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 is generally confessed that the energy consumption in large passenger and cruise ships cannot be predicted and that is a complex and also difficult issue. Intending to resolve it, this chapter reports on an unique technique that improves an innovative agent-based simulation model, which considers varied specifications such as the size, kind as well as behavior of the different categories of passengers onboard, the energy consuming centers as well as devices of a ship, spatial data concerning the layout of a ship's decks, and also alternate ship procedures. Based on the suggested approach, results acquired from multiple simulation runs are then used up by appropriate Machine Learning algorithms to extract purposeful patterns in between the structure of passengers as well as the corresponding energy needs in a ship. By doing this, our approach is able to predict different energy consumption situations as well as activate significant insights concerning the total power management in a ship. On the whole, the proposed approach may handle the hidden unpredictability by blending the process-centric character of a simulation model and the data-centric character of Machine Learning algorithms. The chapter also describes the general architecture of the suggested remedy, which is based upon the microservices strategy.

[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

Unquestionably, energy saving in big ships has lots of benefits, both for the environmental protection and the reduction of a ship's operating expense. In this direction, the International Maritime Organization intends to reduce ship emissions by at least 50% by 2050, while ships themselves be constructed by 2025 are anticipated to be a massive 30% more energy efficient than those built some years ago [11]. A specific ship classification is that of big passenger and cruise liner, which apparently consume a big quantity of energy and thus make up a fascinating area for examining varied energy consumption and energy saving solutions. Surprisingly enough, while such solutions have been thoroughly examined when it comes to buildings, very limited research has been performed up until now for the abovementioned ship classification.

This chapter reports on advancement of a novel approach that builds on a sophisticated agent-based simulation model. The findings of this study contribute to this research gap. The model takes into consideration the size, attributes (e.g. age, special needs etc.) and habits of the different categories of passengers onboard, in addition to the energy consuming facilities and of a ship. Moreover, the simulation model taken advantage of the spatial information corresponding to a detailed layout of the decks of a specific ship, thus using 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 following proposed agent-based simulation model has been carried out using the AnyLogic simulation software (https://www.anylogic.com/). The application provides the users a great graphical interface in order to model complex environments and it also allows the extension of its simulation models through Java code.

A novelty of our system approach concerns the use of the results gotten from several simulation runs by prominent Machine Learning (ML) algorithms to draw out meaningful patterns in between the composition of passengers and the corresponding energy demands in a ship. Because of this, our proposed system approach can forecast alternative energy consumption scenarios and trigger insights concerning the overall energy management in a ship. In addition, it deals with the hidden unpredictability 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 focuses on leveraging existing technological services to establish an incorporated energy consumption and energy saving management system for the requirements of large passenger and cruise liner. A significant challenge of the project is about the advancement of efficient algorithms for the analysis and synthesis of the associated multifaceted information, which may significantly enhance the quality of the associated decision-making problems during the operation of a vessel. These algorithms will trigger suggestions about the management of energy usage, allowing stakeholders to gain energy conserving insights.

In the remainder of this chapter : Section 2 reports on related work. Section 3 presents the overall architecture of the proposed solution. Section 4 describes the suggested approach that is based on the strengths of simulation and machine learning. Section 5 presents significative 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 substantial research has been carried out so far on the optimization of various energy consumption issues in buildings (being they smart or not), extremely minimal work has actually been reported up until now when it comes to large ships. For example, an agent-based model for office energy consumption is explained in [17]. This work elaborates the aspects that are accountable for energy consumption and presents a mathematical model to discuss the energy consumption inside a workplace. The proposed model is verified through 3 sets of experiments giving promising results.

Adopting another point of view, an evaluation of Machine Learning (ML) models for energy consumption and performance in buildings is presented in [16]; the inspiration of this work was the exploitation of modern technologies, consisting of network interaction, smart devices and sensors, towards boosting the precision of prediction in the above energy management problems. On a comparable research direction, a combination of mathematical stats and neural network algorithms to fix diverse energy consumption issues is proposed in [10]; this work examines the associated big data intending to assist in energy consumption predictions for different kinds of buildings.

