A Novel Approach to Energy Management in Large Passenger and Cruise Ships: Integrating Simulation and Machine Learning Models

Eirini Barri[1][a], Christos Bouras[1][b], Apostolos Gkamas[2][c], Nikos Karacapilidis[3][d], Dimitris Karadimas[4][e], Georgios Kournetas[3][f] and Yiannis Panaretou[4][g]

[1]Department of Computer Engineering and Informatics, University of Patras, 26504 Rio Patras, Greece
[2]University Ecclesiastical Academy of Vella, Ioannina, Greece
[3]Industrial Management and Information Systems Lab, MEAD, University of Patras, 26504 Rio Patras, Greece
[4]OptionsNet S.A, Patras, Greece
ebarri@ceid.upatras.gr, bouras@cti.gr, gkamas@aeavellas.gr, karacap@upatras.gr, karadimas@optionsnet.gr, kgiorgos837@gmail.com, panaretou@optionsnet.gr

Abstract. It has been broadly admitted that the prediction of energy consumption in large passenger and cruise ships is a complex and challenging issue. Aiming to address it, this chapter reports on the development of a novel approach that builds on a sophisticated agent-based simulation model, which takes into account diverse parameters such as the size, type and behavior of the different categories of passengers onboard, the energy consuming facilities and devices of a ship, spatial data concerning the layout of a ship's decks, and alternative ship operation modes. According to the proposed approach, outputs obtained from multiple simulation runs are then exploited by prominent Machine Learning algorithms to extract meaningful patterns between the composition of passengers and the corresponding energy demands in a ship. In this way, our approach is able to predict alternative energy consumption scenarios and trigger meaningful insights concerning the overall energy management in a ship. Overall, the proposed approach may handle the underlying uncertainty by blending the process-centric character of a simulation model and the data-centric character of Machine Learning algorithms. The chapter also describes the overall architecture of the proposed solution, which is based on the microservices approach.

[a] https://orcid.org/0000-0001-8733-983X
[b] https://orcid.org/ 0000-0001-9160-2274
[c] https://orcid.org/ 0000-0003-0966-5140
[d] https://orcid.org/0000-0002-6581-6831
[e] https://orcid.org/ 0000-0002-8816-862X
[f] https://orcid.org/0000-0001-8668-296X
[g] https://orcid.org/0000-0003-0609-1604

Keywords: Agent-based simulation, Energy Consumption, Large Passenger and Cruise Ships, Machine Learning Algorithms, Microservices.

1 Introduction

Undoubtfully, energy saving in large ships has many benefits, both for the environmental protection and the reduction of a ship's operating costs. In this direction, the International Maritime Organization aims to reduce ship emissions by at least 50% by 2050, while ships to be built by 2025 are expected to be a massive 30% more energy efficient than those built some years ago [11]. A particular ship category is that of large passenger and cruise ships, which reportedly consume a large amount of energy and thus constitute an interesting area for investigating diverse energy consumption and energy saving solutions. Interestingly enough, while such solutions have been thoroughly investigated in the case of buildings, very limited research has been conducted so far for the abovementioned ship category.

Aiming to contribute to this research gap, this chapter reports on the development of a novel approach that builds on a sophisticated agent-based simulation model. The model takes into account the size, characteristics (e.g. age, special needs etc.) and behavior of the different categories of passengers onboard, as well as the energy consuming facilities and devices of a ship. In addition, the simulation model exploits spatial data corresponding to a detailed layout of the decks of a specific ship, thus offering customized visualizations. Finally, the model caters for alternative ship operation modes, corresponding to cases where the ship cruises during the day or night, or is anchored at a port. The proposed agent-based simulation model has been implemented with the use of the AnyLogic simulation software (https://www.anylogic.com/), which provides a nice graphical interface for modeling complex environments and allows the extension of its simulation models through Java code.

A novelty of our approach concerns the exploitation of the outputs obtained from multiple simulation runs by prominent Machine Learning (ML) algorithms to extract meaningful patterns between the composition of passengers and the corresponding energy demands in a ship. In this way, our approach is able to predict alternative energy consumption scenarios and trigger insights concerning the overall energy management in a ship. In addition, it handles the underlying uncertainty and offers highly informative visualizations of the energy consumption.

The work reported in this chapter is carried out in the context of the ECLiPSe project (http://www.eclipse-project.upatras.gr). The project aims at leveraging existing technological solutions to develop an integrated energy consumption and energy saving management system for the needs of large passenger and cruise ships. A major task of the project concerns the development of efficient algorithms for the analysis and synthesis of the associated multifaceted data, which may considerably enhance the quality of the related decision-making issues during the operation of a vessel. These algorithms will trigger recommendations about the management of energy consumption, enabling stakeholders to gain energy saving insights.

The remainder of this chapter is organized as follows: Section 2 reports on related work. Section 3 presents the overall architecture of the proposed solution. Section 4 describes the proposed approach that builds on the strengths of simulation and machine learning. Section 5 presents indicative experiments and corresponding results from the application of the proposed approach, which are then analyzed in Section 6. The statistical validation of the insights produced by the machine leaning models is discussed in Section 7. Finally, concluding remarks and future work directions are outlined in Section 8.

This chapter is an extended version of the work described in [2]; it enhances the information reported in most sections, while it includes two new sections that describe the overall architecture of the proposed solution, and advance the experimental evaluation of the proposed approach by reporting on the statistical validation of the insights produced by the machine leaning models.

2 Related Work

While considerable research has been conducted so far on the optimization of various energy consumption issues in buildings (being they smart or not), very limited work has been reported so far in the case of large ships. For instance, an agent-based model for office energy consumption is described in [17]. This work elaborates the elements that are responsible for energy consumption and presents a mathematical model to explain the energy consumption inside an office. The proposed model is validated through three sets of experiments giving promising results.

Adopting another perspective, a review of Machine Learning (ML) models for energy consumption and performance in buildings is presented in [16]; the motivation of this work was the exploitation of contemporary technologies, including network communication, smart devices and sensors, towards enhancing the accuracy of prediction in the above energy management issues. On a similar research direction, a combination of mathematical statistics and neural network algorithms to solve diverse energy consumption problems is proposed in [10]; this work analyzes the associated big data aiming to facilitate energy consumption predictions for various types of buildings.

