Design and planning of flexible mobile Micro-Grids using Deep Reinforcement Learning

Cesare Caputo [a], Michel-Alexandre Cardin [a,*], Pudong Ge [b], Fei Teng [b], Anna Korre [c], Ehecatl Antonio del Rio Chanona [d]

[a] Dyson School of Design Engineering, Imperial College London, United Kingdom
[b] Department of Electrical and Electronics Engineering, Imperial College London, United Kingdom
[c] Department of Earth Science and Engineering, Imperial College London, United Kingdom
[d] Department of Chemical Engineering, Imperial College London, United Kingdom

HIGHLIGHTS

- The impacts of climate change on global migratory patterns must be addressed in light of their energy access needs.
- Plug and Play control developments allow nomadic communities to operate in stand-alone mode or connect to the main grid.
- Flexibility in Design and Deep Reinforcement Learning are implemented for their energy systems planning under uncertainty.
- The proposed method outperforms baseline systems on expected cost, equivalent emissions and unmet load over 30 years.
- Results have important implications for policy, design, planning, and adaptable operations of future mobile energy systems.

ARTICLE INFO

ABSTRACT

Ongoing risks from climate change have significantly impacted the livelihood of global nomadic communities and are likely to lead to increased migratory movements in coming years. As a result, mobility considerations are becoming increasingly important in energy systems planning, particularly to achieve energy access in developing countries. Advanced "Plug and Play" control strategies have been recently developed with such a decentralized framework in mind, allowing easier interconnection of nomadic communities, both to each other and to the main grid. Considering the above, the design and planning strategy of a mobile multi-energy supply system for a nomadic community is investigated in this work. Motivated by the scale and dimensionality of the associated uncertainties, impacting all major design and decision variables over the 30-year planning horizon, Deep Reinforcement Learning (DRL) Flexibility Analysis is implemented for the design and planning problem. DRL based solutions are benchmarked against several rigid baseline design options to compare expected performance under uncertainty. The results on a case study for ger communities in Mongolia suggest that mobile nomadic energy systems can be both technically and economically feasible, particularly when considering flexibility, although the degree of spatial dispersion among households is an important limiting factor. Additionally, the DRL based policies lead to the development of dynamic evolution and adaptability strategies, which can be used by the targeted communities under a very wide range of potential scenarios. Key economic, sustainability and resilience indicators such as Cost, Equivalent Emissions and Total Unmet Load are measured, suggesting potential improvements compared to available baselines of up to 25%, 67% and 76%, respectively. Finally, the decomposition of values of flexibility and plug and play operation is presented using a variation of real options theory, with important implications for both nomadic communities and policymakers focused on enabling their energy access.

---

* Corresponding author at: Imperial College London, Dyson School of Design Engineering, Royal College Observatory Building, Room 1M03, SW7 2AZ, United Kingdom.
E-mail address: m.cardin@imperial.ac.uk (M.-A. Cardin).

Nomenclature

Acronyms

ACER	Actor Critic with Prioritized Experience Replay
BD1/2/3/4	Baseline Rigid Design 1/2/3/4
BESS	Battery Energy Storage System
DGs	Distributed Generators
DNNs	Deep Neural Networks
DRL	Deep Reinforcement Learning
EH	Electric Heating/Heater
ENPC	Expected Net Present Cost
EoS	Economies of Scale
FD	Flexible Design
FND	Flexible Nanogrid Design
GA	Genetic Algorithm
GBM	Geometric Brownian Motion
LFS	Load Following Strategy
MDP	Markov Decision Process
ML	Machine Learning
MS	Mobility Scenario
MSSP	Multi-Stage Stochastic Programming
NPC	Net Present Cost
PP	Plug and Play
REAP	Rural Electricity Access Project
RES	Renewable Energy System
SDG	Sustainable Development Goal
SHS	Small Home Solar
SWS	Small Wind Systems
TUL	Total Unmet Load
UB	Ulaanbaatar

Indices and Sets

$t \in T$	Index and set of *general* time steps
$m \in M$	Index and set of *monthly* decision time steps
$h \in H$	Index and set of operational simulation time steps
$s \in S$	Index and set of scenarios
$i \in I$	Index and set of technological components
$x \in X$	Index and set of GBM modeled variables
$n \in N$	Index and set of nanogrids/nodes in PP microgrid

Model Variables and Parameters

T	Evaluation horizon (months)
ξ	Uncertainty vector
θ	Installed nominal capacity (kW or kWh)
λ	Financial discount rate
$\mathscr{R}$	Revenue Function
$\mathscr{C}$	Cost Function
$\mathscr{H}$	Expansion Cost Function
$CAPEX$	Initial Capital Investment
NCF	Net cash flow (USD)
NPC	Net present cost (USD)
$ENPC_{es}$	Energy system ENPC (USD)
C	Energy system cost (USD)
s	State vector
a	Action vector
A	Action space
R	Reward
γ	DRL discount rate
π	DRL agent policy
VoF	Value of flexibility

Model Variables and Parameters (continued)

$P_{ul,es}$	Unmet total load for *PP microgrid* (kWh)
$P_{dist,es}$	Total distributed electricity in *PP microgrid* (kWh)
$\theta_{PP,es}$	Total *PP microgrid* distribution capacity limits (kW)

