5000 kW and similar trends. The maximum values, of about 19,000 kW, are reached at highest ship speeds.

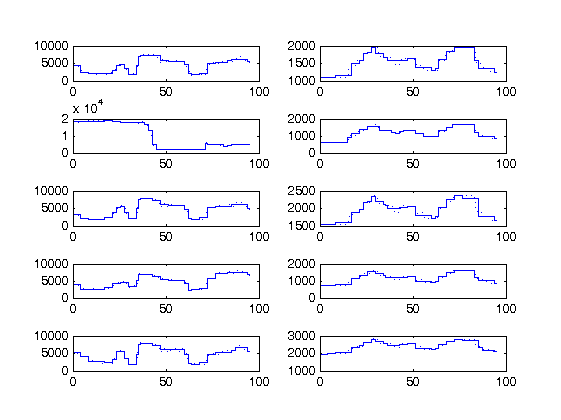


Figure [number]: Power (left) and heat (right) profiles (dotted line) and segmentations (thin line) for the 5 typical days and duration curves.

The same findings can be deduced from the segmentation of the heat profiles as all present similar tendencies. The only significant difference is the range between the minimum and maximum values reached in a single day (e.g. 2000-2800 kW, 1000-2000 kW, 700-1700 kW, etc.). These differences are correlated to changes of the outer temperature and are thus seasonal variations. Heat storage from low-demand to high-demand days is not feasible, but may be implemented, if responsive enough, on single days.

A data clustering was applied on the *total power demand and exergy destruction*, following the same approach as above. The optimum number of clusters is only 2, which shows the direct relation between the power consumption and exergy destruction, and, consequently, the weak relation with the heating needs. This result is as expected, since the engines and subsequent components are responsible for the largest irreversibilities in the system. High accuracy of the clustering for even such a small number of typical days is reached, as shown, for instance, with the relative load differences (ΔMLDC) are 0.18 for both power and exergy. A further segmentation of these two typical days shows that these two variables follow the same trend in both regular and extreme conditions. The data clustering shows that, for ship energy systems, it is critical to reduce the power demand or to improve the power generation system for enhancing the system performance. Such findings are valid for all types of operation days.

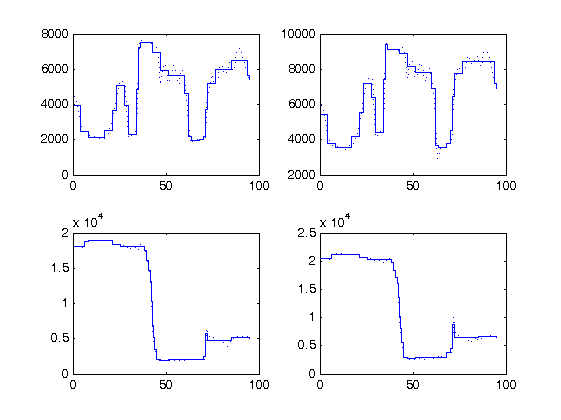
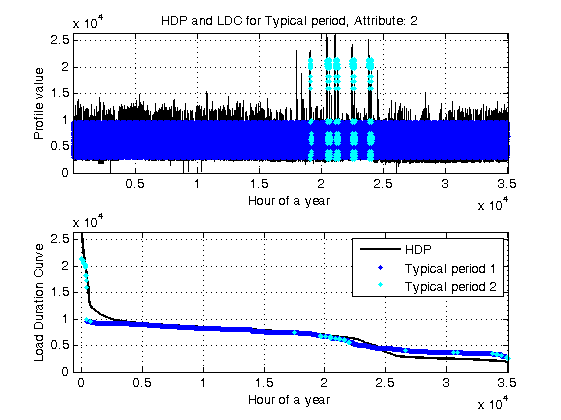


Figure [number]: Exergy destruction profile and duration curve (left), represented with the 35,040 datapoints (black curve) and two typical days (dark and clear blue). Segmented power and exergy destruction profiles for the two typical days (right).