A comparative analysis of energy saving solutions in buildings appears in [5]; the proposed tool for assessing the effectiveness of energy saving technologies implementation allows not only to evaluate individual decisions, but also to compare and rank them according to the breakeven rate for the efficiency implementation decline. A combination of Nearest Neighbors and Markov Chain algorithms for the implementation of a system that is able to support decision making about whether to turn on or off a device in a smart home setting, thus handling the related energy management issues, is described in [14].

Research on the energy consumption of ships during four different transatlantic cruises over the period of one month is reported in [13], through the elaboration of 250 samples of ship data concerning ship speed, wind speed, ship draft, latitude and longitude, etc. Data considered also concern devices that produce power, such as the ship's oil and heat recovery boilers. Based on all these data, a huge database containing thousands of files has been built, which in turn feeds a simulation environment that enables a ship operator to estimate the energy consumption of cruise ships.

A new method to model the ship energy flow and thus understand the dynamic energy distribution of the marine energy systems is introduced in [9]; using the Matlab/Simscape environment, a multi-domain simulation method is employed. As reported, the proposed method can help people better monitor the ship energy flow and give valuable insights about how to efficiently operate a vessel. In a similar research line, aiming to provide a better understanding of the use of energy, of the purpose it serves, and of the efficiency of its conversion on board, an analysis of the energy system of a cruise ship operating in the Baltic Sea is provided in [4]; being based on a combination of direct measurements and computational models of the energy system of the ship, the proposed approach ensures to provide a close representation of the real behavior of the system.

Finally, an optimization framework that improves the efficiency of the energy systems employed in ships has been proposed in [1]. The framework is based on genetic algorithms and aims to maximize the energy efficiency and minimize the fuel consumption and the thermal energy dissipation by optimizing the load allocation of the ship energy systems. To this purpose, different strategies for the energy systems on board of an existing cruise ship are proposed and analyzed.

3 The architecture of the overall solution

The overall architecture of the solution developed in the context of the ECLiPSe project is based on the microservices approach, thus enabling the development of independent services as well as their easy scaling and upgrading (see Figure 1). Microservices provide a way to escalate the development and delivery of large, complex applications, allowing individual components to evolve independently of each other. The microservices architecture offers greater flexibility through service independence, allowing organizations to become more flexible when offering new business opportunities or responding to changing market conditions. Microservices allow the right tool to be used to get the job done, which means applications can be developed and delivered by technology that is better for the project, rather than locked into a single technology, runtime or framework.

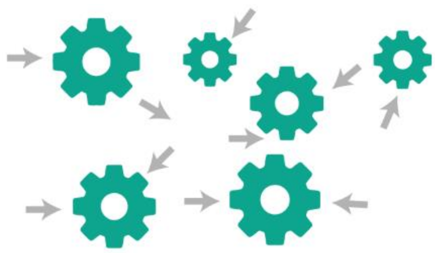

Fig. 1. A schematic description of the microservices concept

Based on the above, the adoption of a microservices-based approach presents significant benefits, in that it allows:

- The integration of models from the AnyLogic simulation software.
- The implementation of modular services and user interfaces.
- The integration of adapters for importing data from third-party systems, such as future energy consumption recording systems or passenger localization systems.
- The integration of adapters for the activation of third-party systems, e.g. future systems with semi-automatic or automatic energy controllers, Internet of Things (IoT)-based smart metering systems, etc.

The individual subsystems can be developed in parallel without affecting each other, while it is particularly important that each service component is able to scale according to the individual requirements, without affecting the operation or performance of the other subsystems. Finally, at the level of system interconnection with existing and/or future third-party systems, the selected architectural approach is able to support the integration of any number and types of adapters that will contribute to the final system. A contribution can be considered either a data stream, e.g. from real sensors, or even a trained algorithm for processing and drawing conclusions on data circulating within the system.

3.1 Example Use Cases

The first use case our system is envisaged to cover concerns data aggregation for passenger localization systems in order to allow future data processing towards correlation of energy prediction with actual passenger locations and spatial distribution. Scenarios for passenger localization as well as for the simulation of passengers movement and spatial distribution on a large passenger ship that have been developed and implemented in the AnyLogic simulation model can be replaced by a service that will connect the proposed system with a passenger positioning system i.e. LYNCEUS2MARKET - An innovative people localization system for safe evacuation of large passenger ships (http://www.lynceus-project.eu/). Without changes to the architecture of the overall solution, this case will be able to display energy forecasts and recipes based on the movements coming from the abovementioned service.

Another use case concerns data feeding of the system from smart energy metering systems. The current trend in the field of information technology and energy is smart metering devices based on the concept of the IoT. This sector is highly developed in both industrial and building facilities; however, there are currently no relevant facilities on passenger ships, while related research shows that this area will be strongly developed in the coming years. The proposed architecture will be able to be enriched with suitable adapters that will undertake both the collection of energy consumption data and the provision of commands for automatic or semi-automatic control of devices, based on the "energy saving recipes" of the system.

3.2 Architecture Details

Figure 2 illustrates the overall solution foreseen in the ECLiPSe project, which allows for interconnections with third party systems and services, e.g. through cloud computing infrastructures for further data processing and/or use of diverse machine learning algorithms.

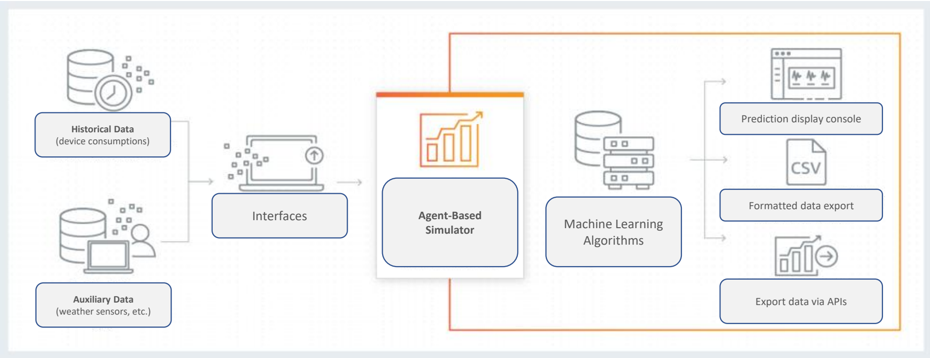

Fig. 2. The overall solution of the ECLiPSe project.