A comparative analysis of energy conserving services in buildings appears in [5]; the proposed tool for examining the efficiency of energy conserving technologies implementation allows not just to examine individual decisions, but also to compare and rank them according to the breakeven rate for the efficiency application decrease. A mix of Nearest Neighbors and Markov Chain algorithms for the implementation of a system that has the ability to support decision making about whether to turn on or off a device in a smart house setting, therefore handling the related energy management problems, is described in [14].

Research on the energy consumption of ships throughout four different transatlantic cruises over the duration of one month is reported in [13], through the elaboration of 250 samples of ship information concerning ship speed, wind speed, ship draft, latitude and longitude, etc. Data considered likewise concern devices that produce power, such as the ship's oil and heat recovery boilers. Based on all these data, a big database containing a big number of files has been developed, which in turn feeds a simulation environment that makes it possible for a ship operator to approximate the energy consumption of cruise liner.

A new method to design the ship energy circulation and therefore comprehend the dynamic energy circulation of the marine energy systems is presented in [9]; utilizing the Matlab/Simscape environment, a multi-domain simulation approach is employed. As reported, the proposed method can help people better keep track of the ship energy flow and offer important insights about how to efficiently run a vessel. In a comparable research line, aiming to offer 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 offered in [4]; being based on a combination of direct measurements and computational models of the energy system of the ship, the proposed approach makes sure to offer a close representation of the real habits 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.

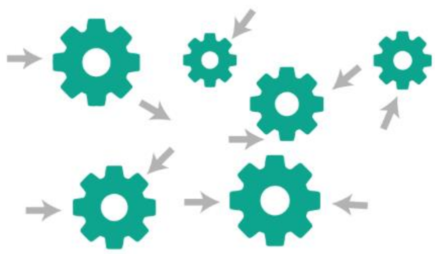

Fig. 1. A schematic description of the microservices concept

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.

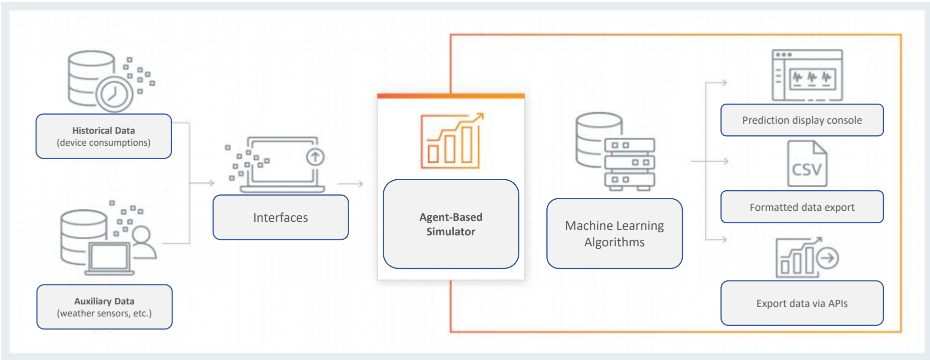

Fig. 2. The overall solution of the ECLiPSe project.

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.

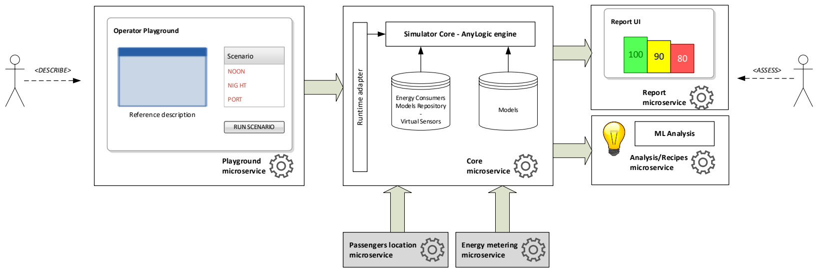

Fig. 3. Functional view of the proposed solution.

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.

