ORIGINAL RESEARCH **Open access**

Optimizing Energy Trajectories in Islanded Microgrids through Smart Management and Behavioural Modelling

Saeed Al-Mansoori[1*], Hessa Al-Hammadi[1], Mohammed Al-Blooshi[2], Daniel Green[3]

Abstract

As the modern world becomes increasingly interconnected and reliant on electricity, generating renewable power, ensuring reliable distribution, and adopting an energy-restraint lifestyle at both individual and societal levels will be crucial for maintaining global stability. Therefore, designing effective energy storage solutions combined with resilient energy management frameworks is highly valuable; however, many existing studies isolate the human behavioural element. Although human conduct may have a negligible effect in massive industrial facilities, its influence is critical and cannot be ignored in residential neighbourhoods. To address this, the current study introduces a novel, adaptable population dynamics model drawing on epidemiological techniques to simulate a wide array of societal conditions. We establish a comprehensive framework by linking this population model with an intelligent energy management approach. Specifically, in addition to theoretically pinpointing sustainable operational parameters across numerous setups, our empirical results highlight the effectiveness of reinforcement learning agents in executing optimal energy management. In this specific scenario, the optimized agent expands the boundaries of sustainability by up to 200% compared to conventional strategies, suggesting a path toward less disruptive societal transitions.

Keywords Reinforcement learning, Optimization, Energy storage and management, Population dynamics

*Correspondence:
Saeed Al-Mansoori
saeed.mansoori@ku.ac.ae
[1] Department of AI for Desert Solar Energy Systems, College of Engineering, Khalifa University, Abu Dhabi, United Arab Emirates
[2] Department of Intelligent Energy Analytics and Smart Cities, College of Engineering, United Arab Emirates University, Al Ain, United Arab Emirates
[3] Department of AI for Smart Grid Resilience, Faculty of Engineering, University of New South Wales, Sydney, Australia

Introduction

Abundant scientific literature [1-5] emphasizes the critical need for prompt and profound reductions in emissions across every sector to prevent extreme global warming. Failing to meet the targets established at the Paris Agreement will probably result in severe global consequences for humanity. Consequently, this worldwide predicament demands rapid and substantial transformations within society, involving both organizational bodies and ordinary citizens, since individual daily choices unequivocally add to greenhouse gas output [6-8].

Furthermore, despite sharp increases in energy costs and geopolitical conflicts over energy resources, significant segments of the public remain highly reluctant to embrace lifestyle shifts necessary for combating climate change. This reluctance renders the creation and calibration of national policies both delicate and complicated, as substantial public engagement and compliance are strictly required [9].

Within this framework, governing authorities will probably depend on computational simulations to investigate a multitude of scenarios and their potential impacts. As proven frequently throughout the COVID-19 crisis [10], such simulation

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

models offer a valuable, cost-effective, and highly adaptable mechanism for strategizing and predicting the outcomes of specific systemic conditions. However, without interdisciplinary modelling [11, 12], these approaches are often deployed in isolation and may overlook broader systemic variables—an issue that the methodology proposed herein directly addresses.

Using the introduced model, isolated microgrids can represent individual structures, districts, or entire urban centres, based on the scale the policy planner intends to analyze.

Specifically, building upon findings in Refs. [13] and [14], which demonstrated the significant applicability of Reinforcement Learning (RL) for micro-grid energy management systems (EMS), this paper presents a combined framework that integrates demographic modelling with intelligent energy coordination. As illustrated in **Figure 1,** the framework assesses a theoretical isolated setup containing a community divided into distinct segments with differing levels of ecological awareness, labelled as A, P, and N. According to the specific environment under review, analysts and policymakers can scale the model to buildings, local districts, or whole municipalities linked to a renewable power generator and its integrated energy storage infrastructure. This allows for robust scenario analysis by establishing the rules that control social dynamics within the community and their responses to outside influences.

Practically, several shifting dynamics among the different community cohorts are implemented—including ideological shifts via social interaction or financial sensitivity—and these can be weighted variably to reflect unique societal structures. These configurations can subsequently be simulated over a specific period using the suggested dynamic model, determining if such variables result in manageable energy consumption for a selected storage strategy, as depicted in **Figure 2.**

Thus, the primary innovations of this research are rooted in the novel demographic model that merges intelligent EMS with adaptable population dynamics. It reveals distinct benefits of utilizing RL-driven policies to broaden the sustainability threshold amidst external limitations. More precisely, the comprehensive experimental framework indicates that advanced, data-oriented approaches can greatly facilitate societal energy transitions; ultimately, they empower officials to design more gradual social regulations that enjoy higher public acceptance.

To achieve this, the subsequent section reviews relevant literature alongside the various hurdles and constraints inherent to the overlapping disciplines connected to this study. Following that, the discussion shifts to demographic simulation methods, offering a detailed breakdown of the formalized transition frameworks. The subsequent segment details the experimental design: initially validating the utility of reinforcement learning algorithms for the energy management system compared to standard procedures, before displaying the outcomes from exploring various scenarios. In addition to pinpointing viable demographic parameters across diverse starting conditions, these findings highlight the way the introduced methodology expands the scope of sustainable policies, thereby smoothing public endorsement of the evolution required to meet energy limitations.

Related Works

As mentioned in the opening section, strong scientific data [4] are driving societal shifts aimed at reducing the hazards and consequences of global warming. Consequently, the past few years have seen a massive surge in eco-friendly, sustainable funding, alongside a growing share of renewable power in the worldwide electricity mix [15-18]. A crucial component of this progression is the development and implementation of micro-grids, which serve as a foundational pillar in shifting from centralized power generation to distributed network configurations [19]. This modern structural approach to energy distribution and access, tailored to meet local consumption adaptably, introduces several difficulties [20, 21]. Issues like volatility and intermittency arise because, unlike traditional production units that offer consistent and easily forecasted energy yields, these localized grids possess finite supply capabilities.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

To realize the diverse advantages promised by micro-grid systems, numerous studies have proposed methodologies for engineering the necessary, highly resilient energy management frameworks [22-24]. For example, the research in Taha and Mohamed [25] outlines a robust Model Predictive Control (MPC) system aimed at reducing factors such as operational expenses and pollutant emissions, whilst maintaining the daily charging cycles of batteries within appropriate bounds. Although yielding promising outcomes, these strategies depend heavily on domain expertise and simplifications to model environmental dynamics, which can potentially restrict the maximum achievable efficiency.

On a broader scale, despite impressive technological advancements in researching and utilizing micro-grids as the primary power source for localized areas, most approaches are evaluated under static conditions, typically relying on historical data sets [26, 27]. Proceeding in this manner, however, fails to offer an accurate perspective should human behavioural adaptation occur—an outcome that is highly probable in the context of energy transition and sobriety.

Importantly, the analysis of both singular and collective group behaviour forms the core of an extensive body of literature focused on the dynamic simulation of demographics [12, 28, 29]. For instance, Hellmann [30] details multiple non-linear interaction systems among different species, specifically examining how mutualism, competition, and predation shape the overarching trajectories of the considered cohorts. In a comparable context, Estes *et al*. [31] focus more exclusively on the function of predators within an ecosystem and demonstrate—as similarly indicated by Chu [11]—how such individuals can impact the diverse environmental cycles present in ecosystems.

Expanding upon these concepts and the corresponding non-linear models utilized to describe group interactions, numerous elements of this field proved vital during the management of the COVID-19 crisis [10]. Indeed, these methods empowered governing authorities to assess potential trajectories and, by leveraging pertinent data, to plan and formulate local measures to guarantee public safety as effectively as possible [32-33]. In Wilder *et al*. [34], an individual-centric model that accounts for specific geographical and demographic characteristics is proposed and applied to estimate the likely parameters dictating the spread of the epidemic. Specifically, this approach facilitated the simulation of localized contact restrictions and suggested that containment strategies need to be customized for specific populations, further underscoring the critical utility of population models.

From a different perspective, deep learning approaches are highly versatile, spanning a wide array of applications from advanced visual recognition [35] and protein folding [36] to strategic board games [37] and impressive natural language processing [38, 39].

Specifically, the challenge of discovering effective strategies within intricate, dynamic, and responsive settings is primarily addressed through reinforcement learning (RL).

As shown in Degrave *et al.* [40], RL agents can successfully contain and stabilize fusion plasma via tokamak control inside a nuclear reactor. As reported in OpenAI J. *et al.* [41], they can be trained to operate a robotic hand to solve a Rubik's cube.