A functional view of the proposed solution is shown in Figure 3; building on the abovementioned microservices-based approach, our solution may be easily extended to accommodate diverse use cases. The major microservices shown in Figure 3 are:

- *Playground*: The service describing the scenario and the specifications of the desired energy analysis and modelling. It identifies all the necessary specifications of the models used by the AnyLogic simulation tool. In a future version, when data from smart energy consumption devices or from the actual passenger position/movement will be available, this service will be configured and adapted accordingly (without affecting the other services).
- *Core*: The basic service of simulating energy consumption scenarios. It handles the required synthesis of data, as described in the previous sections, either for data that are not available, or because we want to evaluate special cases in order to draw meaningful conclusions about energy consumption.
- *Report*: The service that displays diverse results, in the form of raw data reports or in the form of comparison charts.
- *Analysis/Recipes*: The data analysis service that is based on machine learning algorithms and tries to detect trends and behaviors in order to provide insights to ship managers regarding its energy consumption and possible ways to optimize it. The basic principles of this service are analyzed in the next section.
- Potential third party services to allow expansion of the system usability, i.e. passengers location, and energy metering.

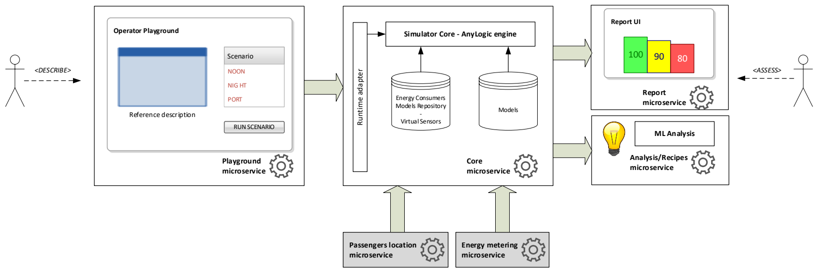

Fig. 3. Functional view of the proposed solution.

Figure 4 shows the proposed system architecture and the interoperability of its individual components, which are:

- *AnyLogic Engine:* the AnyLogic simulation software properly adapted for the needs of the proposed solution.
- *Predictive Models Evaluator:* the evaluator of the energy consumption models predicted by the proposed solution.
- *High Level Architecture (HLA) Support:* the component that supports and adapts system data to and from the AnyLogic simulation tool.
- *Runtime Adapter:* the communication adapter of the individual microservices, i.e. the scenario description service and the energy consumption scenario simulation service.
- *Integration Middleware:* it includes the data analysis service.
- *End-User UI and External/Internal APIs (application programming interfaces):* it includes the configuration and scenario description services, as well as the results display service. Functionalities for data exporting are also included in this subsystem.

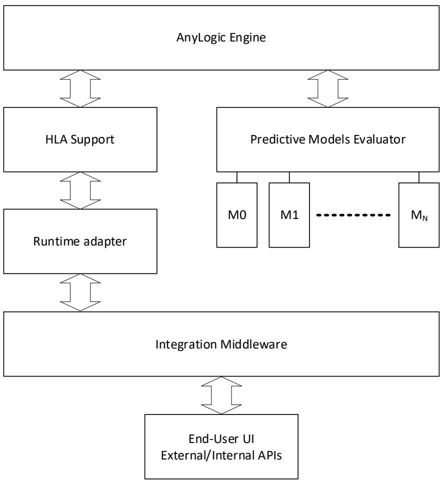

Fig. 4. The proposed architecture enables a simulator-as-a-service environment.

The above architecture has been developed and deployed by utilizing modern software virtualization technologies, namely the Docker platform. Unlike virtual machines that require the installation of an entire operating system regardless of their workload, the Docker platform was adopted as the most mature containerization technology for the realization of the foreseen solution. Figure 5 illustrates an extended component break-down of the overall solution, which includes existing components as well as future extensions and connectors to third-party systems and APIs.

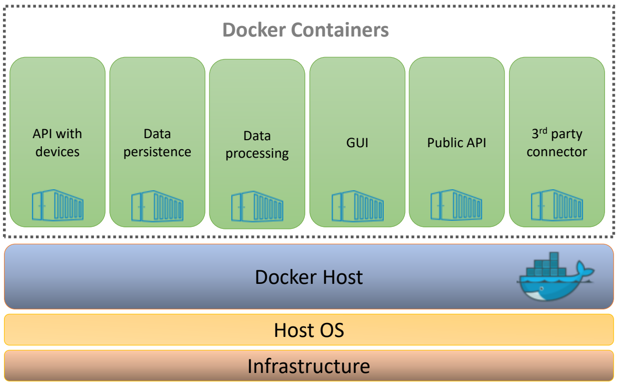

Fig. 5. Deployment stack and containerized orchestration of the architecture.

3.3 Integration Aspects

The integration between the Runtime Adapter, UIs, APIs and other components is managed by a Queue-based component based on RabbitMQ (https://www.rabbitmq.com/) Message broker and utilizing the Advanced Message Queuing Protocol (AMQP). AMQP is an ISO standard based on publish-subscribe messaging. It runs over the TCP/IP network protocol and was designed to interconnect remote devices with a small code footprint on low permeability networks. These AMQP specifications make it ideal for interconnecting low-power network devices and for this reason AMQP has significant penetration of modern object-based web applications. The structure of an AMQP messaging network consists of devices / clients that have the ability to publish and / or receive messages and the server that hosts the broker. The broker is essentially the core element of AMQP, as it is the entity that is responsible, among other things, for identifying the connection of the devices / clients, and at the same time managing all the exchanged messages and routing them to the clients who request to receive messages.