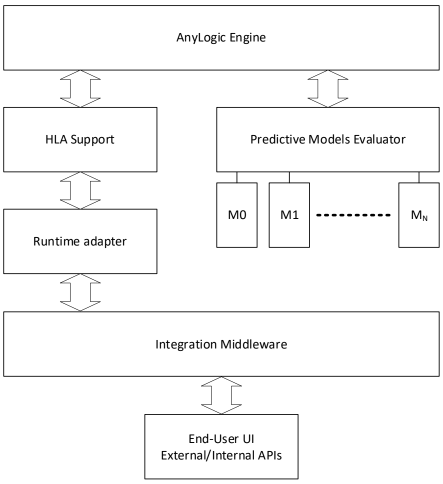

Fig. 4. The proposed architecture enables a simulator-as-a-service environment.

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.

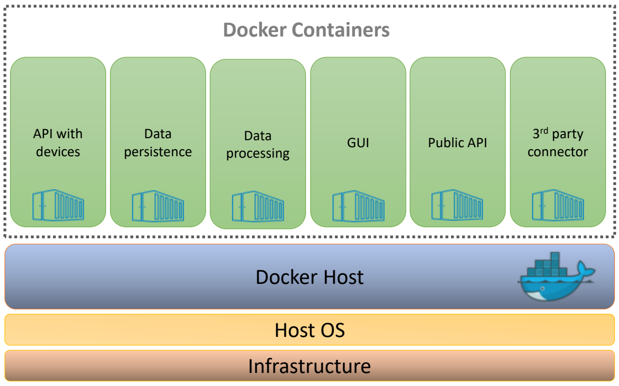

Fig. 5. Deployment stack and containerized orchestration of the architecture.

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 method adopts the Action Research paradigm [6], which aims support people in a troublesome circumstance; it concerns the enhancement of practices and techniques in the complex setting under consideration, as well as the acquisition of additional understanding to enhance the method shipping stakeholders address issues and fix problems. Based on the strengths of already existing related work, as it is reported in the previous section, our proposed approach comprises two main phases: (i) agent-based simulation of the energy usage in numerous sites of a ship, and (ii) usage of popular ML algorithms on the outputs of several simulation runs to draw out significant insights about the relation between the corresponding energy needs and the passenger composition. By means of these phases, our approach is able to collect, aggregate and examine heterogeneous information representing both the energy consumption in varied devices and facilities and the concentration of passengers in various areas of a ship.

To tweak our approach, a series of meetings with shipping companies were conducted; through them, we identified the kinds of devices and facilities that primarily impact energy consumption in the ship classifications under consideration, and gotten valuable info concerning the criteria to be taken into account in energy consumption models (such as that energy supply in a ship is offered by a variety of electrical power generators, which are frequently of different capacity and cannot work in parallel; estimates of energy needs according to the number of passengers were also obtained through similar meetings). In addition, details gathered concerned the layout of ship decks and its relation to the energy management problems investigated. Lastly, we clarified problems associated with the alternative kinds of passengers and how these might influence alternative energy usage and energy saving scenarios [3].

4.1 Agent-based Simulation

Our method intends to allow stakeholders forecast the energy needs of a ship (e.g. to recommend the suitable number of power generators to run each time), assist in predictive maintenance concerns (impacting the related devices), and hopefully decrease the energy related operating costs. To fulfil these goals, our simulation model considers the passengers' habits and its dependences with a ship's facilities, devices and resources.

A fundamental assumption of our approach is that the energy demands depend on the number of the passengers who gather in numerous sites of a ship (such as the restaurant, the nightclub, the kindergarten etc.) at a given time, in addition to their structure in terms of type (customer or crew member), age, gender etc. We consider that various age groups have different paths and habits (differences among passenger groups might even impact the speed of a moving agent). To approximate the populations gathered in these sites, we count on the behavioral preferences that big subgroups of passengers have. For example, we presume that young passengers prefer to spend their time at nightclub from 10pm to 3am, while elderly passengers choose to eat dinner at a fancy restaurant. Our model might likewise simulate the behavior of persons with special needs (PWSN); in particular, we assume that these people move at a lower rate and remain in most cases accompanied by another person. Such assumptions allow us to predict the collected populations and, appropriately, the energy needs during day and night. This method facilitates the modeling of energy consumption, especially for ships that do not have sophisticated energy consumption tracking and control systems.

