2023 IEEE 3rd International Conference on Sustainable Energy and Future Electric Transportation (SEFET) | 979-8-3503-1997-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/SEFET57834.2023.10245183

Reinforcement Learning Based Cost-Effective Smart Home Energy Management

Arpita Benjamin
Dept. of Electrical Engineering
National Institute of Technology Warangal
Warangal, India
abee21205@student.nitw.ac.in

Altaf Q. H. Badar
Dept. of Electrical Engineering
National Institute of Technology Warangal
Warangal, India
altafbadar@nitw.ac.in

***Abstract*—Demand Response (DR) techniques are regarded as the most economical and reliable way to smooth the load curve in context of the rising energy demand. In this paper, using Fuzzy Reasoning (FR) and Reinforcement Learning (RL), we have proposed a cost-effective strategy for residential demand response. This algorithm employs Q-learning, a reinforcement learning technique based on a reward system, to schedule shiftable/controllable loads optimally so that they are shifted from peak to off-peak hours of tariff. This reduces the overall electricity expenditure of a smart home while taking user comfort into account. FR is used for reward matrix generation. The suggested method works with one agent to operate 8 home appliances and makes use of fuzzy logic for rewards functions and a smaller number of state-action pairs to assess the action taken for a specific state. The Smart Home Energy Management System (SHEMS) demonstrates the application of the suggested DR scheme through MATLAB. The findings indicate that the cost of the electricity bill was reduced by 38.28%, showing the efficacy of the suggested strategy.**

***Index Terms*—Reinforcement learning, Demand response, Q-learning, Smart home energy management system, Fuzzy reasoning**

I. Introduction

The demand for power has been increasing continuously. DR is regarded as an essential component that can assist customers in managing their energy consumption. Consumer participation in such strategies will lower power use at times of high demand, i.e., peak hours, and result in lower electricity costs. DR is characterised as a shift in end-users pattern of electricity consumption in response to variations in tariff and other monetary incentives provided by the energy supplier.

DR programs are categorised into: price-based programs and incentive-based programs [1]. Consumers that participate in an incentive-based DR strategy receive financial incentive for switching their consumption from peak (high demand) hours to off-peak (low demand) hours. In exchange for their participation in the program, these customers receive a discounted rate or bill credit payment. All tariff programs that offer clients financial incentives or rebates in exchange for reducing their electricity use during particular hours are considered price-based programs. Such programs offer various electricity tariff pricing to assist customers in obtaining optimized power. By implementing tariff price programs like critical peak pricing (CPP), time of use (TOU) etc., consumers actively change the amount of electricity used in their homes based on tariff [1].

SHEMS enables consumers to reduce their electricity costs and utilize energy optimally. It can be considered a potential system for realizing DR strategies. A detailed review on SHEMS is compiled in [2]. A practical application of SHEMS is demonstrated in [3]. In [4], authors have concentrated on the HEMS algorithm while considering appliance priority, user preferences and comfortable lifestyle. The development of HEMS controllers using machine learning and computational intelligence approaches has received a lot of interest recently. Reinforcement Learning (RL) has recently come to light as a promising machine learning method for decision-making, control, and energy management. Due to its capability to solve issues without requiring prior knowledge of the environment, RL models offer exceptional decision-making abilities [5]. In [6], reinforcement learning technique has been proposed, using multiple agents where every appliance has its own agent and environment. In [7], [8], authors have concentrated on scheduling home appliances such that operating time of shiftable devices is shifted using SARSA (State-Action- Reward-State-Action) in Q-learning, but these methods slows down the Q-values' ability to converge.

This study proposes a flexible SHEMS which uses one agent and a smaller number of state-action pairs where fuzzy reasoning is used for the generation of reward function without sacrificing user's preferences and comfort.

II. Proposed Smart Home Energy Management System

The suggested model installs a SHEMS at the consumer end that can monitor, calculate, and optimize the use of energy and minimize electricity bills. Fig. 1 shows different components of the system considered. The system includes smart meter, communication network (LAN), EV, and home appliances. We consider that the smart meter gets the grid provided energy pricing signals and communicates the values to SHEMS. A particular SHEMS module's fundamental functions are device data collection, data processing, and load control. The strategies for load control and the analysis of the collected data are both parts of processing and intend to schedule the devices using various optimization methods. Finally, the developed

schemes serve as the basis for the HEMS load control. Existing LAN access protocols, such as Wi-Fi and Zigbee, which can accommodate a variety of communication applications, can be used to facilitate communication between the devices and HEMS for the purposes of data collection and load control [9].