The overall solution hosts a broker entity with the aim of simultaneously connecting to other system components, i.e. the Runtime Adapter for communicating with the agent-based simulator and monitoring/control devices of a ship installation in a future version of the system. These devices can be both stand-alone sensor / actuator systems and gateways that aim to connect peripherals of heterogeneous network technologies (zigbee, z-wave, Bluetooth, etc.) to the IP network.

4 The Proposed Approach

Our approach adopts the Action Research paradigm [6], which aims to contribute to the practical concerns of people in a problematic situation; it concerns the improvement of practices and strategies in the complex setting under consideration, as well as the acquisition of additional knowledge to improve the way shipping stakeholders address issues and solve problems. Building on the strengths of existing related work, as reported in the previous section, the proposed approach comprises two main phases: (i) agent-based simulation of the energy consumption in various sites of a ship, and (ii) utilization of prominent ML algorithms on the outputs of multiple simulation runs to extract meaningful insights about the relation between the passenger composition and corresponding energy demands. Through these phases, our approach is able to gather, aggregate and analyze heterogeneous data representing both the energy consumption in diverse devices and facilities and the concentration of passengers in different areas of a ship.

To fine tune our approach, a series of meetings with shipping companies were conducted; through them, we identified the types of devices and facilities that mainly affect energy consumption in the ship categories under consideration, and obtained valuable information concerning the parameters to be taken into account in energy consumption models (such as that energy supply in a ship is provided by a number of electric power generators, which are often of different capacity and do not work in parallel; estimations of energy demands according to the number of passengers were also obtained through such meetings). In addition, information collected concerned the layout of ship decks and its relation to the energy management issues investigated. Finally, we clarified issues related to the alternative types of passengers and how these may influence alternative energy consumption and energy saving scenarios [3].

4.1 Agent-based Simulation

Our approach aims to enable stakeholders predict the energy needs of a ship (e.g. to recommend the appropriate number of power generators to operate each time), facilitate predictive maintenance issues (affecting the related equipment), and hopefully reduce the energy related operating costs. To fulfil these aims, our simulation model takes into account the passengers' behavior and its dependencies with a ship's facilities, devices and resources.

A basic assumption of our approach is that the energy demands in many sites of a ship (such as the restaurant, the nightclub, the kindergarten etc.) depend on the number of passengers who gather at these sites at a given time, as well as their composition in terms of type (customer or crew member), age, gender etc. We consider that different age groups have different paths and habits (differences among passenger groups may even affect the speed of a moving agent). To estimate the populations gathered in these sites, we relied on the behavioral preferences that large subgroups of passengers have. For instance, we assume that young passengers prefer to spend their time at nightclub from 10pm to 3am, while elderly passengers prefer to eat dinner at a fancy restaurant. Our model may also simulate the behavior of persons with special needs (PWSN); in particular, we assume that these people move at a slower pace and are in most cases accompanied by another person. Such assumptions enable us to predict the gathered populations and, accordingly, the energy demands during day and night. This approach facilitates the modeling of energy consumption, especially for ships that do not have sophisticated energy consumption monitoring and control systems.

In addition, according to our approach, the passengers' behavior is being considered and modelled through three basic scenarios corresponding to the ship (i) being moved during the day, (ii) being moved during the night, and (iii) being anchored at a destination or port. In the above scenarios, we assume different behaviors from passengers, which may result to different energy demands. Finally, to accommodate the spatial particularities of each ship, our approach pays much attention to the layout of each deck. These layouts provide us with the spatial data that are needed to calculate the movement of passengers inside the ship. AnyLogic offers a user-friendly import of sectional plans (views), thus enabling the production of a more realistic model of the distribution of ship passengers, facilities and devices. Taking into account what our models predict in terms of energy needs, we suggest different policies of energy management, aiming to reduce energy consumption.

4.2 ML Algorithms

Having thoroughly assessed the palette of broadly used ML algorithms for the needs of our approach, we decided to utilize two classification algorithms, namely the Decision Trees (DT) and the K-Nearest Neighbors (K-NN) algorithms. This is due to the fact that these algorithms provide high interpretability of their results, they have low computational cost, and they fit well to our data structure.

Decision Trees is one of the simplest and widely used classifiers in the field of Data Mining. They constitute a non-parametric supervised learning method, aiming to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. DT demonstrates excellent applicability in datasets with either categorical or continuous variables. In addition, it requires little data preparation and it is able to process large amounts of data [15].

K-NN is a simple supervised ML algorithm that can be used for both classification and regression problems, and has been extensively applied in diverse disciplines, such as Economics and Health [7]. It relies on labeled input data to learn a function that produces an appropriate output when given new unlabeled data. In most cases, K-NN yields competitive results and has significant advantages over other data mining methods. It differs from other classifiers in that it does not build a generic classification model; instead, whenever a new record is being inserted in the system, it tries to find similar records (nearest neighbors) from past data stored in its memory and assigns it the value of the dependent variable that its neighbors have.

5 Experimental evaluation

To demonstrate the applicability and potential of the proposed approach, this section presents a particular set of experiments carried out for a specific vessel. In particular, we elaborate energy demands that are associated with four popular facilities of a ship, namely (i) the night club, (ii) the kindergarten, (iii) the casino, and (iv) the restaurant. For the case under consideration, we consider and import in the simulation software the original deck layouts, where all ship facilities and passenger cabins are mapped. Moreover, we assume a total population of 3100 passengers onboard, belonging to four distinct age groups (i.e. 1-14, 15-34, 35-54, ≥55 years old). Table I summarizes sample data concerning the populations of each age group in the facilities considered. For each individual group of passengers, we create a simple linear behavioral model in which each individual group remains in a specific facility for some time. We do this for every group of passengers and every time period to create a comprehensive routine for all passengers throughout the day. In this way, we are able to simulate diverse scenarios, which may be easily aggregated to create an illustrative energy consumption map for the whole vessel.