Moreover, according to our method, the passengers' habits is being thought about and modelled through three fundamental scenarios representing the ship (i) being moved throughout the day, (ii) being moved throughout the night, and (iii) being anchored at a destination or port. In the above circumstances, we presume different behaviors from passengers, that may result to various energy needs. Finally, to accommodate the spatial particularities of each ship, our method pays much attention to the layout of each deck. These layouts provide us with the spatial information that are required to compute the motion of passengers inside the ship. AnyLogic offers an easy to use import of sectional strategies (views), therefore enabling the production of a more realistic model of the distribution of ship passengers, facilities and equipment. The result of our system is to suggest different policies, intending to decrease energy consumption, taking into account what our models predict in terms of energy needs.

4.2 ML Algorithms

Having completely assessed the palette of broadly used ML algorithms for the needs of our method, we chose to make use of two classification algorithms, particularly the Decision Trees (DT) and the K-Nearest Neighbors (K-NN) algorithms. This is due to the truth that these algorithms provide high interpretability of their outcomes, they have low computational cost, and they fit well to our data structure.

Among the easiest and extensively utilized classifiers in the field of Data Mining are the Decision Trees. They constitute a non-parametric monitored learning approach, aiming to develop a model that predicts the value of a target variable by finding out basic choice guidelines inferred from the data features. DT shows exceptional applicability in datasets with either categorical or continuous variables. In addition, it needs little data preparation and it is able to process large amounts of data [15].

K-NN is a basic supervised ML algorithm that can be used for both classification and regression problems, and has actually been extensively used in diverse disciplines, such as Economics and Health [7]. It depends on identified input data to find out a function that produces a suitable output when provided new unlabeled data. In many cases, K-NN yields competitive results and has substantial benefits over other data mining techniques. It differs from other classifiers in that it does not build a generic classification model; rather, whenever a new record is being inserted in the system, it tries to find similar records (nearest neighbors) from previous data saved in its memory and assigns it the value of the dependent variable that its neighbors have.

5 Experimental evaluation

To show the applicability and potential of the proposed method, this section presents a particular set of experiments carried out for a specific vessel. In particular, we elaborate energy needs that are connected with four popular facilities of a ship, particularly (i) the night club, (ii) the kindergarten, (iii) the casino, and (iv) the restaurant. For the case under consideration, we think about and import in the simulation software the initial deck layouts, where all ship facilities and passenger cabins are mapped. Furthermore, it is assumed that a total population of 3100 passengers onboard, belonging to four distinct age groups (i.e. 1-14, 15-34, 35-54, ≥ 55 years of ages). Table I summarizes sample data concerning the populations of each age group in the facilities thought about. For each individual group of passengers, we develop an easy linear behavioral model in which each individual group remains in a specific facility for some time. We do this for every single group of passengers and every time period to develop a thorough routine for all passengers throughout the day. In this way, we have the ability to simulate varied scenarios, which might be easily aggregated to develop an illustrative energy consumption map for the entire 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 created random samples of 500 passengers, presuming that the percentage of passengers visiting this facility is between 15% and 17%. This facility operates from 11pm to 5am. The conditional possibility of somebody going to the night club is shown in Table 1. We likewise 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 particular deck, where detailed spatial data about the cabins and the possible pathways leading to the night club area are explained. By running the matching simulations, we are able to visualize the possible concentration of passengers during the night at this location of the ship (see Figure 6). As a result, by estimating the energy requirements of the night club with respect to the number of passengers hosted, we can compute the possible energy needs for the particular time period and facility (see Figure 7). Such estimations can be utilized for future forecasts of energy usage in cases where passengers are distributed in a similar method. Moreover, the obtained data can be statistically analyzed to expose the data patterns and mechanisms that may trigger the particular energy needs.