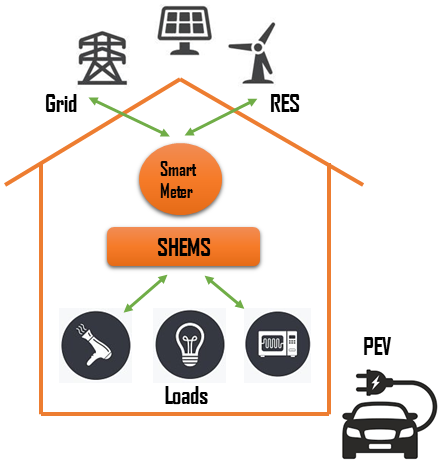

Fig. 1. Proposed Smart Home Energy Management Framework

III. Reinforcement Learning

One of the primary Machine Learning (ML) methods for selecting the best decisions in a non-deterministic setting is RL [10]. As shown in Fig. 2, an agent learns the appropriate action based on the current state of the environment when it interacts with it and a newly learnt action is delivered to the environment. The environment in return, gives the agent a reward and the updated state (next state) of the environment. Until the agent optimizes the total cumulative rewards acquired from the environment, this learning process will continue. The agent's main objective is to maximize the reward by identifying the best policy. A policy is defined for the way an agent behaves in a particular condition. A mapping from the observation of the current environment to a probability distribution of the actions to be taken is referred to as a reinforcement learning policy.

Fig. 2. Reinforcement Learning Architecture

Q-learning determines the optimal policy $\nu*$ of a decision-making problem. In Q-learning technique, a Q-value $(Q(s_t, a_t))$ is calculated where s_t is state and a_t is action at a discrete time t and the Q-value is updated using the maximum reward value using the Bellman equation as shown below:

$$Q_\nu *^* (s_t, a_t) = r(s_t, a_t) + \mu \cdot \max Q(s_{t+1}, a_{t+1}) \quad (1)$$

According to Eq. (1), the best Q-value is obtained by adding the maximum future reward which is discounted, $\mu \cdot \max Q(s_{t+1}, a_{t+1})$, and the present reward, $r(s_t, a_t)$, depending on the most optimal policy, $\nu*$ where μ is the discounting factor and is given as $\mu \in [0, 1]$. When μ=0, suggests that only current reward is considered, while μ=1 suggests that agent will focus on future rewards. The calculated or defined reward $r(s_t, a_t)$, will be obtained once the action a_t is executed in accordance with policy $\nu*$, and the agent will then enter a new state (s_{t+1}). Using the following Eq. 2, the action value in $Q(s_t, a_t)$ is changed concurrently:

$$Q(s_t, a_t) \leftarrow (1-\alpha) Q(s_t, a_t) + \alpha \begin{bmatrix} r(s_t, a_t) + \\ \mu \cdot \max Q(s_{t+1}, a_{t+1}) \end{bmatrix} \quad (2)$$

$$\nu* = \text{argmax}\, Q(s_t, a_t) \quad (3)$$

where α is the rate of learning, which establishes the extent to which the previous Q-value is influenced by the current reward. α=0, indicates that agent has learned nothing and makes use of the previous Q-value in the process of learning. While when α=1 indicates the most recent information is considered. Finally, by iteratively updating $Q(s_t, a_t)$ using Eq. (2), the Q-value $(Q(s_t, a_t))$ will become bigger and bigger.The agent will eventually acquire the optimal policy $(\nu*)$ with the biggest Q-value using Eq. (3).

IV. Demand Response Strategy using Q Learning Model

A. Home Energy Management Model

Typically, home appliances can be categorised into non-shiftable & shiftable. Appliances categorized as shiftable are washing machines, dishwashers, and clothes dryers that can be operated at any time during the user's specified time period. Appliances like refrigerators, water heaters, and lighting that require a constant supply of electricity to carry out their functions are referred to as non-shiftable. The non-shiftable load needs to be utilized only during predetermined times. As a result, in DR schemes only the shiftable appliances may participate.