Table 1. Distribution of age groups in various ship facilities [2]

Ship's Cite	Age Group	1-14	15-34	35-54	≥55
Nightclub	**Percentage**	0%	60%	30%	10%
	Population	0	300	150	50
Kindergarten	**Percentage**	35%	10%	55%	0%
	Population	53	15	82	0
Restaurant	**Percentage**	12%	8%	35%	45%
	Population	46	30	134	172
Casino	**Percentage**	0%	0%	35%	65%
	Population	0	0	112	208

5.1 Night Club

For the case elaborated in this chapter, we generated random samples of 500 passengers, assuming that the percentage of passengers visiting this facility is between 15% and 17%. This facility operates from 11pm to 5am. The conditional probability of someone visiting the night club is shown in Table 1. We also set the time spent there (from passengers of all age groups) to follow a triangular distribution with a lower limit equal to 50 minutes, mode equal to 95 minutes, and upper limit equal to 110 minutes. Finally, we imported the layout of a specific deck, where detailed spatial data about the cabins and the possible pathways leading to the night club area are described. By running the corresponding simulations, we are able to visualize the possible concentration of passengers during the night at this area of the ship (see Figure 6). Consequently, by estimating the energy requirements of the night club with respect to the number of passengers hosted, we can calculate the possible energy needs for the particular time period and facility (see Figure 7). Such estimations can be used for future predictions of energy consumption in cases where passengers are distributed in a similar way. Furthermore, the derived data can be statistically analyzed to reveal the data patterns and mechanisms that may cause the particular energy demands.

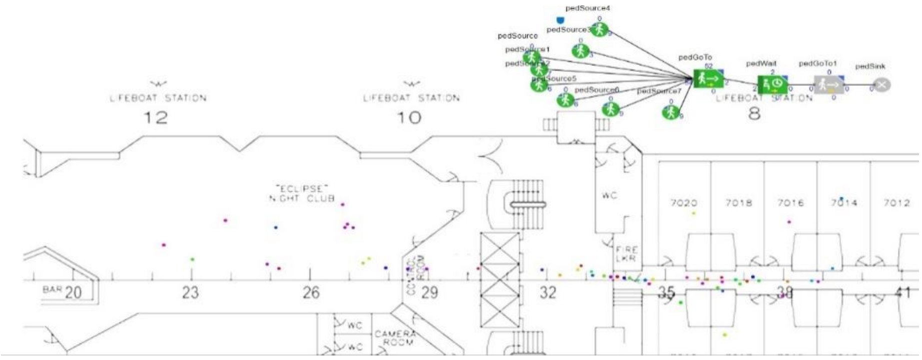

Fig. 6. An instance of a simulated energy consumption scenario in the nightclub [2]

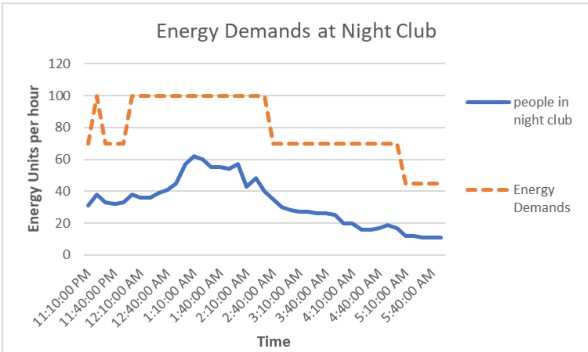

Fig. 7. Energy demands corresponding to passengers' concentration in the nightclub [2]

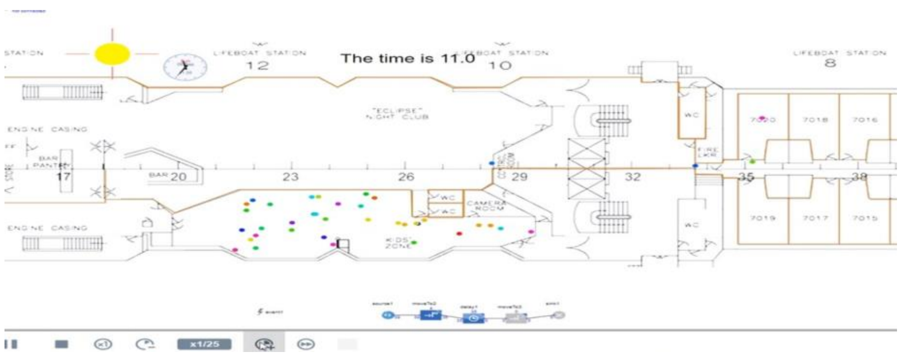

Fig. 8. An instance of a simulated energy consumption scenario in the kindergarten [2]

5.2 Kindergarten

For this facility (see Figure 8), we considered that the passengers who visit it are mainly children (1-14 years old) and their parents (who may belong into the age groups of 15-34 and 35-54 years old). The opening hours of this facility are from 11am to 2pm. We assumed that the kindergarten is not the only choice that the above groups have for entertainment purposes. Also, compared to other areas on the ship, the kindergarten is not large enough to accommodate all parents with their children. We have therefore assumed that the proportion of passengers visiting it daily ranges from 4% to 5.5%, i.e. from 120 up to 176 persons. The time people spend while visiting this facility is described by a triangular distribution with a minimum time of 50 minutes, a maximum time of 110 minutes, and a dominant value of 80 minutes. The experiments carried out gave the concentration of passengers shown in Figure 9.

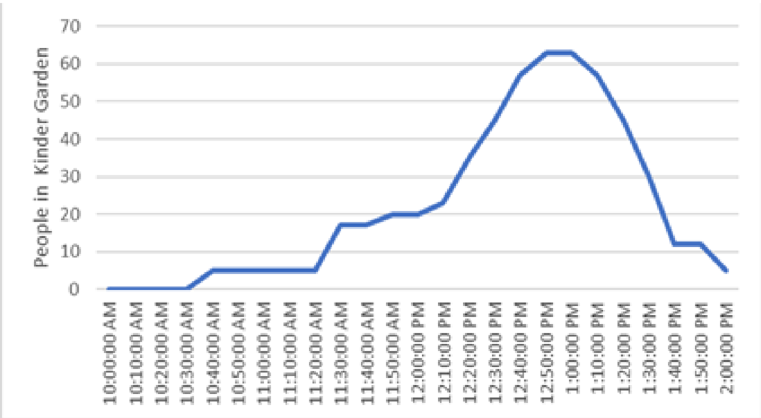

Fig. 9. Passengers' concentration in the kindergarten [2]