Nevertheless, RL agents frequently exhibit brittleness [42, 43] because they are primarily trained and tested within stationary environments—a characteristic similarly present in several micro-grid studies. Specifically, agents optimized for rewards rely heavily on the independent and identically distributed (i.i.d.) assumption [44, 45] and typically suffer severe performance degradation when applied outside their training distribution. As highlighted in Chen *et al.* [46], numerous contemporary research efforts focus on improving agents' capacity to handle unfamiliar conditions, yielding promising yet still constrained outcomes.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

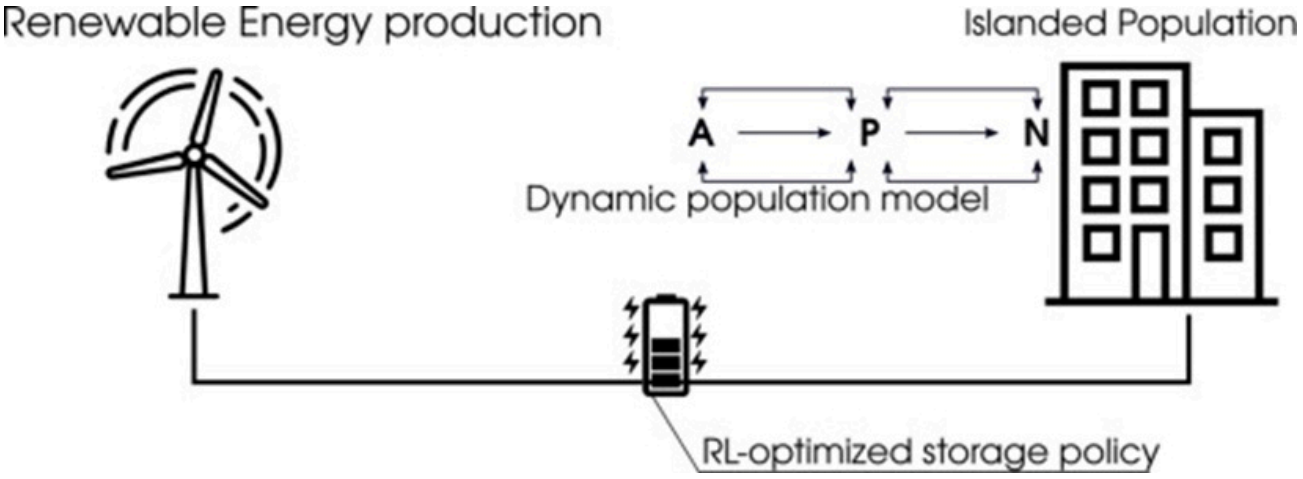


Figure 1. The present study models a theoretical isolated community, comprising diverse demographic groups and behavioural paradigms, linked to a renewable energy generator via a storage mechanism and its corresponding management policy.

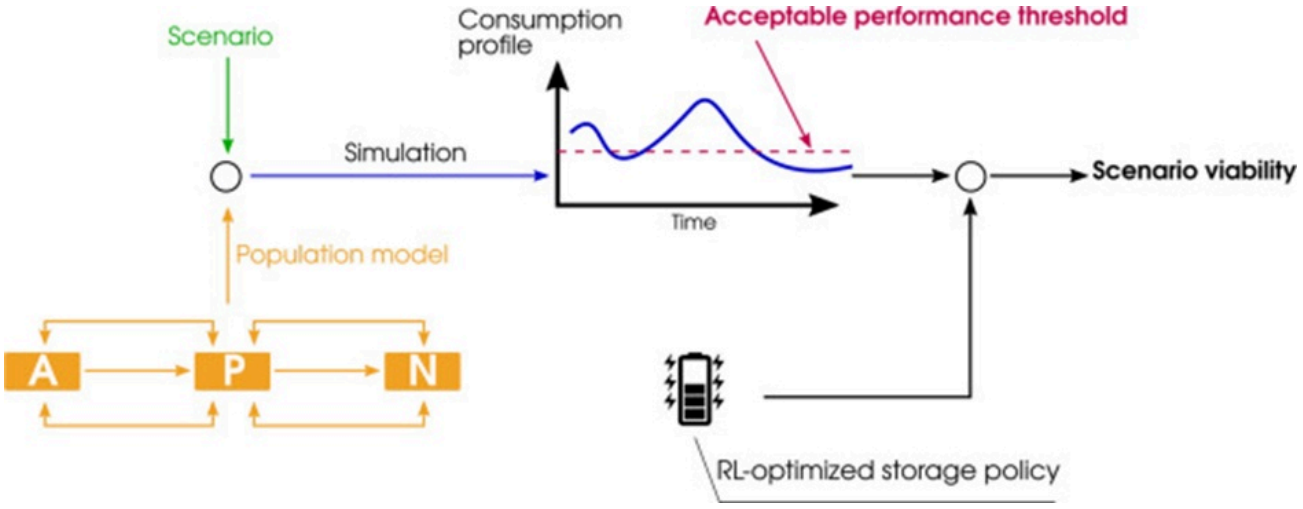


Figure 2. S3SAME integrates a dynamic demographic simulation with an RL-driven EMS approach to assess the impact of human habits on the overall system efficacy and to determine the compatibility of different scenarios with renewable energy generation.

Taking a broader perspective, it becomes evident that while micro-grid engineering and demographic simulation are highly distinct disciplines, they could mutually profit from integrating each other's principles. Specifically, demographic simulation could facilitate the precise forecasting of a micro-grid's necessary operational capacity, driven by predicted shifts in local consumption and its sustainability through an intelligent, potentially RL-guided EMS framework. Nonetheless, these domains are seldom investigated and applied collaboratively, neglecting crucial optimization opportunities and potentially vital insights. This gap consequently inspires the framework proposed herein, facilitating adaptable simulation and analysis of societal trends within renewable energy generation and coordination.

Dynamic Population Modelling

This segment outlines the primary theoretical assumptions used to generate and analyze the scenarios. It primarily details the population modelling approach, the various demographic cohorts participating in the dynamics, and elaborates on the transition mechanics that drive the movement of individuals between these cohorts.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Population modelling

As previously noted, overall energy usage is profoundly shaped by societal habits which, despite frequently being depicted as a generalized average, possess a much higher degree of diversity and nuanced complexity. Therefore, the architectural framework proposed in this study draws upon epidemiological simulation methods to establish a streamlined, yet realistic, demographic model comprising three distinct segments, each characterized by a unique consumption profile. Specifically, the model incorporates the following classifications:

- Environmentally active people (A): The demographic segment most committed to minimizing their energy usage.
- Environmentally passive (P): The transitional or moderate segment.
- Neutral (N): The segment that remains unengaged with climate and energy conservation concerns.

Transition mechanisms

As illustrated in **Figure 3**, several mechanisms can prompt individuals to modify their habits and subsequently transition between demographic groups. More precisely, a distinction is made between internal group dynamics stemming from interpersonal interactions and influences originating from external variables, both of which are thoroughly described in the subsequent sections.

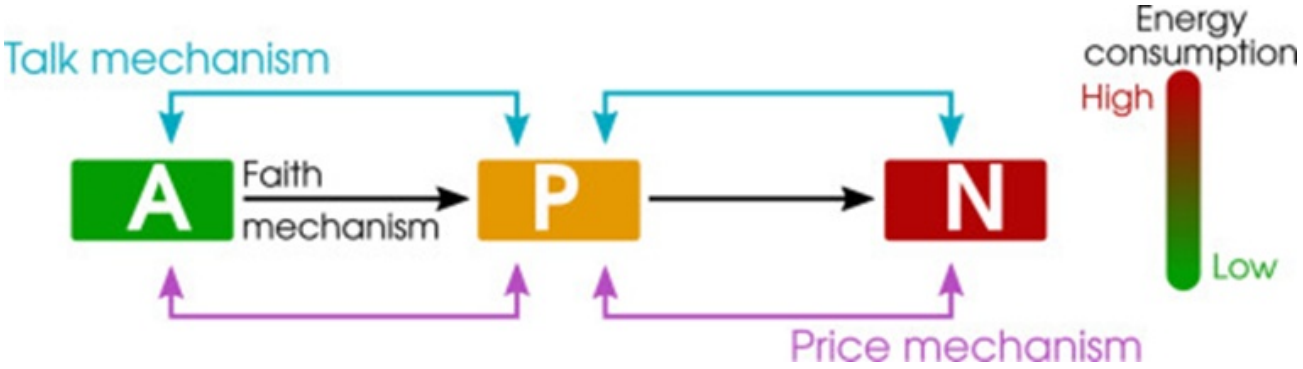


Figure 3. Population dynamics modelling: Members are influenced by various transitional triggers that motivate them to alter their mindset and embrace different behavioural routines, shaped by the prevailing circumstances.

Group dynamic factors

The advanced simulation architecture is founded on the premise that exchanges between members of distinct cohorts inevitably entail a sharing of perspectives and ideologies, occasionally leading one participant to adopt the other's group affiliation. Furthermore, internal mechanisms of pessimism are also incorporated. These reflect the notion that individuals might experience sufficient frustration or disillusionment to regress into the nearest less environmentally conscious cohort. Under this structure, the neutral demographic remains unimpacted by this regression, functioning effectively as an absorbing state.

External factors

While a broad spectrum of external variables and their consequences could theoretically influence the designed demographic model, the present study concentrates exclusively on the impact of price variations. Specifically, to mirror

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

the extreme price instability currently witnessed in global energy markets and the corresponding shifts in consumer habits, this research defines specific price sensitivity thresholds. Once exceeded, these thresholds trigger the migration of individuals either toward a more conservative group (during periods of exorbitant energy costs) or toward the nearest less conservative group (when energy prices remain accessible).

Reinforcement Learning Background

As highlighted in earlier sections, a primary contribution of this research is the integration of an intelligent, RL-driven EMS with the previously outlined demographic model. Consequently, this section supplies the fundamental theoretical concepts necessary to understand reinforcement learning. Although various policy optimization techniques could be applied to this domain, prior investigations [13, 47] into comparable problems validate the implementation of reinforcement learning methodologies [48].