In this study, 8 household appliances of a smart home are chosen that will be managed using the best scheduling model [11]. The Power Demand of shiftable appliances can be defined as follows:

$$PD_t^{tot} = \sum_{x=1}^{X} E_t^{tot} * I_t^x \quad (4)$$

where PD_t^{tot} is the total power demand considering only the shiftable loads for every hour, E_t^{tot} is rated power of each appliance, I_t^x is ON (1) or OFF (0) status of each appliance

at a specific hour $t \in [1, 2, 3. . . .24]$, and X is the summation of all shiftable appliances wherein $x \in [1, 2. . . . .X]$.

The data of shiftable household appliances in a smart home are listed in Table I [11], including the rated power of each appliance, required time period of operation, and the priority of operation taking user's comfort into account.

TABLE I
DATA OF SHIFTABLE HOME APPLIANCES

Appliance	Rated Power (W)	Duration cycle (min)	Priority Order
Washing Machine	500	60	1
Water Pump	1800	180	2
Dishwasher	800	60	3
Tumble dryer	750	120	4
Microwave	1200	60	5
Electric Water Heater(EWH)	2000	120	6
EV1	6000	300	7
EV2	7000	300	8

B. Q Learning Model

An intelligent agent uses RL to make the best decision possible in a particular stochastic setting with variable patterns of energy usage and electricity prices. The dynamic system can be controlled by the agent by carrying out a series of actions. Such a system is defined by a financial incentive and state-space.

The Power Demand (PD) and the Electricity Cost signal (EC) for electricity serve as models of the state-space in this case. PD is categorized into low, medium, and high. EC is categorized into less and expensive as follows:

$$PD^{\text{total}}_{t,\text{ index}} = \begin{cases} PD^{\text{tot}}_{\text{low}}, & \text{if } PD^{\text{tot}}_t \leq 7.5 \text{ kW} \\ PD^{\text{tot}}_{\text{medium}}, & \text{if } 7.5 < PD^{\text{tot}}_t < 9.24 \text{ kW} \\ PD^{\text{tot}}_{\text{high}}, & \text{if } PD^{\text{tot}}_t \geq 9.24 \text{ kW} \end{cases} \tag{5}$$

$$EC^{\text{index}}_t = \begin{cases} EC^{\text{less}}_t, & \text{if } EC_t \leq 1¥/kWh \\ EC^{\text{expensive}}_t, & \text{if } EC_t > 1¥/kWh \end{cases} \tag{6}$$

Table. II shows state index that can be created from PD and EC.

TABLE II
INDEX OF STATES BASED ON POWER DEMAND AND ELECTRICITY COST

Power Demand	Electricity Cost	State
PD_{low}	EC_{less}	1
PD_{low}	$EC_{expensive}$	2
PD_{medium}	EC_{less}	3
PD_{medium}	$EC_{expensive}$	4
PD_{high}	EC_{less}	5
PD_{high}	$EC_{expensive}$	6

One action is selected by the agent from action space; A as given below:

$$A = [Transfer, Fill\ Valley, Stay\ Idle] \tag{7}$$

where Transfer action transfers the least priority load and this action is taken during periods of high demand. The goal of Fill Valley action is to switch on the home appliance of highest priority that was shifted, typically during times of low demand. When set to Stay Idle, the system operates normally, and no appliance is shifted.

In Q-learning model, the reward's $(r(s_t, a_t))$ objective is to assess the extent to which the action taken was appropriate for a specific state. Fuzzy Reasoning deals with approximations rather than precise values. In order to assess the action taken for a particular state, fuzzy logic is utilized here. Fuzzy inference is a technique that assigns values to the output vector based on the interpretation of the input vector values, using a set of rules. In this paper, Mamdani method is utilized because it provides a smoother output. In this Fuzzy Inference System(FIS), the input variables are PD and EC as shown in Fig. 3 whereas the output variables are Transfer, Fill Valley, Stay Idle. Table. III shows the fuzzy rules of FIS. Action taken for each state is evaluated using FIS. The fuzzy sets are defined as Poor Action (PA), Fine Action (FA) and Super Fine Action (SFA) for each output(action), as shown in Fig. 4.