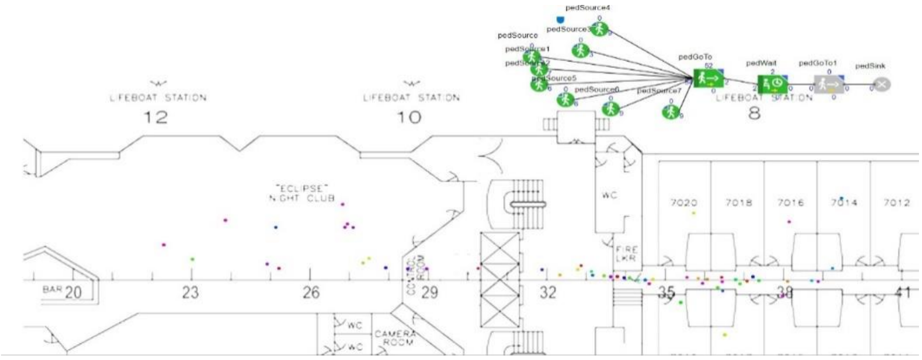

Fig. 6. An instance of a simulated energy consumption scenario in the nightclub [2]

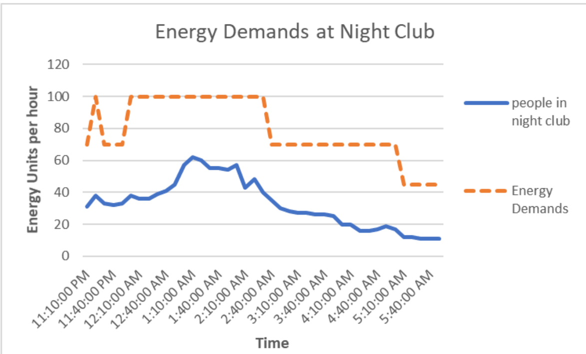

Fig. 7. Energy demands corresponding to passengers' concentration in the nightclub [2]

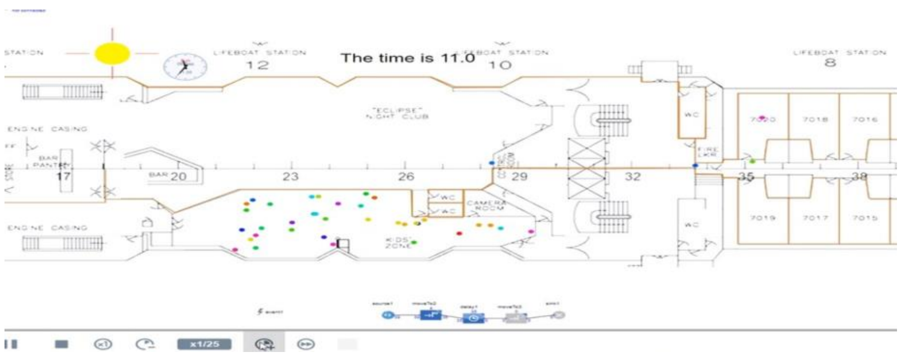

Fig. 8. An instance of a simulated energy consumption scenario in the kindergarten [2]

5.2 Kindergarten

For this facility (see Figure 8), we thought about that the passengers who visit it are primarily children (1-14 years of ages) and their parents (who may belong into the age groups of 15-34 and 35-54 years of ages). The opening hours of this facility are from 11am to 2pm. We presumed that the kindergarten is not the only choice that the above groups have for home entertainment functions. In addition, compared to other areas on the ship, the kindergarten is not that to accommodate all the families (parents with their kids). We have actually therefore presumed that the percentage of passengers visiting it daily varieties from 4% to 5.5%, i.e. from 120 approximately 176 persons. The time individuals 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 performed provided the concentration of passengers displayed in Figure 9.

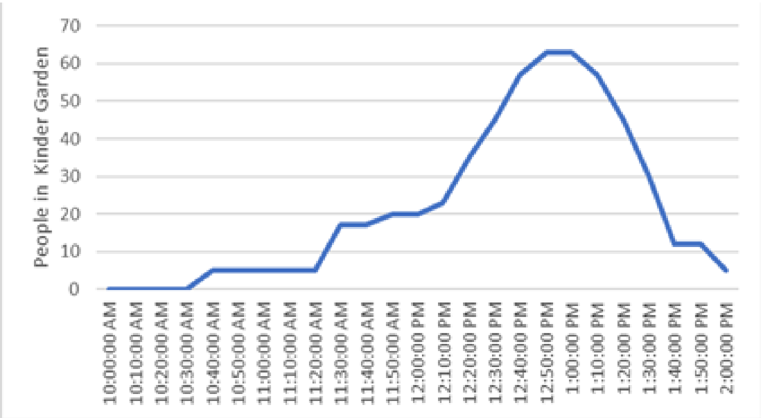

Fig. 9. Passengers' concentration in the kindergarten [2]