In essence, the objective of reinforcement learning is to generate a policy that maximizes a defined success metric for a specific task. By interacting with the designated environment, the policy refines its sequence of chosen actions based on the received observations, aiming to produce operational trajectories that yield the highest possible cumulative rewards from the environment. Mathematically, the environment is typically formulated as a Markov Decision Process (MDP), comprising:

- States $s \in S$
- Actions $a \in A$
- Reward function $rt = f(st, at)$
- Transition probability function: $p(st+1|st, at)$

The growing fascination with reinforcement learning techniques has birthed several different frameworks [49-51]. Specifically, the current investigation utilizes Proximal Policy Optimization (PPO) [52], an on-policy, actor-critic algorithm featuring a trust region that has shown strong suitability for the complexity and specificities of the modeled environment [40]. Principally, the PPO approach seeks to adjust action selection probabilities in alignment with the advantage function of a specific state-action pairing. This mechanism is mathematically represented through the minimization of the following objective:

$$L(s, a, \theta_{\text{old}}, \theta_k) = \min(r_t(\theta_k) A^{\pi_{\theta_{\text{old}}}}(s, a), \text{clip}(r_t(\theta), 1-\varepsilon, 1+\varepsilon) A^{\pi_{\theta_k}}(s, a)) \quad (1)$$

Where $rt(\theta k)$ denotes the probability ratio between the preceding policy and the updated one for a specific state and action combination. This ratio is defined as:

$$\rho_t(\theta_k) = \frac{\pi_{\theta k}(a|s)}{\pi_{\theta_{\text{old}}}(s, a)} \quad (2)$$

RL Agent As an Efficient Energy Storage Policy

The industrial and economic landscape targeted by this research is the subject of intense and ongoing investigation, featuring a wide array of entities addressing the complexities of energy storage and distribution. Nonetheless, the exceptionally fine granularity of the existing methodologies and datasets complicates the establishment of a standardized benchmark. To our knowledge, there is no universally compatible, publicly accessible framework available to benchmark the techniques introduced in this paper against other leading-edge results. Given this limitation, and because rigorous evaluations are essential to prove the value and viability of the proposed methods, the subsequent sections detail the

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

framework utilized to develop baseline strategies. Essentially, these constitute the optimal policies the researchers could formulate without utilizing reinforcement learning algorithms, which are subsequently compared to the RL-optimized policy. Following this, we evaluate the scaling parameters across the entire range of demographic distributions.

Environment for policy generation

Before detailing the baselines, this segment outlines the core MDP components necessary for training the reinforcement learning-driven storage policy, which is the cornerstone of our framework. RL environments depend on an observation space, an action space, a reward signal, and a transition dynamic. However, because the current study examines a deterministic sequence, akin to Ref. [13], the transition function is omitted from the formalization. The other components are strictly defined as follows:

- The state space supplies an observation serving as the input for the agent to choose an action s = concat (s_{historic}, s_{storage}, s_{latent}) $\in \mathbb{R}^{H+1+L}$, where:
  - $s_{\text{historic}} \in \mathcal{R}^N$ indicates the historical ratio of production to consumption over the previous N intervals (the current study utilizes n = 8).
  - $s_{\text{storage}} \in \mathcal{R}^1$ represents the normalized status of current storage.
  - $s_{\text{latent}} \in \mathcal{R}^L$ denotes a latent vector supplied by the predictive model with dimension L (e.g., L = 16 in the present study).

- The action space is a continuous, single dimension: $at \in$ [-1, 1], and it governs the complete energy delivery and storage sequence:

  - In particular, for positive values, the framework banks the incoming energy proportionally to the action's magnitude and routes the remainder to the grid for user consumption. Mathematically:

$$\begin{cases} e_{\text{grid}} = (1 - a_t).p_{\text{turb}} \\ e_{\text{storage}} = \min[e_{\text{storage}} + a_t \times p_{\text{turb}}, e_{\text{max storage}}] \end{cases} \quad (3)$$

Where e_{grid}, e_{storage}, $e_{\text{max storage}}$, p_{turb}, at designate, respectively, the energy routed to the distribution grid, the currently stored energy, the absolute maximum storage capacity, the power generated in the preceding time interval, and the action dictated by the agent.

- Conversely, negative values direct all energy harvested in the previous interval to the grid while simultaneously adding the system's maximum energy output adjusted by the action's proportion. Mathematically, this is formulated as:

$$\begin{cases} e_{\text{grid}} = p_{\text{turb}} + \min[a_t * \Delta_{\max}, e_{\text{storage}}] \\ e_{\text{storage}} = \max[e_{\text{storage}} - e_{\text{grid}}, 0] \end{cases} \quad (4)$$

Where Δmax signifies the absolute peak energy volume extractable within a single environment time interval, a parameter strictly dictated by the underlying physical hardware.

- Lastly, the instituted reward function is structured to encourage the agent to dispatch a volume of energy that tightly matches the subsequent consumption demand. Mathematically, this is defined as $rt = ot + os + oc$, where:

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

- The principal objective: the energy supplied to the distribution grid must exceed the actual demand.

$$o_t = \begin{cases} 1 \text{ if } e_{\text{grid}} \geq e_{\text{consumption}} \\ 0 \text{ otherwise} \end{cases} \quad (5)$$

- $o_s = \frac{e_{\text{storage}}}{e_{\max \text{storage}}}$ incites the agent to maximize the storage system's capacity.
- $o_c = \exp \alpha \times (e_{\text{grid}} - e_{\text{consumption}})$ provides bonus points tied to accuracy, adjusted by a scaling factor.

Heuristics-designed baseline storage strategy

In practical terms, the formulated manual baseline is a reactive protocol, parameterized by $\gamma m \in \mathbb{R}2m$. Operationally, it incorporates the latest m historical production-to-consumption ratios as data inputs, along with an extra bias variable, to calculate the scalar action that is subsequently integrated into the MDP.

Mathematically, the manual policy's selected action at, γ is derived via the relationship illustrated in **Eq. 6.**

$$\begin{aligned} \vec{v} &= {}_{i\in[1..m]}[\gamma_i - R_{(n-i)}] \in \mathbb{R}^m \\ a_t &= \text{clip}(\vec{\gamma}_{i\in[m..2\ m]} \cdot \vec{v}, -1., 1.) \in \mathbb{R} \end{aligned} \quad (6)$$

Where n denotes the dimensionality of the observation, making $Rn\text{-}i$ the ith most recent ratio of energy production to consumption.

In execution, multiple manual strategies are developed, utilizing $m \in \{1 \ldots 5\}$, with their specific parameters refined via gradient ascent to yield the highest possible reward.

Ultimately, the findings indicate that the configuration employing m = 2 delivers the optimal average performance, and it is therefore the version retained for benchmarking against the reinforcement learning approach.

RL agent performance evaluation

Following the methodology outlined in Ref. [13], we designed an experimental framework and specific metrics to compare our trained policy against a manually developed baseline, aiming to validate the efficacy of these advanced methodologies. The proposed actor-critic model is trained using historical European electricity prices and average individual consumption data from 2015 to 2018, with 2019 serving as the testing period. During evaluation, each policy executes 200 episodes, recording the total hours of successful energy provision per episode. As illustrated in **Figure 4**, the reinforcement learning (RL) policy (light blue) consistently outperforms the manual baseline (red), supplying power to the population over significantly longer durations. Furthermore, given the high volatility characteristic of renewable energy sources, **Figure 5** depicts the per-episode distribution of energy availability when factoring in direct production. The trained agent again demonstrates superior storage efficiency, sustaining energy loads for extended periods. These outcomes substantiate the application of RL-driven techniques and subsequent agents for downstream operational tasks.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

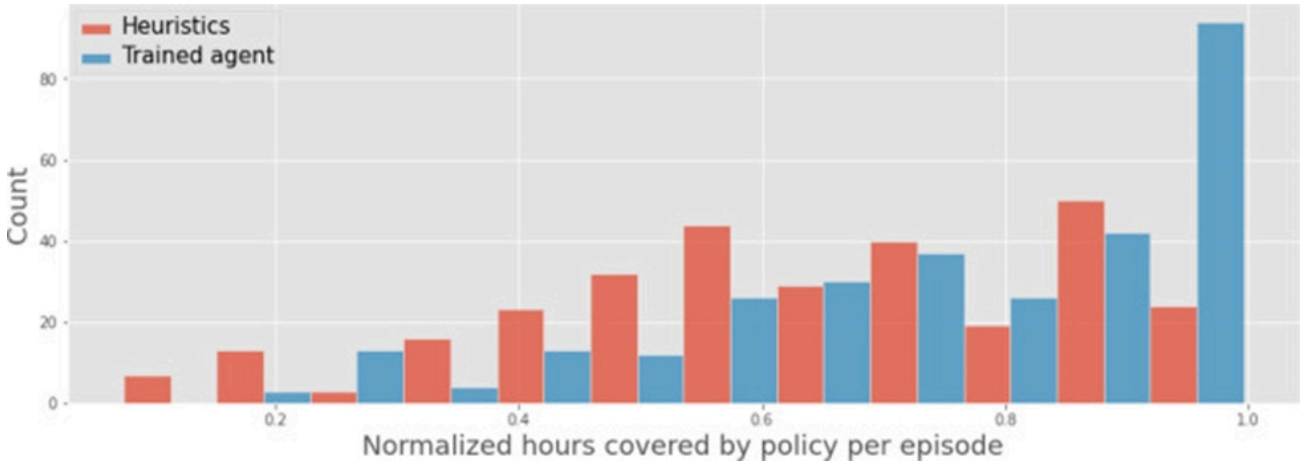


Figure 4. Distribution of normalized policy performance per episode. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