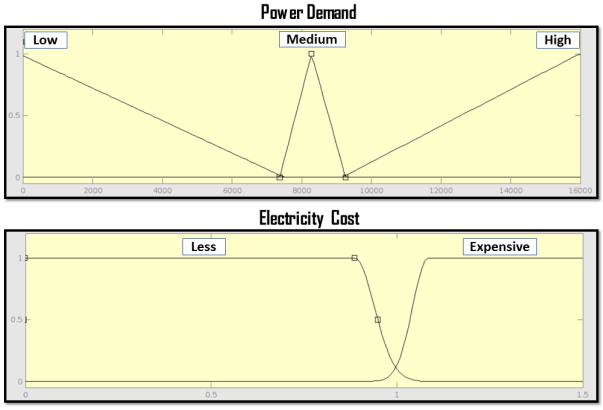

Fig. 3. Input Variables to FIS (Power Demand, Electricity Cost)

TABLE III
FUZZY RULES OF FIS

Power Demand	Electricity Cost	Transfer	Fill-Valley	Stay Idle
Low	Less	PA	SFA	FA
Low	Expensive	PA	FA	FA
Medium	Less	PA	FA	SFA
Medium	Expensive	PA	FA	SFA
High	Less	PA	PA	SFA
High	Expensive	SFA	PA	PA

V. Q LEARNING ALGORITHM FOR SHEMS

The [states*actions] dimension of the Q-matrix should be initialized to zero. Once the agent has interacted with the environment after taking action in a given state, it will update each pair in the Q-matrix using Eq. (2). The optimal Q-values will be achieved only after Q-matrix has converged. In this paper, an adequate number of epochs are used to

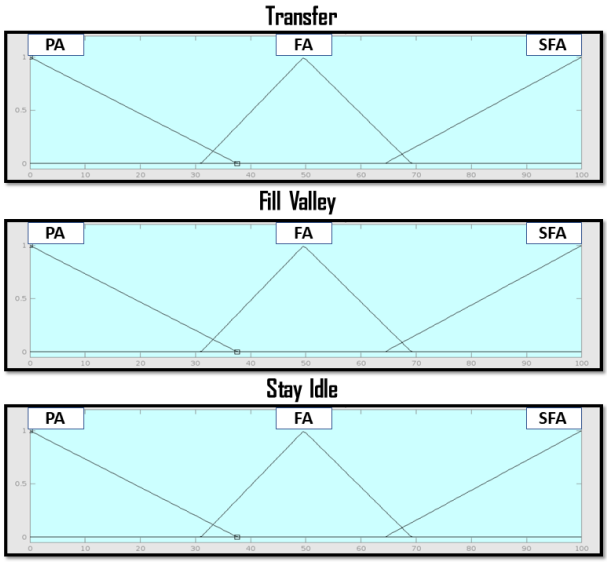

Fig. 4. Output Variables of FIS (Transfer, Fill Valley, Stay Idle)

investigate and update the values of $Q(s_t, a_t)$. This is known as "exploring" in which random action is taken at each step. The algorithm for Q Learning model is given in Fig. 5. The parameters μ and α are assigned values 0.9 and 0.1 respectively.

Q Learning Algorithm	
1.	Initialise $Q(s_t, a_t)$, $s \forall S$, $a \forall A$
2.	Set μ and α parameters and rewards in matrix as in Table 3.
3.	**for** each time step t **do**
4.	Choose random initial state
5.	**while hour= 1:24**
6.	Determine all available actions and select random action for current state.
7.	Execute selected action a_t and observe the state s_{t+1} and numerical reward $r(s_t, a_t)$
8.	Determine the maximum Q-value for next state in Q-matrix.
9.	Update the $Q(s_t, a_t)$ using Equation (2)
10.	Set the next state as current state
11.	End **while**
12.	End **for**

Fig. 5. Algorithm

VI. Result And Discussion

In SHEMS, smart meters collect power data from all home appliances and get pricing signals from the grid. The electricity price signals (TOU) in ¥/kWh is shown in Fig. 6 as received from the grid. Fig. 7 shows the cumulative power demand (Watts) of a smart home comprising of shiftable home appliances. At each time step, the agent receives PD and EC values, which define the state of the system. The converged Q-Matrix shown in Table. IV is used by the SHEMS to make an optimal decision. This involves shifting the lowest-priority appliance operating time, during high demand hours and turning on the highest-priority appliance that was shifted earlier, during low demand hours. For the current state, depending on the maximum Q-value in the converged Q-matrix, the chosen action will be taken. The actions to be taken over 24 hours is shown in Fig. 8. This method is based on the correlation between the TOU pricing & the total power consumption of all home appliances while accounting for consumer comfort preferences, and the load priority.