5.3 Casino

The samples of passengers utilized in the particular set of experiments concerned 320 people (i.e. 10% of average passengers' population). We presumed that this facility operates from 7pm to 7am and generally attracts passengers that are older than 35 years of ages (65% of them belonging to the ≥ 55 age group and the remaining 35% to the 35-54 age group). Additionally, passengers that go to the casino are divided into two categories, those who pick to waste their time solely in the casino throughout the night (20%) and those who visit the casino for a specific time period (they might leave and return to the casino throughout the night). The very 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). Likewise, 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 thought about one of the available ship restaurants (using an "à la carte" menu, therefore not being an economic one), running from 7pm to 11pm. This facility concerns all passengers, regardless of age group. We presumed that 10% -12% of passengers (320-380 individuals) 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 explained above show diverse features and options provided by the proposed simulation model. To forecast energy usage in big passenger and cruise ships, our approach aggregates results acquired from each specific facility of a ship and produces a matching time series diagram, in which the dependent variable is the energy consumption measured in energy units per hour and the time period is 10 minutes. Figure 10 highlights the total energy needs with regards to the estimated gathering of passengers in the facilities discussed in the previous section throughout the day. Obviously, our experiments have ruled out the totality of facilities and energy consumers readily available on a ship (such as air condition, lighting, heating and so on); nevertheless, all of them can be easily aggregated to our model and thus offer a comprehensive mapping of the total energy usage.

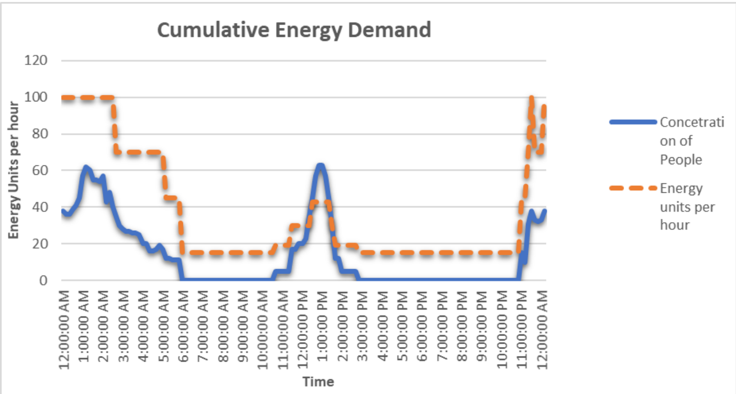

Fig. 10. Cumulative concentration of passengers in four major ship's facilities and corresponding energy demand [2]

Building on the proposed agent-based simulation model that assists in the development of alternative energy usage scenarios, we can produce realistic data that can be additional elaborated by popular machine learning algorithms to supply meaningful insights for handling diverse energy consumption patterns [8] Specifications taken into account by the proposed machine learning algorithms likewise consist of the variety of ship generators (categorical variable), the alternative age groups and their populations (as specified for each ship), and the time slots considered each time (the ones adopted in our method 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 specified, is presented in Table 3 (the variety of generators that run 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 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 integrated in our approach. Through the utilization of these algorithms, one might forecast the required variety 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 needed. As it can be observed, the energy usage 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 might be that older individuals use to be more active in the morning (compared to young populations). Results shown in Figure 12 provide extra evidence in favor of the above insight; as depicted, the correlation between the number of generators being utilized in the morning and the variety of elderly passengers is positive.