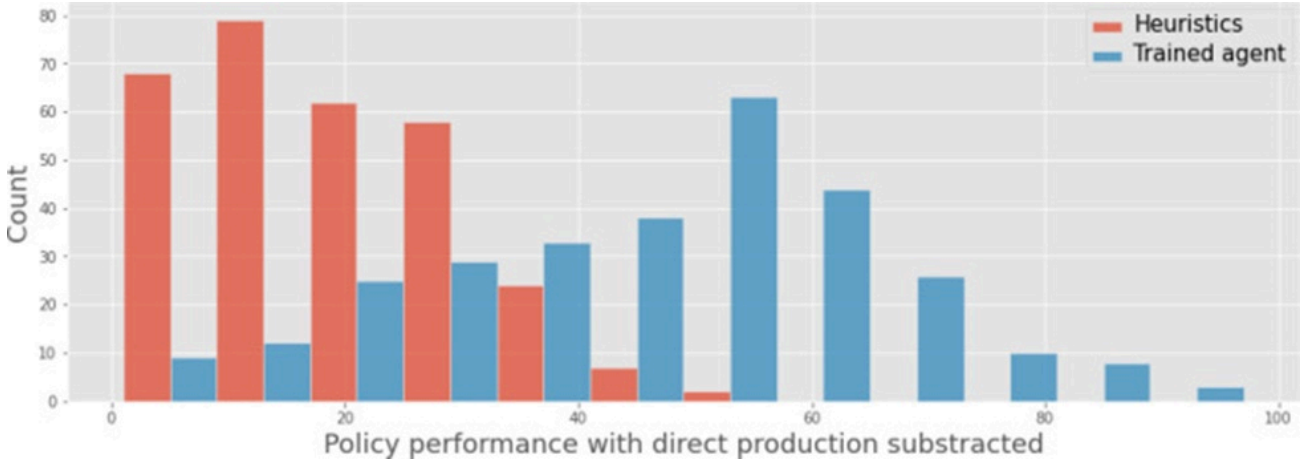


Figure 5. Distribution of normalized policy performance per episode.

Storage policy performances as a function of the general population consumption profile

Having validated the RL-trained agent as the optimal policy within this energy storage framework, delineating its performance boundaries is essential for informed decision-making. Since the policy is trained assuming a default consumption profile (i.e: α = 1), this section evaluates its efficacy across a broad spectrum of consumption variations. **Figure 6** illustrates system performance relative to the mean consumer profile. For every analyzed consumer factor, the policy processes 1,000 episodes, recording the operational performance ratio (hours with power versus total hours). The pronounced variance in performance across these episodes stems from highly diverse energy production conditions. Nonetheless, a strong negative correlation is evident between system performance and both the mean consumer profile and its corresponding consumption factor, as reflected by the mean performance curve. To provide comparative reliability context, baseline policy performance is also plotted (in red).

For a defined operational standard—governed by national regulations, designated energy sources, and other variables alongside the storage strategy—this established slope can dictate an α threshold beyond which average system performance cannot be assured. This α value subsequently informs the evaluation of social scenarios detailed below.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

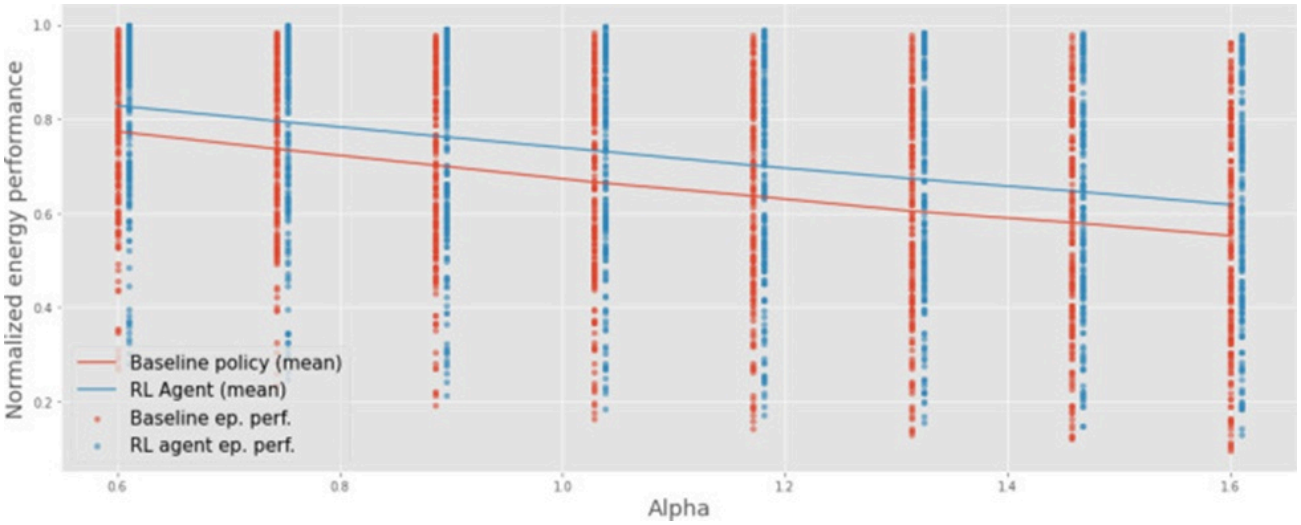


Figure 6. Mean performance level as a function of consumption profile for a given RL-trained storage policy (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Scenarios Definition and Evaluation

Population dynamics modelling

This methodology relies heavily on formulating a population dynamic model to simulate the distribution of individuals across three distinct behavioral profiles over a designated timeframe. Drawing on standard epidemiological frameworks [10, 32, 53] such as the SIR model, the following differential equations delineate the evolution of the active **(Eq. 7)**, passive **(Eq. 8)**, and neutral **(Eq. 9)** cohorts:

$$\begin{aligned}\frac{dA}{dt} = &- F_{A\to N}\times A \\ &+ (T_{N\to A} - T_{A\to N})\times A\times N \\ &+ (T_{P\to A} - T_{A\to P})\times A\times P \\ &+ H_{\text{upgrade}}(\text{price},\varepsilon_P)\times P\times H_{\text{speed}} \\ &- H_{\text{downgrade}}(\text{price},\varepsilon_A)\times A\times H_{\text{speed}}\end{aligned} \tag{7}$$

$$\begin{aligned}\frac{dP}{dt} = &- F_{P\to N}\times P \\ &+ (T_{N\to P} - T_{P\to N})\times P\times N \\ &+ (T_{A\to P} - T_{P\to A})\times A\times P \\ &- H_{\text{upgrade}}(\text{price},\varepsilon_P)\times P\times H_{\text{speed}} \\ &+ H_{\text{upgrade}}(\text{price},\varepsilon_N)\times N\times H_{\text{speed}} \\ &+ H_{\text{downgrade}}(\text{price},\varepsilon_A)\times A\times H_{\text{speed}} \\ &- H_{\text{downgrade}}(\text{price},\varepsilon_P)\times P\times H_{\text{speed}}\end{aligned} \tag{8}$$

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

$$\begin{aligned}\frac{dN}{dt} = &+ F_{A\to N} \times A \\ &+ F_{P\to N} \times P \\ &+ (T_{A\to N} - T_{N\to A}) \times A \times N \\ &+ (T_{P\to N} - T_{N\to P}) \times P \times N \\ &- H_{\text{upgrade}}(\text{price}, \varepsilon_N) \times N \times H_{\text{speed}} \\ &+ H_{\text{downgrade}}(\text{price}, \varepsilon_P) \times P \times H_{\text{speed}}\end{aligned} \quad (9)$$

Here, F denotes the internal mechanism capturing the loss of conviction, driving individuals toward a less sober demographic. T represents the communication mechanism governing inter-group interaction, quantifying the persuasive influence of each cohort. Lastly, H_{upgrade} and $H_{\text{downgrade}}$ are Heaviside functions representing the price thresholds (eg: εP) over or below which individuals shift demographics in response to price signals. Accordingly, the passive group dynamics defined by **Eq. 9** are structured as follows:

1. Subtract individuals whose diminishing conviction causes them to transition to less sober groups.
2. Add individuals from the neutral (less sober) cohort, calculated as the persuasion differential between the passive and neutral groups. This quantity can be negative.
3. Execute a parallel interaction process involving the active (more sober) cohort.
4. Subtract individuals migrating to the more sober cohort if prices exceed the tolerance threshold εP. This occurs at the rate Hspeed.
5. Add individuals upgrading from the neutral cohort.
6. Add individuals downgrading from the active cohort.
7. Subtract individuals downgrading from this current cohort toward the neutral group.

Scenario definition and critical configuration identification

The ultimate objective of this research is to develop an adaptable, modular framework for scenario generation, exploration, and the detection of critical system states under a specific storage policy. Once the linear relationship between system efficacy and mean consumer behavior is established and the critical threshold is identified, this policy outcome integrates into the population simulation module. Specifically, formalizing a scenario (i.e., establishing governing parameters for the population model) enables the projection of the mean consumer profile across the temporal horizon. This projected metric then serves as a benchmark for evaluating scenario viability. Operationally, this study utilizes the ratio of days exhibiting a specific mean consumer profile relative to the total timeframe, directly linking system performance to the scenario's foundational parameters. This can be compared against the threshold value to ascertain configuration acceptability, as depicted in **Figure 1.**

Results and Discussion

To demonstrate the utility of this framework, this section outlines multiple scenarios, their respective parameters, and potential analytical methodologies. It provides quantitative evidence that deploying an RL-based energy storage policy substantially expands the parameter space supporting sustainable operations, offering distinct advantages for social policy formulation. Additionally, it presents examples of how these findings can inform public policy, confirming the tool's practical relevance. Specifically, analyzing outcomes and state-space visualizations across these varied scenarios highlights strategic interventions capable of maximizing energy goals while averting critical system failures.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Because empirical data precisely aligning with this distinct population model was unavailable at the time of writing, we applied a broad spectrum of parameter values. Fundamentally, each simulation depends on the following parameters:

- Conversion speed (Price): The daily transition rate of a demographic when its defined price tolerance threshold is breached (positively or negatively).
- Faith loss: The daily proportion of individuals migrating to a less sober cohort due to diminished ideological conviction.
- Price thresholds: The specific financial boundaries that prompt population migration toward more sober (upgrade) or less sober (downgrade) cohorts. The conversion speed regulates this migration.
- Conversion (A → P): The daily transition rate from the actively sober demographic to the passively sober one, correlating with increased consumption. Conversely, Conversion (P → A) dictates the reverse migration flow.
- Conversion (P → N): Quantifies the rate at which passively sober individuals abandon sustainable practices to join the neutral cohort, while (P → N) also defines the rate of inverse conversion.