5.3 Casino

The samples of passengers used in the particular set of experiments concerned 320 people (i.e. 10% of average passengers' population). We assumed that this facility operates from 7pm to 7am and mainly attracts passengers that are older than 35 years old (65% of them belonging to the ≥55 age group and the remaining 35% to the 35-54 age group). Moreover, passengers that visit the casino are divided into two categories, those who choose to waste their time exclusively in the casino during the night (20%) and those who visit the casino for a certain time period (they may leave and re-enter the casino during the night). The first category concerns the 20% of the casino visitors (their stay follows a triangular distribution with a minimum time of 250 minutes, a maximum of 300 minutes and a dominant value of 270 minutes). Similarly, for the rest 80% of casino visitors we considered that their time spent follows a triangular distribution with a minimum time of 20 minutes, a maximum time of 80 minutes and a dominant value of 35 minutes).

5.4 Restaurant

We considered one of the available ship restaurants (offering an "à la carte" menu, thus not being an economic one), operating from 7pm to 11pm. This facility concerns all passengers, regardless of age group. We assumed that 10%-12% of passengers (320-380 people) choose this particular restaurant; their stay is described by a triangular distribution with a minimum time of 75 minutes, a maximum time of 150 minutes and a dominant value of 120 minutes.

6 Data Analysis and Synthesis

The experiments described above demonstrate diverse features and options offered by the proposed simulation model. To predict energy consumption in large passenger and cruise ships, our approach aggregates results obtained from each particular facility of a ship and produces a corresponding time series diagram, in which the dependent variable is the energy consumption measured in energy units per hour and the time interval is 10 minutes. Figure 10 illustrates the overall energy demands with regards to the estimated gathering of passengers in the facilities discussed in the previous section throughout the day. Obviously, our experiments have not considered the entirety of facilities and energy consumers available on a ship (such as air condition, lighting, heating etc.); however, all of them can be easily aggregated to our model and thus provide a detailed mapping of the overall energy consumption.

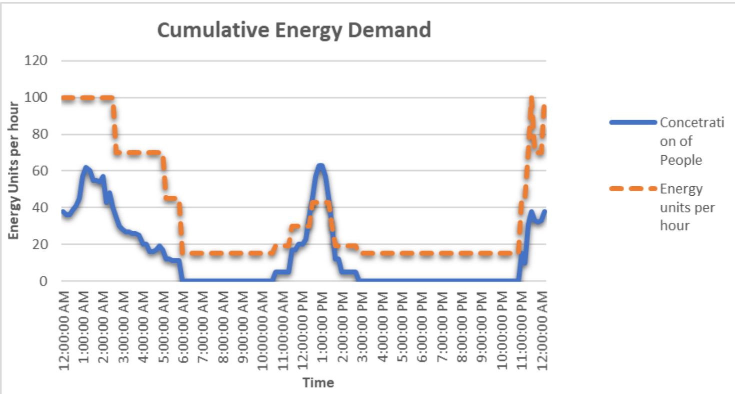

Fig. 10. Cumulative concentration of passengers in four major ship's facilities and corresponding energy demand [2]

Building on the proposed agent-based simulation model that facilitates the creation of alternative energy consumption scenarios, we can produce realistic data that can be further elaborated by prominent machine learning algorithms to provide meaningful insights for managing diverse energy consumption patterns [8]. Parameters taken into account by the proposed machine learning algorithms also include the number of ship generators (categorical variable), the alternative age groups and their populations (as defined for each ship), and the time slots considered each time (the ones adopted in our approach are shown in Table 2).

Table 2. Time slots considered in our approach [2]

Time interval	Time slot
7:00am – 11:59am	Morning
12:00pm – 4:59pm	Midday
5:00pm – 9:59pm	Evening
10:00pm – 6:59am	Night

Table 3. Sample of our dataset [2]

Composition ID	Age Groups				PWSN	Time slot	Number of gen. in simultaneous operation
	1-14	15-34	35-54	≥55			
1	290	535	945	1432	97	Morn.	3
						Mid.	2
						Even.	4
						Night	3
2	200	750	1200	1100	75	Morn.	2
						Mid.	3
						Even.	4
						Night	4
3	175	700	1150	1150	20	Morn.	2
						Mid.	3
						Even.	4
						Night	4
4	48	885	1890	550	100	Morn.	1
						Mid.	3
						Even.	4
						Night	3

In our experiments, we generated a large dataset of 919 different passenger compositions for each time slot. A small sample of this dataset, concerning only four of these compositions for the time slots defined, is presented in Table 3 (the number of generators that operate for each data combination is calculated upon the definition of a set of energy unit intervals and their association with the energy produced by the simultaneous operation of a certain number of generators). A big part of this dataset (70%) was used as the training set of the two ML algorithms incorporated in our approach. Through the utilization of these algorithms, one may predict the required number of generators per time slot for a specific passenger composition.

Focusing on the ‘morning’ time slot, Figure 11 illustrates the output of the Decision Tree algorithm, which classifies alternative passenger compositions into different numbers of power generators required. As it can be observed, the energy consumption of the ship in this time slot is being affected by (i.e. positively correlated to) the ratio of passengers that are older than 55 to those that are younger than 35 years old. The interpretation of this may be that older people use to be more active in the morning (compared to young populations). Results shown in Figure 12 provide additional evidence in favor of the above insight; as depicted, the correlation between the number of generators being used in the morning and the number of elderly passengers is positive.