Fig. 9 shows different states detected, based on PD and EC at different hours. Like at 6:00 pm, the state index is 4 because PD is medium (8.2 KW) and EC is expensive (1.1 ¥) in that hour. Using the algorithm mentioned in Fig. 5 and Table. IV, it can be observed that the maximum Q-value (10.1571) is for Stay Idle, hence that action should be taken. Similarly, at 1:00 am, the state is 1, and from Table. IV it can be observed that the maximum Q-value is for Fill Valley, hence load is increased at this hour.

Fig. 10 shows the final power consumption profile of a smart home considering all appliances over 24 hours after using RL Algorithm for the management of loads, along with PD. On calculating the electricity bill using the final power consumption profile and TOU pricing, it is observed that there is a 38.28% reduction in bill cost considering all the shiftable appliances.

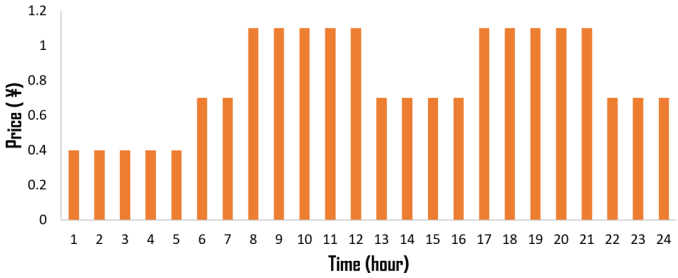

Fig. 6. TOU Pricing Signal

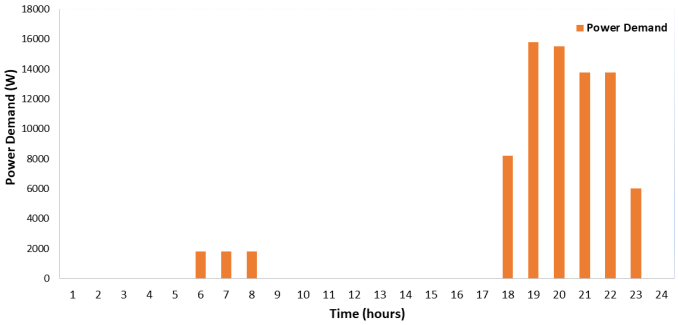

Fig. 7. Power Demand Profile of Smart home

VII. Conclusion

This study proposed a DR strategy to shift load demand of a smart home from peak hours to off-peak hours in response to users' priority of home appliances & electricity cost signal. In this paper, a Q-learning-based SHEMS is created to manage various power consumption patterns and fluctuating electricity

TABLE IV
CONVERGED Q MATRIX(10000 ITERATIONS)

State Index	Action		
	Transfer	*Fill-Valley*	*Stay Idle*
1	2.3375	9.908	6.7789
2	3.0756	6.0733	6.162
3	2.3427	6.0723	10.5454
4	3.4072	2.736	10.1571
5	2.3385	2.3881	9.8834
6	9.8582	2.4255	2.4237