In practice, four primary scenarios are examined below, each featuring two variations (primarily differentiated by income levels to assess the influence of wealth). Their foundational narratives are summarized as follows:

- Military conflict (MC): Anticipating military engagement, the depletion of natural resources spurs severe inflation, instability, and uncertainty. This indirectly heightens public sensitivity to environmental constraints, accelerating transitions toward highly sober demographics.
- Economic resurrection (EC): The resurgence of post-crisis confidence prioritizes short-term objectives, leisure, and entertainment, consequently diminishing broader environmental awareness.
- Social tension (COTC): This model investigates escalating social friction and its resultant impact on aggregate energy demand. Drawing from contemporary dynamics, it assumes political mandates and societal shifts face heterogeneous public reception, heavily influenced by individual metrics such as income and geographic distribution.
- Winter is coming (WC): Confronted with ominous forecasts regarding winter energy supplies, the population exhibits heightened environmental sensitivity, demonstrating a pronounced tendency to adopt highly sober consumption behaviors.

Quantitative comparison of storage strategies

Although previous section demonstrated that RL-based storage strategies yield superior performance and returns compared to the suggested baselines, those prior experiments utilized a static setup without providing a broader perspective linked to behavioral variables. Consequently, assessing their effectiveness within a dynamic population framework is complex and might not immediately promote their implementation. Therefore, this section presents a comparative analysis of the resulting state space scores for both the manual baseline and the RL-based storage policies.

Accordingly, **Figure 7** depicts the performance discrepancies between the two approaches across all evaluated scenarios (four scenarios, each including one variant). Here, the state space scores in rows 1 and 3 are generated by the RL agent, while rows 2 and 4 reflect the manual policy. The sustainable configurations (indicated by yellow regions) under the RL strategy encompass a substantially larger portion of the state space, with improvements varying from a modest 4% (in the ER1 scenario) to over 200%.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

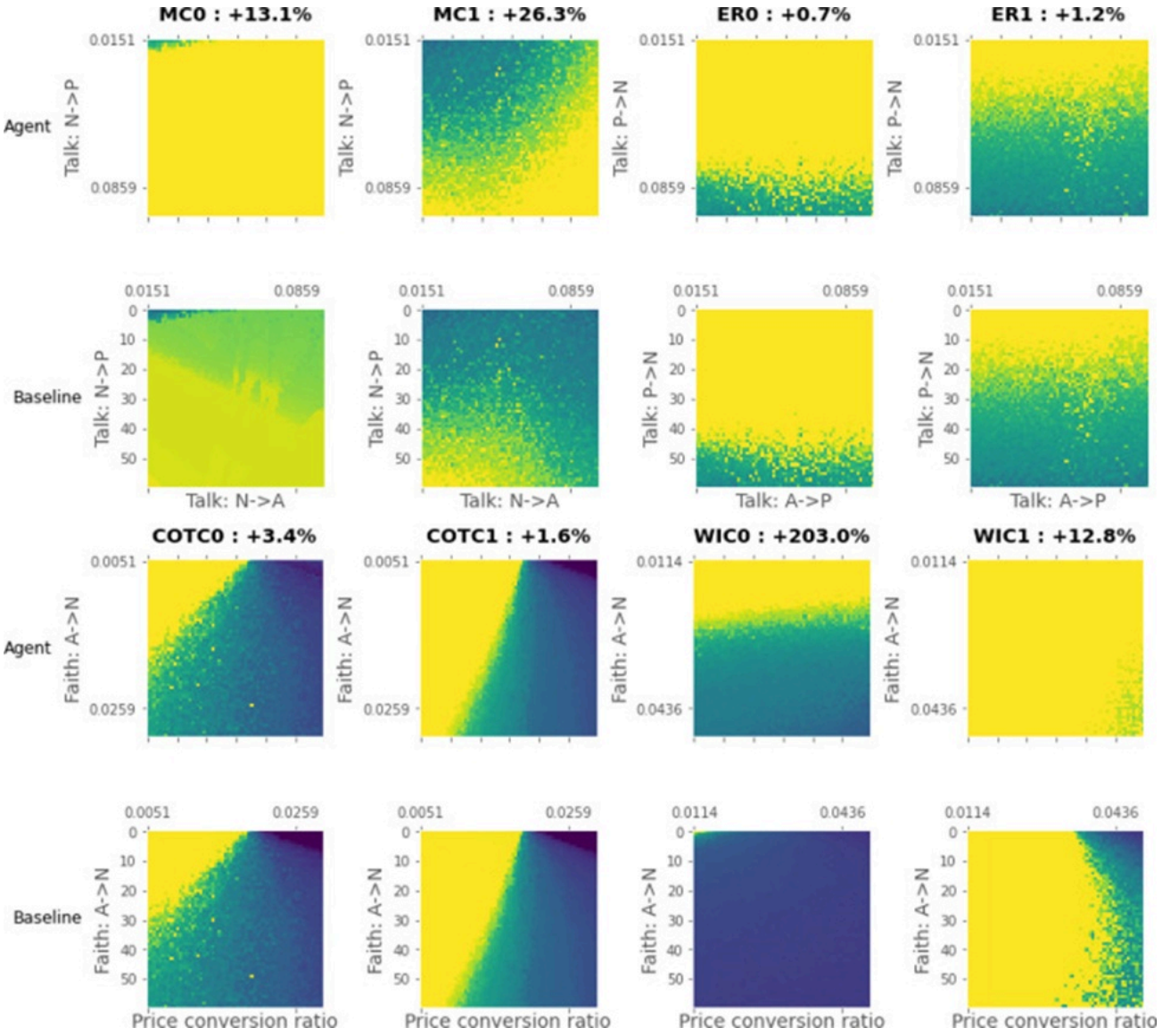


Figure 7. Comparison of state space policy scores for multiple scenarios (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Specific case study

As established before, utilizing policies developed through reinforcement learning is beneficial across diverse scenarios, as such techniques considerably expand the sustainable area within the state space. Following this validation, this section introduces several practical applications for the developed simulation framework. Although it can be applied to any scenario, the COTC0 scenario is selected for this analysis. Its highly contrasting characteristics help to elucidate the underlying logic more clearly. Specifically, this scenario examines sensitivity concerning both the price conversion threshold and the faith resistance dimensions. As observed, the assessed parameter range produces both highly sustainable states (yellow) and highly energy-intensive configurations (deep blue) **(Figure 8).**