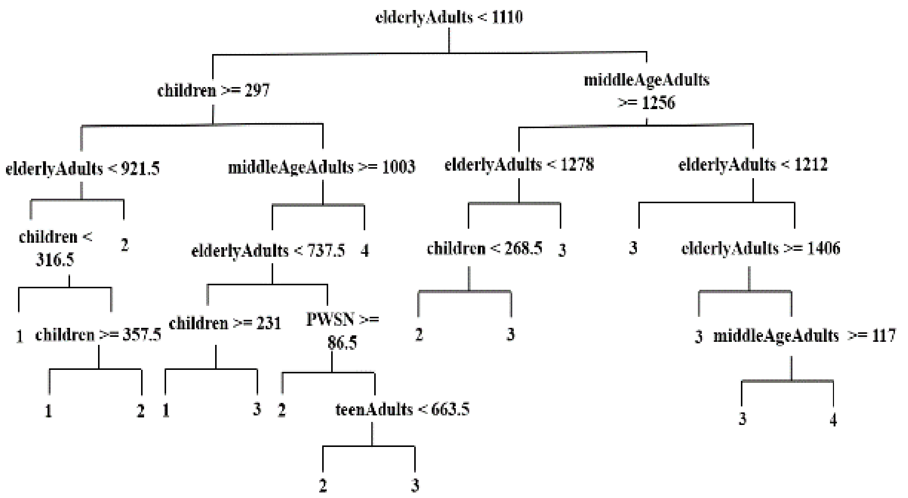

Fig. 11. Decision Tree classification ('children', 'teenAdults', 'middleAgeAdults' and 'elderlyAdults' correspond to the 1-14, 15-34, 35-54 and >=55 age groups, respectively) [2]

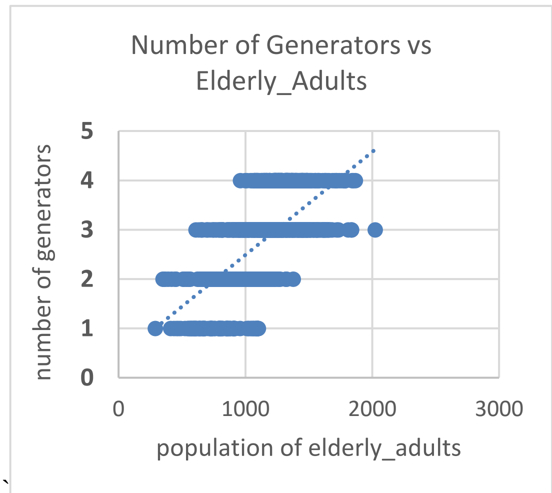

Fig. 12. Scatter plot - number of generators vs population of elderly passengers [2]

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

Table 4 sums up a little set of forecasts produced by the K-NN algorithm for the cases of one or four generators running simultaneously. It is noted that for these cases K-NN produces really comparable results to those acquired by the Decision Tree, i.e. the energy needs are positively associated to the ratio of passengers that are older than 55 to those that are younger than 35 years of ages. Such insights, arising from several simulation runs, were also confirmed by shipping stakeholders. According to their recognition feedback, adjustments to the initially set parameters and energy de-mand limits were carried out.

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 liner is definitely a tough issue. This is generally due to the requirement to at the same time consider the interaction between multiple criteria and agent behaviors. To deal with this problem, the proposed method mixes the process-centric character of a simulation model and the data-centric character of ML algorithms. Initially, by building on a comprehensive and informative agent-based simulation model, it helps with the generation and assessment of alternative energy usage scenarios that integrate vast amounts of realistic data under various conditions. Second, it promotes using popular machine learning algorithms to aid the finding, understanding and interpretation of patterns that are implicit in this data, ultimately intending to offer meaningful in-sights for shaping energy saving services in a ship. In addition, the general architecture of the integrated system is based upon the microservices approach. Microservices offer a way to intensify the development and delivery of big and complicated applications; they likewise enable individual parts to progress individually of each other.

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

Another research study direction concerns the examination of alternative modes to integrate simulation and artificial intelligence in our method. Specifically, we plan to consider the application of ML algorithms prior to and within the simulation. In the previous case, we will need real data to develop rules and heuristics that our agent-based simulation design can then use. In the latter, we may recycle previously trained ML-based models or train the ML models as the simulation is occurring. In addition, we prepare to investigate extra ML algorithms.

Lastly, we plan to broaden the proposed agent-based simulation model with problem specific algorithms and , intending to make it possible for shipping stakeholders perform a progressive synthesis and numerous criteria comparative evaluation of alternative energy consumption setups (a similar method has actually 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).