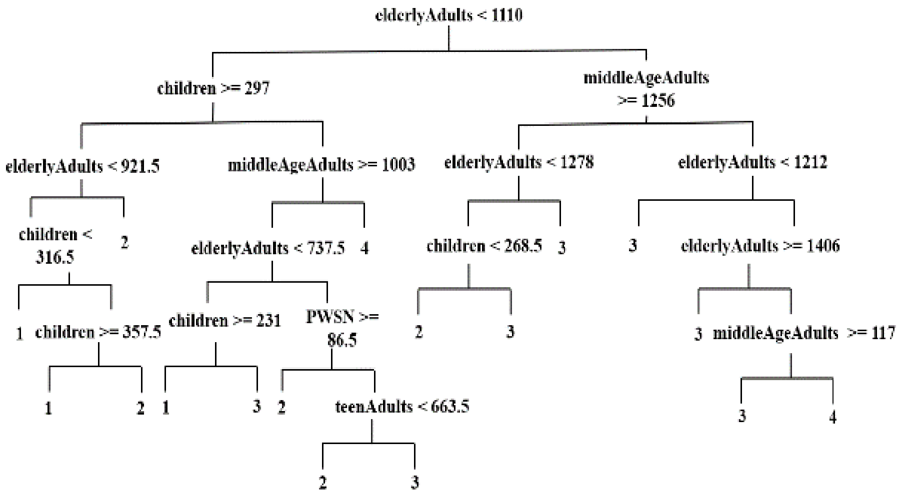

Fig. 11. Decision Tree classification (‘children’, ‘teenAdults’, ‘middleAgeAdults’ and ‘elderlyAdults’ correspond to the 1-14, 15-34, 35-54 and >=55 age groups, respectively) [2]

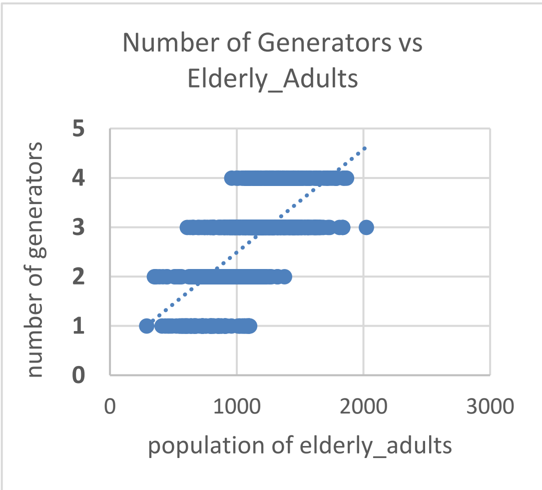

Fig. 12. Scatter plot - number of generators vs population of elderly passengers [2]

For the abovementioned time slot, we also applied the K-NN algorithm. The confusion matrix produced (this matrix is actually a technique for summarizing the performance of a classification algorithm) showed us limited reliability. In particular, K-NN performed very well (with more than 95% accuracy) when classifying compositions of passengers that were associated with the operation of one or four generators, while this was not the case for compositions associated with the operation of two or three generators (in these cases, the accuracy was about 45% and 55%, respectively).

Table 4 summarizes a small set of predictions produced by the K-NN algorithm for the cases of one or four generators operating simultaneously. It is noted that for these cases K-NN produces very similar results to those obtained by the Decision Tree, i.e. the energy needs are positively correlated to the ratio of passengers that are older than 55 to those that are younger than 35 years old. Such insights, resulting from multiple simulation runs, were also validated by shipping stakeholders. According to their validation feedback, adjustments to the initially set parameters and energy demand thresholds were performed.

Table 4. Predictions produced by K-NN algorithm [2]

Age Groups				**PWSN**	**Number of generators in simultaneous operation**
1-14	**15-34**	**35-54**	**≥55**		
100	755	1100	1300	75	4
270	668	916	1570	43	4
174	968	865	1021	40	4
243	755	1412	656	41	1
328	686	1450	678	82	1
410	995	1425	780	10	1

7 Statistical validation of the insights produced by machine learning models

To validate the results derived from the machine learning models used in our approach, we conduct a series of statistical tests to prove and measure the strength of the suggested correlations. Specifically, for the portion of data that concerns the morning time slot and elderly adults, we conduct tests for normality (Kolmogorov-Smirnov and Shapiro-Wilk tests) to investigate whether our data follow a normal distribution.

Table 5. Test for normality; the case of elderly adults' sample

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Elderly Adults	.031	919	.032	.996	919	.023
a. Lilliefors Significance Correction						

Table 5 shows the results from tests of normality. Both tests' P-values (0.032 and 0.023, respectively) reveal that the null hypothesis of being normally distributed is rejected; thus, our data cannot be considered as normally distributed.

In this stage of our analysis, we are able to conduct non-parametric statistical tests to verify that the number of elderly adults in the passenger's composition is related to the number of generators that operate simultaneously, corresponding to the morning time slot. A Kruskal-Wallis H test is conducted with the number of generators acting as the grouping variable and the number of elderly adults as the testing variable.

Table 6. Kruskal-Wallis H test for correlation between elderly adults and number of generators

Test Statistics[a,b]	
	Elderly Adults
Kruskal-Wallis H	452.048
df	3
Asymp. Sig.	.000
a. Kruskal Wallis Test	
b. Grouping Variable: number of generators	

As shown in Table 6, the P-value (Asymp. Sig) of Kruskal-Wallis H test was estimated at 0.000. This result led us to reject the null hypothesis of Kruskal-Wallis H test; thus, we can assume that there is a statistically significant correlation between the number of elderly adults and the number of generators needed in the morning slot. To measure the strength and ascertain the sign of this correlation, we estimate the Kendall's tau b coefficient, which performs well when the correlation under consideration is between a numerical variable (elderly adults) and an ordinal variable (number of generators).

Table 7. Calculation of Kendall's tau b coefficient

Correlations			
			Elderly Adults
Kendall's tau_b	number of generators	Correlation Coefficient	.537**
		Sig. (2-tailed)	.000
		N	919
**. Correlation is significant at the 0.01 level (2-tailed).			

As shown in Table 7, the Kendall's tau b coefficient is estimated at 0.537, which indicates that there is a moderate positive correlation between the number of elderly adults and the number of generators. This result confirms and further validates the outcomes derived from the machine learning models used in our approach and also provides an indication of how strong the discovered correlations are.