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

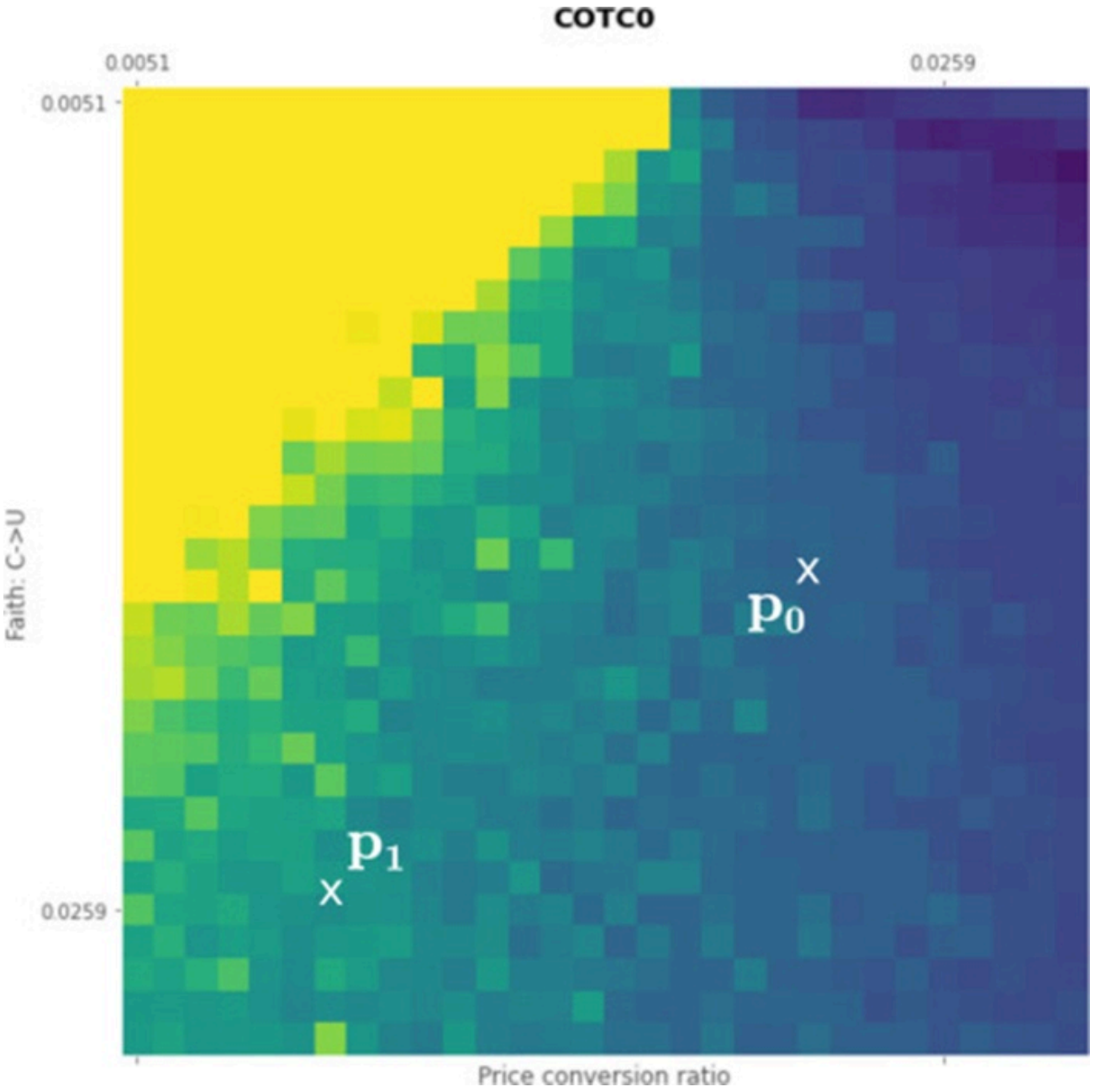


Figure 8. Heatmap for the social tensions (COTC0) scenario. Specific starting points for subsequent policy tests are represented in white (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Subsidizing energy price

Let us consider a demographic group p_0, which has been mapped (e.g., via polling data) to location 1. This specific group possesses a moderate level of conviction but exhibits higher reactivity to price fluctuations than the general public. It is quantitatively represented as p_0 = [0.5, 0.7] across the respective axis limits. Even with the performance enhancement afforded by the RL-based storage mechanism (+11%), this setup still falls outside sustainable bounds.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Assuming no supplementary energy sources are accessible, several strategic responses remain viable. A critical factor among these is the price conversion threshold, which heavily influences personal energy consumption choices. Within this context, municipal authorities might offer financial subsidies tied to this variable to promote sustainability. For example, **Figure 9** illustrates the resultant growth of highly energy-sober populations and their corresponding configuration scores when the price conversion threshold receives financial subsidization. In detail, while a baseline conversion price of 40 euros generates a sustainability score that mandates reliance on external energy grids, amplifying public financial support rapidly bolsters the environmentally responsible cohort, thereby alleviating strain on the overall energy infrastructure.

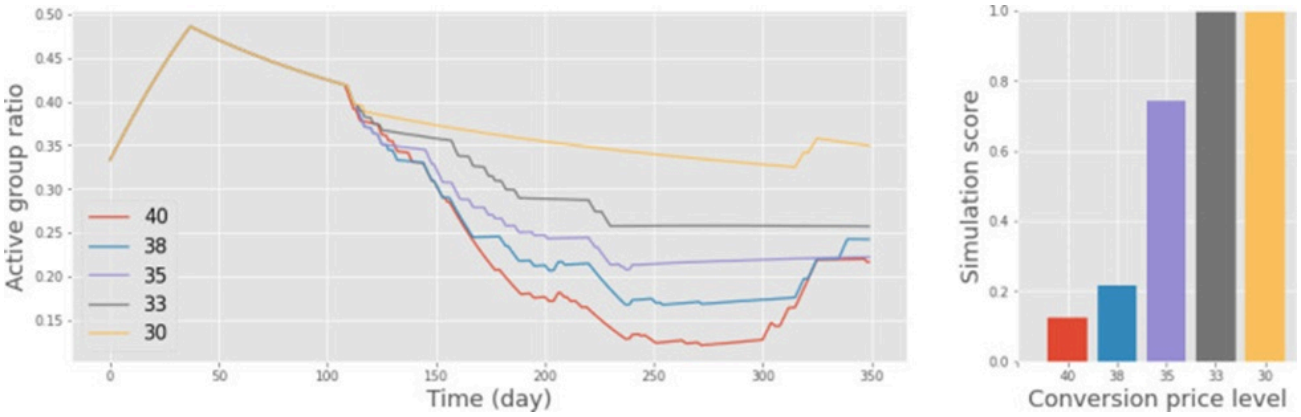


Figure 9. Evaluation of sustainability levels resulting from a subsidiary policy to financially support individuals by offsetting their price conversion threshold.

Energy tax usage and magnitude

Alternatively, regional governments might implement taxation mechanisms to persuade individuals exhibiting the most intensive energy consumption patterns to align their habits with available energy supplies. Utilizing the identical initial coordinate p_0, this section investigates the impacts of decreasing the upper price conversion threshold. Functionally, this represents the price point at which users begin transitioning into demographics characterized by reduced energy usage. Because taxation frameworks are inherently intricate and fall beyond the purview of the present study, **Figure 10** employs a streamlined model to depict only the ultimate price boundary.

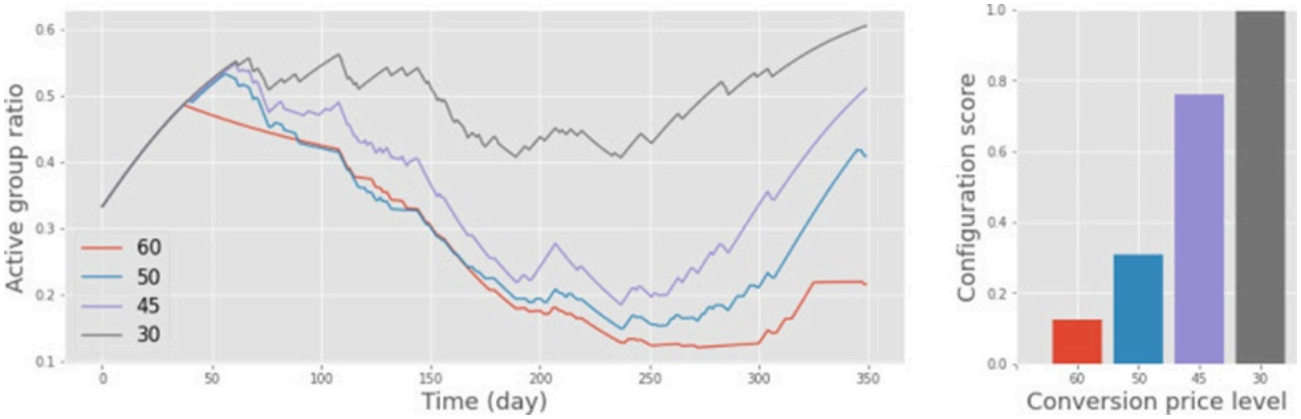


Figure 10. Evaluation of sustainability levels resulting from a tax policy to financially incentivize individuals to lessen their consumption by artificially increasing their price conversion threshold to more sober groups.

Local exploration

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Let us examine an alternative demographic setup, p_1= [0.8, 0.18], which demonstrates greater resilience or inertia against rapid price fluctuations but exhibits a heightened deterioration of belief and commitment. Here, individuals from group A are heavily assimilated into the N cohort, meaning that members who were previously energy-conscious shift toward more intensive consumption habits.

Although the most sustainable region for this specific scenario and its corresponding parameter set resides in the upper-west quadrant, isolated pockets of sustainability are evident across various areas within the state space. This subsection specifically highlights a high-scoring cluster that maintains an identical price sensitivity but features a distinct faith resistance metric.

Specifically, analyzing setups featuring minimal faith resistance alongside adjacent faith metrics ranging between 0.18 and 0.19 reveals the population behaviors mapped in **Figure 11**. Several of these evaluated points yield substantial sustainability scores. Generally speaking, when originating from p_1, designing governmental initiatives that gently shift the demographic toward these localized zones might be more economically and energetically viable than attempting to force a transition to the uppermost sustainable quadrant.