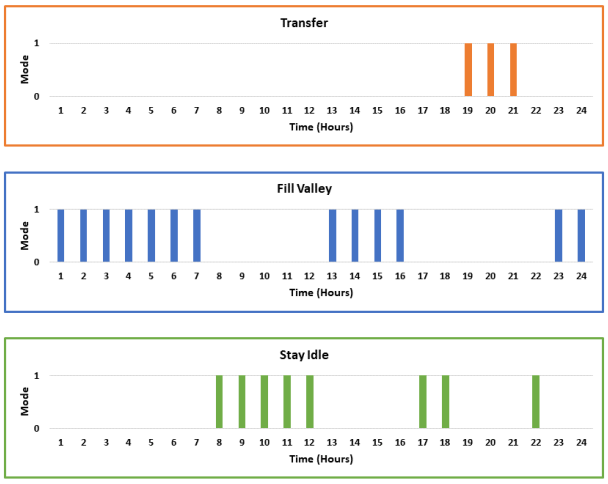

Fig. 8. Actions taken depending on current state and converged Q Matrix

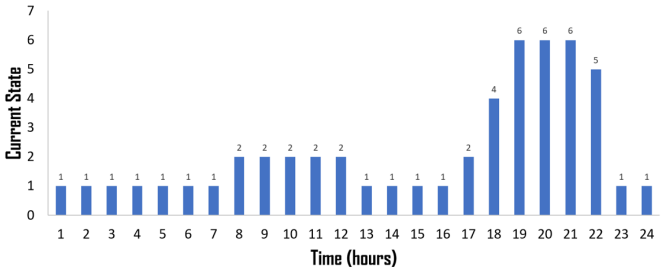

Fig. 9. Different States during one day

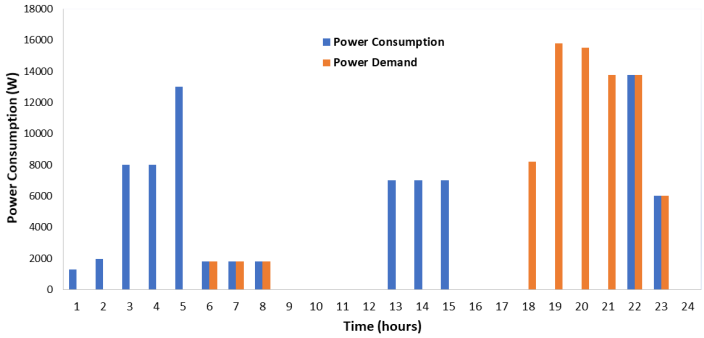

Fig. 10. Power Consumption profile after RL implementation

prices without sacrificing the customers' comfort. The suggested method uses one agent to operate 8 home appliances using fuzzy reasoning for reward generation for a specific state. MATLAB is used for implementation and demonstration of proposed DR strategy in a smart home. The findings indicate that the cost of the electricity bill was reduced by 38.28%, showing the efficacy of the suggested strategy.

REFERENCES

[1] H. Shareef, M. S. Ahmed, A. Mohamed, and E. Al Hassan, "Review on home energy management system considering demand responses, smart technologies, and intelligent controllers," IEEE Access, vol. 6, pp., 2018.

[2] Badar, Altaf QH, and Amjad Anvari-Moghaddam. "Smart home energy management system–a review." Advances in Building Energy Research 16.1 (2022): 118-143.

[3] Chauhan, Rajeev Kumar, Kalpana Chauhan, and Altaf QH Badar. "Optimization of electrical energy waste in house using smart appliances management System-A case study." Journal of Building Engineering 46 (2022): 103595.

[4] M. Ahmed, "A home energy management algorithm in demand response events for household peak load reduction," Przeglad Elektrotechniczny, vol. 1, no. 3, pp. , Mar. 2017.

[5] H. Zhang, D. Wu, and B. Boulet, "A review of recent advances on reinforcement learning for smart home energy management," in Proc. 2020 IEEE Electric Power and Energy Conference (EPEC), pp. 1–6, 2020.

[6] Lu, R., Hong, S. H., and Yu, M. (2019). Demand response for home energy management using reinforcement learning and artificial neural network. IEEE Transactions on Smart Grid, 10(6), 6629-6639.

[7] S. Kim and H. Lim, "Reinforcement learning based energy management algorithm for smart energy buildings," Energies, vol. 11, no. 8, p. 2010, 2018.

[8] N. Chauhan, N. Choudhary, and K. George, "A comparison of reinforcement learning based approaches to appliance scheduling," in Proc. 2nd Int. Conf. Contemp. Comput. Inform. (ICI), Dec. 2016, pp. 253-258.

[9] D.-M. Han and J.-H. Lim, "Design and implementation of smart home energy management systems based on zigbee," IEEE Trans. Consum. Electron., vol. 56, no. 3, pp. 1417–1425, Aug. 2010.

[10] Lee, Sangyoon, and Dae-Hyun Choi. "Reinforcement learning-based energy management of smart home with rooftop solar photovoltaic system, energy storage system, and home appliances." Sensors 19.18 (2019): 3937.

[11] Lu, X., Zhou, K., Chan, F. T., and Yang, S. (2017). Optimal scheduling of household appliances for smart home energy management considering demand response. Natural Hazards, 88, 1639-1653.