8 Concluding remarks and future work directions

The prediction of energy consumption in large passenger and cruise ships is certainly a hard problem. This is mainly due to the need to simultaneously consider the interaction between multiple parameters and agent behaviors. To deal with this problem, the proposed approach blends the process-centric character of a simulation model and the data-centric character of ML algorithms. First, by building on a comprehensive and informative agent-based simulation model, it facilitates the generation and assessment of alternative energy consumption scenarios that incorporate vast amounts of realistic data under various conditions. Second, it advocates the use of prominent machine learning algorithms to aid the finding, understanding and interpretation of patterns that are implicit in this data, ultimately aiming to provide meaningful insights for shaping energy saving solutions in a ship. In addition, the overall architecture of the integrated system is based on the microservices approach. Microservices provide a way to escalate the development and delivery of large and complex applications; they also allow individual components to evolve independently of each other.

In any case, we need to compare the outputs of the proposed approach with real data. As far as the outcomes produced by the agent-based simulation model are consistent with real data, our machine learning algorithms will be better trained, which in turn will enhance the accuracy of the associated energy consumption predictions. Such reinforcement learning activities consist one of our future work directions.

Another research direction concerns the investigation of alternative modes to combine simulation and machine learning in our approach. Specifically, we plan to consider the application of ML algorithms prior to and within the simulation. In the former case, we will need real data to develop rules and heuristics that our agent-based simulation model can then employ. In the latter, we may reuse previously trained ML-based models or train the ML models as the simulation is taking place. In addition, we plan to investigate additional ML algorithms.

Finally, we plan to expand the proposed agent-based simulation model with problem-specific algorithms and interfaces, aiming to enable shipping stakeholders perform a progressive synthesis and multiple criteria comparative evaluation of alternative energy consumption configurations (a similar approach has been proposed in [12]).

9 Acknowledgements

The work presented in this chapter has been co-financed by the European Union and Greek national funds through the Regional Operational Program “Western Greece 2014-2020”, under the Call “Regional research and innovation strategies for smart specialization (RIS3) in Energy Applications” (project: 5038607 entitled “ECLiPSe:

Energy Saving through Smart Devices Control in Large Passenger and Cruise Ships”).

References

1. Ancona, M. A., Baldi, F., Bianchi, M., Branchini, L., Melino, F., Peretto, A., & Rosati, J., Efficiency improvement on a cruise ship: Load allocation optimization. Energy Conversion and Management, 164, 42-58 (2018).
2. Barri, E., Bouras, C., Gkamas, A., Karacapilidis, N., Karadimas, D., Kournetas, G., & Panaretou, Y. Blending simulation and Machine Learning models to advance energy management in large ships. In Proc. of the 10th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2020), 2020, pp. 101-109 (2020).
3. Barri, E., Bouras, C., Gkamas, A., Karacapilidis, N., Karadimas, D., Kournetas, G., & Panaretou, Y. Towards an informative simulation-based application for energy saving in large passenger and cruise ships. In Proc. of the 6th IEEE International Energy Conference (ENERGYCON 2020), Gammarth, Tunisia (2020).
4. Baldi, F., Ahlgren, F., Nguyen, T.-V., Thern, M., & Andersson, K.. Energy and Exergy Analysis of a Cruise Ship. Energies, 11(10), 2508. doi: 10.3390/en11102508 (2018)
5. Chebotarova, Y., Perekrest, A., & Ogar, V. Comparative Analysis of Efficiency Energy Saving Solutions Implemented in the Buildings. In Proc. of the IEEE International Conference on Modern Electrical and Energy Systems (MEES 2019), pp. 434-437. doi: 10.1109/mees.2019.8896691 (2019).
6. Checkland, P.B. & Holwell, S. Action Research: Its Nature and Validity. Systemic Practice and Action Research, 11(1), pp. 9–21 (1998).
7. Cunningham, P., & Delany, S. J. k-Nearest neighbour classifiers. Multiple Classifier Systems, 34(8), 1-17 (2007).
8. Deist, T., Patti, A., Wang, Z., Krane, D., Sorenson, T. & Craft, D. Simulation assisted machine learning. Bioinformatics. 35. 10.1093/bioinformatics/btz199 (2018).
9. Guangrong, Z., Kinnunen, A., Tervo, K., Elg, M., Tammi, K. & Kovanen, P. Modeling ship energy flow with multi-domain simulation. In Proc. of the 27th CIMAC World Congress on Combustion Engines, Shanghai, China (2013).
10. Guzhov, S., & Krolin, A. Use of big data technologies for the implementation of energy-saving measures and renewable energy sources in buildings. In Proc. of the Renewable Energies, Power Systems & Green Inclusive Economy Conference (REPS-GIE 2018), pp. 1-5. doi: 10.1109/repsgie.2018.8488861 (2018).
11. International Maritime Organization. Current awareness bulletin, No. 5. Available at: http://www.imo.org/en/KnowledgeCentre/CurrentAwarenessBulletin/Documents/CAB%20258%20MAY%202018.pdf (2018).
12. Karacapilidis, N. & Moraitis, P. Building an agent-mediated electronic commerce system with decision analysis features. Decision Support Systems, 32(1), 53-69 (2001).
13. Marty, P., Corrignan, P., Gondet, A., Chenouard, R. & Hetet, J-F. Modelling of energy flows and fuel consumption on board ships: application to a large modern cruise vessel and comparison with sea monitoring data. In Proc. of the 11th International Marine Design Conference (IMDC 2012), June 2012, Glasgow (2012).
14. Rajasekaran, R. G., Manikandaraj, S., & Kamaleshwar, R. Implementation of Machine Learning Algorithm for predicting user behavior and smart energy management. In Proc.

of the International Conference on Data Management, Analytics and Innovation (ICDMAI 2017), pp. 24-30. doi: 10.1109/icdmai.2017.8073480 (2017).

15. Rokach, L. & Maimon, O. Data mining with decision trees. Theory and applications. World Scientific (2008).
16. Seyedzadeh, S., Pour Rahimian, F., Glesk, I., & Roper, M. Machine learning for estimation of building energy consumption and performance: a review. Visualization in Engineering 6, p. 5. doi: 10.1186/s40327-018-0064-7 (2018).
17. Zhang, T., Siebers, P. & Aickelin, U. Modelling Office Energy Consumption: An Agent Based Approach. SSRN Electronic Journal. Available at: http://dx.doi.org/10.2139/ssrn.2829316 (2010).