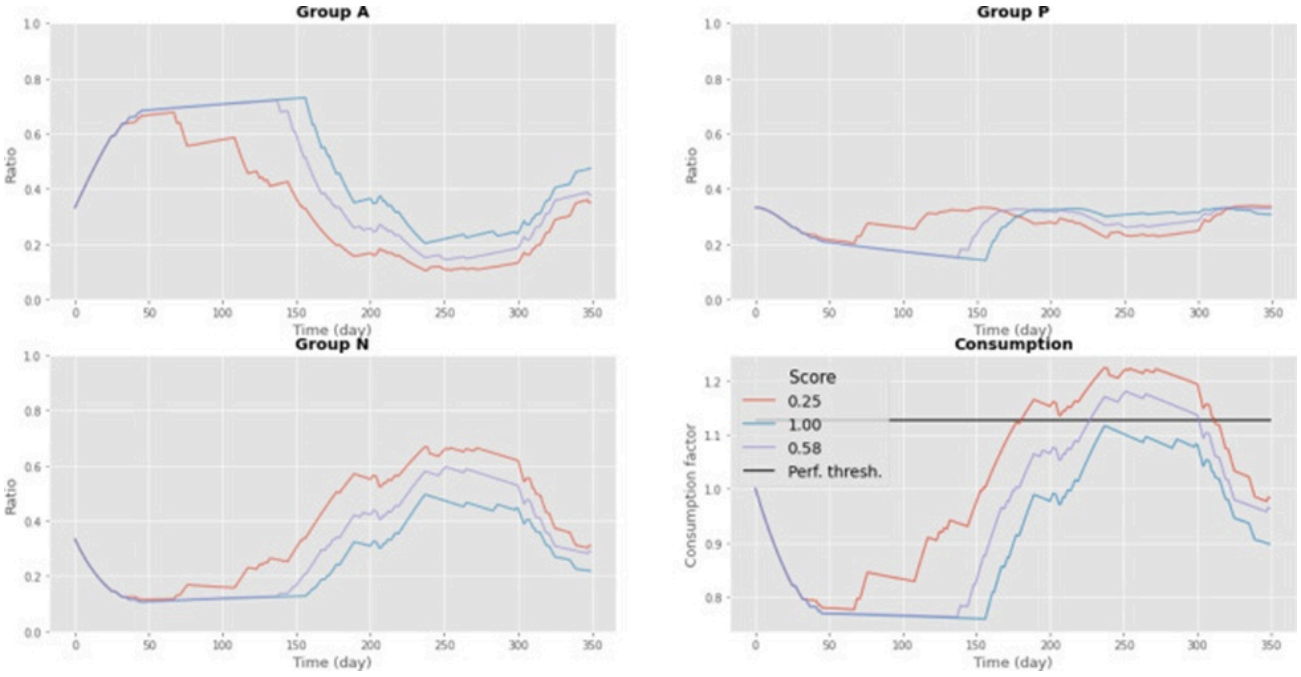


Figure 11. Heatmap for the clash of the classes scenario. Specific starting points for subsequent policy tests are represented in white.

Nevertheless, as further illustrated by **Figure 11**, the price sensitivity axis exhibits significant score volatility. This instability could render such a strategy infeasible in practical applications, where data is typically aggregated and where societal rigidities and behavioral inertia prevent pinpoint precision. This phenomenon is additionally demonstrated in **Figure 12**, which examines the entire spectrum of faith parameters. In this analysis, the full range of conviction is assessed against price sensitivities strictly confined to the [0.18, 0.20] bracket. Although the shaded region—which portrays the maximum and minimum outcomes for every coordinate—features prominent spikes indicating zones of high sustainability, the average score suggests that achieving a stable equilibrium is difficult, unlike scenarios characterized by more robust faith metrics.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

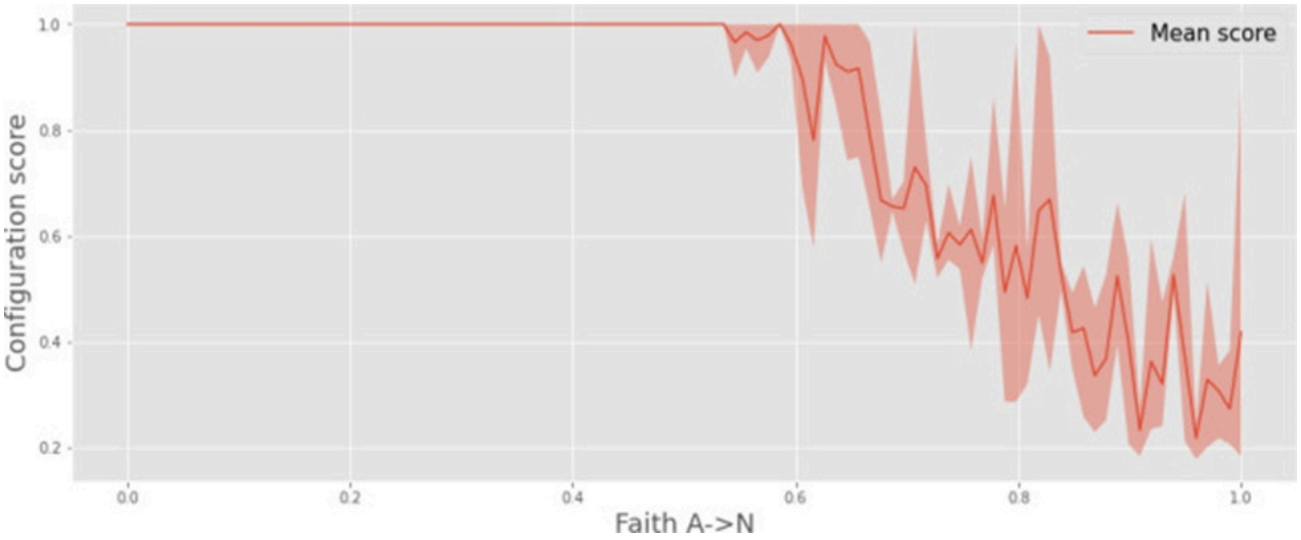


Figure 12. Price sensitivity between 0.18 and 0.20.

As detailed in the preceding sections, the introduced forward-looking model yields valuable insights regarding the imbalance between electrical generation and consumption. It also estimates the annualized systemic performance anticipated for a specific demographic, conditional upon the chosen energy storage architecture and its governing protocol. Although the currently used parameter boundaries were chosen primarily to facilitate theoretical exploration, applying precise empirical data to the quantitative strategic comparisons could yield profound practical consequences. The assessment of multiple governance structures clearly indicates that RL-driven models reliably expand the boundaries of sustainability; this equates to state space arrangements where aggregate energy demands remain harmonized with the infrastructure's operational capacity. Consequently, keeping the public within the limits of sustainable renewable energy utilization will necessitate fewer severe civic regulations, thereby proportionally increasing the likelihood of widespread societal acceptance.

However, under the present formulation, systemic efficiency is calculated as an average across the entire temporal duration, a method that might be sub-optimal. This approach intrinsically assumes that scenario variables—including individual ideological resolve or attitudes regarding energy usage—remain completely impervious to the environmental or systemic realities encountered throughout the year. Realistically, if operational conditions deteriorate, the demographic parameters are highly likely to shift. In practical terms, should an initial configuration subject the public to prolonged periods of energy scarcity, personal convictions or susceptibility to alternative behavioral paradigms would inevitably adapt. Mathematically, this phenomenon translates to second-order dynamics, a subject slated for investigation in forthcoming research.

Conclusion

As global demand for accessible, clean, and renewable power accelerates, the natural intermittency of these resources becomes increasingly evident, heavily underlining the vital importance of environmentally responsible and conservative individual habits. Even as such sustainable practices become more prevalent and normalized among the public, an extensive array of behavioral variables remains, all of which must be carefully integrated when designing comprehensive energy frameworks. In response, this study—having validated the applicability of reinforcement learning to the administration of energy reserves—introduces a novel, adaptable, and modular architecture. By exploiting RL-enhanced storage protocols, this tool facilitates the simulation of diverse hypotheses and the rigorous evaluation of their corresponding outcomes. By bridging these distinct domains, the present research diverges from conventional methodologies that typically isolate such subjects, thereby preventing the loss of synergistic analytical perspectives.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

As evidenced by the provided case studies, intricate behavioral dynamics can be seamlessly incorporated and modeled. Analyzing the results of these diverse setups empowers policymakers to pinpoint specific viability thresholds for any chosen parameter, allowing them to avoid critical zones and proactively mitigate severe socioeconomic repercussions. Additionally, by contrasting various storage strategies across a broad array of scenarios, this investigation quantitatively maps the expansion of sustainable operational zones. This directly correlates with reducing the stringency of societal interventions required to keep renewable energy consumption within manageable limits. Furthermore, the proposed architecture has proven effective in appraising the ramifications of civic interventions, including the issuance of financial subsidies to encourage eco-friendly habits. Conversely, the behavioral shifts induced by taxation can also be modeled to a meaningful degree. As illustrated, adopting a comprehensive perspective on state-space configuration scores facilitates pragmatic governance, highlighting stable equilibria that are less complex to achieve.

In subsequent research, these promising findings will be expanded across several dimensions. First, integrating a continuous feedback and behavioral adaptation loop is a critical enhancement for the model's overall fidelity. Theoretically, this upgrade will introduce second-order dynamics into the demographic simulations, enabling synthetic individuals to behave more rationally by directly accommodating historical memory effects. Next, because the primary intent of this framework is to aid in the formulation of public policy, a significant advancement would involve refining the current case studies to include a heuristic search methodology. This would calculate the most economically efficient civic pathway required to shift a precarious state space coordinate back into the sustainable zone. By assigning specific socioeconomic transition costs to movements within the state space, analysts can determine the optimal levers to effectively maneuver a population from unacceptable thresholds toward highly reliable operational states, tailored to the extant storage capacity and renewable generation profile. By integrating authentic empirical datasets into this paradigm, it could ultimately become a premier asset in policy architecture, smoothing the public reception of complex energy transitions. Finally, to satisfy the operational demands of real-world deployment, extensive supplementary data collection is imperative. This will ensure precise model calibration, establish fully realistic boundary parameters, and broaden the simulation's scope by incorporating variables such as heterogeneous starting demographic distributions.

Acknowledgements

None

Conflict of interest

None

Financial support

Akkodis and the Adecco group, France, have sponsored this work.

Ethics statement

None

Received: 09 Oct 2023 Revised: 15 Dec 2023 Accepted: 02 Jan 2024
Published online: 15 June 2025

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Arias P, Bellouin N, Coppola E, Jones R, Krinner G, Marotzke J, et al. Technical summary. In: Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2021. p. 33-144.

2. Canadell J, Monteiro P, Costa M, Cotrim da Cunha L, Cox P, Eliseev A, et al. Global carbon and other biogeochemical cycles and feedbacks. In: Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2021. p. 673-816.

3. Canadell J, Monteiro P, Costa M, Cotrim da Cunha L, Cox P, Eliseev A, et al. Global carbon and other biogeochemical cycles and feedbacks supplementary material. In: Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change; 2021. Available from: https://www.ipcc.ch/.

4. Climate change 2022: impacts, adaptation and vulnerability. Contribution of working group II to the sixth assessment report of the intergovernmental panel on climate change. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2022.

5. Climate change 2022: mitigation of climate change. Contribution of working group III to the sixth assessment report of the intergovernmental panel on climate change. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2022.

6. Reichl J, Cohen JJ, Klöckner CA, Kollmann A, Azarova V. The drivers of individual climate actions in Europe. Glob Environ Change. 2021;71:102390.

7. Akerlof KL, Boules C, Ban Rohring E, Rohring B, Kappalman S. Governmental communication of climate change risk and efficacy: moving audiences toward "danger control". Environ Manag. 2020;65(5):678-88.

8. Boyce C, Czajkowski M, Hanley N. Personality and economic choices. J Environ Econ Manag. 2019;94:82-100.

9. Cohen J, Moeltner K, Reichl J, Schmidthaler M. Effect of global warming on willingness to pay for uninterrupted electricity supply in European nations. Nat Energy. 2018;3:37-45.

10. Prague M, Wittkop L, Clairon Q, Dutartre D, Thiébaut R, Hejblum BP. Population modeling of early COVID-19 epidemic dynamics in French regions and estimation of the lockdown impact on infection rate. 2020. Available from: https://hal.archives-ouvertes.fr/hal-02555100.

11. Cyrus Chu C. Population dynamics: theory of nonstable populations. In: Smelser NJ, Baltes PB, editors. International encyclopedia of the social & behavioral sciences. Oxford: Pergamon; 2001. p. 11771-3.

12. Shorrocks B. Competition, interspecific. In: Levin SA, editor. Encyclopedia of biodiversity. 2nd ed. Waltham: Academic Press; 2001. p. 177-91.

13. Mehdi M, Fabien M, Yassine M, Guilhem M, Enzo M, Dorian F. S2SAME: A multi-modal deep learning system model for automated renewable energy trading. In: 3rd International Conference on Energy and AI; 2022.

14. Levent T, Preux P, le Pennec E, Badosa J, Henri G, Bonnassieux Y. Energy management for microgrids: a reinforcement learning approach. In: 2019 IEEE PES Innovative Smart Grid Technologies Europe; 2019. p. 1-5.

15. Qing-Hua W, Jiehui Z, Zhaoxia J, Xiaoxin Z. Large-scale integrated energy systems. Springer Link; 2016.

16. Zia MF, Elbouchikhi E, Benbouzid M. Microgrids energy management systems: A critical review on methods, solutions, and prospects. Appl Energy. 2018;222:1033-55.

17. Keirstead J, Jennings M, Sivakumar A. A review of urban energy system models: Approaches, challenges and opportunities. Renew Sustain Energy Rev. 2012;16:3847-66.

18. Nigim KA, Lee W-J. Micro grid integration opportunities and challenges. In: 2007 IEEE Power Engineering Society General Meeting; 2007. p. 1-6.

19. Parhizi S, Lotfi H, Khodaei A, Bahramirad S. State of the art in research on microgrids: a review. IEEE Access. 2015;3:1.

20. Su W, Wang J. Energy management systems in microgrid operations. Electr J. 2012;25(8):45-60.

21. Palma-Behnke R, Benavides C, Lanas F, Severino B, Reyes L, Llanos J, et al. A microgrid energy management system based on the rolling horizon strategy. IEEE Trans Smart Grid. 2013;4(2):996-1006.

22. Ahmad Khan A, Naeem M, Iqbal M, Qaisar S, Anpalagan A. A compendium of optimization objectives, constraints, tools and algorithms for energy management in microgrids. Renew Sustain Energy Rev. 2016;58:1664-83.

23. Guo L, Liu W, Li X, Liu Y, Jiao B, Wang W, et al. Energy management system for stand-alone wind-powered-desalination microgrid. IEEE Trans Smart Grid. 2016;7(2):1079-87.

24. Kyriakarakos G, Dounis A, Arvanitis K, Papadakis G. A fuzzy logic energy management system for polygeneration microgrids. Renew Energy. 2012;41:315-27.

25. Taha MS, Mohamed YA-RI. Robust MPC-based energy management system of a hybrid energy source for remote communities. In: 2016 IEEE Electrical Power and Energy Conference; 2016. p. 1-6.

26. Eghtedarpour N, Farjah E. Power control and management in a hybrid AC/DC microgrid. IEEE Trans Smart Grid. 2014;5(3):1494-505.

27. Mokheimer EM, Sahin AZ, Al-Sharafi A, Ali AI. Modeling and optimization of hybrid wind–solar-powered reverse osmosis water desalination system in Saudi Arabia. Energy Convers Manage. 2013;75:86-97.

28. Artzrouni M. Mathematical demography. In: Kempf-Leonard K, editor. Encyclopedia of social measurement. New York: Elsevier; 2005. p. 641-51.

29. Brander JA. Easter island: resource depletion and collapse. In: Cleveland CJ, editor. Encyclopedia of energy. New York: Elsevier; 2004. p. 871-80.

30. Hellmann JJ. Species interactions. In: Levin SA, editor. Encyclopedia of biodiversity. 2nd ed. Waltham: Academic Press; 2013. p. 715-25.

31. Estes J, Crooks K, Holt RD. Predators, ecological role of. In: Levin SA, editor. Encyclopedia of biodiversity. 2nd ed. Waltham: Academic Press; 2013. p. 229-49.

32. Chan T-L, Yuan H-Y, Lo W-C. Modeling COVID-19 transmission dynamics with self-learning population behavioral change. Front Public Health. 2021;9.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

33. Eshragh A, Alizamir S, Howley P, Stojanovski E. Modeling the dynamics of the COVID-19 population in Australia: A probabilistic analysis. PLoS One. 2020;15:e0240153.

34. Wilder B, Charpignon M, Killian JA, Ou H-C, Mate A, Jabbari S, et al. Modeling between-population variation in COVID-19 dynamics in Hubei, Lombardy, and New York City. Proc Natl Acad Sci. 2020;117(41):25904-10.

35. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, editors. Advances in neural information processing systems. Vol. 28. Curran Associates, Inc.; 2015. p. 91-9.

36. Jumper JM, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583-9.

37. Silver D, Huang A, Maddison C, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of go with deep neural networks and tree search. Nature. 2015.

38. Wei J, Wang X, Schuurmans D, Bosma M, Chi EH, Le Q, et al. Chain of thought prompting elicits reasoning in large language models. 2022. arXiv:2201.11903. Available from: https://arxiv.org/abs/2201.11903.

39. Rubin O, Herzig J, Berant J. Learning to retrieve prompts for in-context learning. 2021. arXiv:2112.08633. Available from: https://arxiv.org/abs/2112.08633.

40. Degrave J, Felici F, Buchli J, Neunert M, Tracey B, Carpanese F, et al. Magnetic control of tokamak plasmas through deep reinforcement learning. Nature. 2022;22(1):79-86.

41. OpenAI J, Akkaya I, Andrychowicz M, Chociej M, Litwin M, McGrew B, et al. Solving rubik's cube with a robot hand. 2019.

42. Sahni H. Deep reinforcement learning doesn't work yet. 2018. Available from: https://himanshusahni.github.io/2018/02/23/reinforcement-learning-never-worked.html.

43. Eysenbach B, Gupta A, Ibarz J, Levine S. Diversity is All You Need: Learning Skills without a Reward Function. In: International Conference on Representation Learning; 2019.

44. Mahajan D, Tople S, Sharma A. Domain generalization using causal matching. In: Meila M, Zhang T, editors. Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research. Vol. 139. PMLR; 2021. p. 7313-24.

45. Kaddour J, Lynch A, Liu Q, Kusner MJ, Silva R. Causal machine learning: a survey and open problems. 2022. arXiv:2206.15475. Available from: https://arxiv.org/abs/2206.15475.

46. Chen AS, Sharma A, Levine S, Finn C. You only live once: single-life reinforcement learning. 2022. arXiv:2210.08863. Available from: https://arxiv.org/abs/2210.08863.

47. Yu L, Xie W, Xie D, Zou Y, Zhang D, Sun Z, et al. Deep reinforcement learning for smart home energy management. IEEE Internet Things J. 2020;7(4):2751-62.

48. Sutton RS, Barto A. Reinforcement learning: an introduction. MIT Press; 2018.

49. Haarnoja T, Zhou A, Abbeel P, Levine S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In: International Conference on Machine Learning (ICML); 2018.

50. Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. ICLR (Poster). 2016. Available from: http://arxiv.org/abs/1509.02971.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

51. Fujimoto S, van Hoof H, Meger D. Addressing function approximation error in actor-critic methods. 2018. arXiv:1802.09477. Available from: http://arxiv.org/abs/1802.09477.

52. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. 2017. arXiv:1707.06347. Available from: http://arxiv.org/abs/1707.06347.

53. Beckley R, Weatherspoon C, Alexander M, Chandler M, Johnson A, Bhatt GS. Modeling epidemics with differential equations. 2013.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press