$P_{dist,es,max}$	Constraint on maximum *PP microgrid* distribution (% of ED)
P_{dist}	Active power distribution requirements per *nanogrid* (kW)
L_{dist}	Distance distribution requirements per *nanogrid* (kW)
R_{dist}	Distribution resistance requirements per *nanogrid* (Ω)
V_{dist}	Voltage distribution requirements per *nanogrid* (V)
I_{dist}	Current distribution requirements per *nanogrid* (amp)
α_{cb}	Material temperature coefficient
$P_{dl,es}$	Active power distribution losses for *PP microgrid* (kW)
DL	Total distribution loss %
P_{grid}	Active exchanged with main grid for *PP microgrid* (kW)
θ_{grid}	Total *main grid* distribution capacity limits (kW)
ρ_{grid}	Binary variable on grid availability
$P_{uhl,es,ppp}$	Unmet heating load after *PP microgrid* distribution (kW)
$P_{res,exc,pp}$	Excess electricity after *PP microgrid* distribution (kW)
$P_{coal,es}$	Power coal requirements for *PP microgrid* (kW)
M_{coal}	Coal mass requirements for *PP microgrid* (kg)
$\mu_{localcoal}$	Local lignite coal energy density
$\eta_{coalstove}$	Coal stove net efficiency (%)
$CO_{2eq.emissions,es}$	Equivalent CO_2 emissions generated (tonnes)
EF	Weighted average emission factor
C_{inv}	Total investment costs (USD)
C_{opex}	Operational costs (USD)
C_{grid}	Grid active power exchange costs or revenue (USD)
C_{CO2}	Carbon emissions equivalent costs or revenue (USD)
$C_{es,max}$	Budget constraint (USD)
CC	Communication costs (USD)
EC	Expansion costs (USD)
RC	Replacement costs (USD)
IC	Infrastructure costs (USD)
PEC	Power electronics costs (USD)
c_{cc}	Communication costs factor (USD/ kWh)
c_{ic}	Cabling costs factor (USD/ km)
c_{pe}	Power electronics cost factor (USD/ kW)
K	Investment constant in EoS model
α	EoS factor in EoS model
ε_{pe}	Power electronics requirement ratio
sf_{pe}	Safety factor
M_{batt}	Battery total weight (kg)
M_{pe}	Power electronics total weight (kg)
M_{cb}	Distribution cabling total weight (kg)
EV	Expected Value
$VoPP$	Value of Plug and Play Operation
μ	Drift
σ	Volatility
W	Weiner process stochastic deviation
N_n	Number of community dwellings
$r_{cluster}$	Nomadic community cluster radius (km)
L_{cb}	Cabling length (km)
P_{wind}	Active Wind power (kW)
P_{pv}	Active PV power (kW)
P_{NEL}	Net electricity load (kW)
P_{NHL}	Net heating load (kW)
ED	Electricity demand (kW)
HD	Heating demand (kW)
P_{res}	RES active power (kW)
$P_{eh,max}$	Maximum EH capacity (kW)
η_{inv}	Inverter roundtrip efficiency (%)
$P_{ba,in,max}$	Maximum battery charging capacity (kW)
$P_{ba,out,max}$	Maximum battery discharging capacity (kW)
E_{ba}	Battery storage level (kWh)
$E_{ba,min}$	Minimum battery storage level (kWh)
$E_{ba,max}$	Maximum battery storage level (kWh)

η_{batt}	Battery roundtrip efficiency (%)	c_{coal}	Coal cost factor (USD/kg)
$P_{res,PL}$	Available RES capacity post meeting nanogrid load (kW)	C_{om}	Total O&M costs (USD)
$N_{deficit}$	Nanogrid units with electricity deficit	c_{om}	O&M cost factor (USD/kW)
$N_{surplus}$	Nanogrid units with electricity surplus	C_{ul}	Total unmet load costs (USD)
$P_{eg.es}$	Excess *PP microgrid* electricity generation (kWh)	c_{ul}	Unmet load cost factor (USD/kWh)
L	Functional operational life (years)	C_{grid}	Total grid interaction costs (USD)
SV	Salvage value (USD)	c_{grid}	Grid tariff (USD/kWh)
θ_{sv}	Abandoned capacity (kW)	p_{grid}	Grid RES subsidy (USD/kWh)
k	Depreciation factor	M_{RES}	RES total weight (kg)
C_{coal}	Total coal costs (USD)		

1. Introduction

1.1. Climate change and migration

The climate crisis experienced in recent years is inducing significant changes in temperature, precipitation, and extreme weather event patterns [1]. This is impacting the livelihood of nomadic communities, as it is highly dependent on local environmental conditions in areas unsuitable for sedentary agriculture [2]. As such, most of them are dependent on following an approximately fixed seasonal pattern of movements to sustain their agricultural activities, transporting some form of mobile dwelling. It is estimated that there are currently a least 30–40 million rural nomads, primarily found across Sub-Saharan Africa, South-Asia, and Central America [3]. On the other hand, areas where sedentary agriculture can be sustainably practiced are rapidly decreasing, leading to climate-change induced mobility worldwide [4]. Recent models suggest that around 143 million more people will be displaced by 2050, likely impacting vulnerable communities the most [5,6].

Climate driven migration is also recognized as one of the major causes of urbanisation worldwide [7]. Urban migrants are normally composed primarily of unskilled workers associated with the informal sector, where the lack of proper steady housing has limited their integration into the formal sector. They typically create informal settlements on the outskirts of major cities, housed in small tent-like structures. As such, they are normally unable to access services such as electricity or district heating. The resulting energy poverty impacts the livelihood of urban migrants significantly, having to spend long periods of time and a high proportion of their disposable income in the search for more rudimentary power sources, such as dung or biomass [8]. Furthermore, it was estimated that globally there are at least 208 million urban migrants completely lacking access to energy, primarily due to infrastructure limitations or lack of access to the formal government sector [7,9].

While the UN Sustainable Development Goals (SDGs) do not explicitly link climate change and migration, it is becoming increasingly important to develop policies and technologies which help mitigate the impacts of climate induced mobility [10]. Energy access is very complex in this context, as the current infrastructure is not compatible with such mass migratory movements, particularly in lower income countries [11]. In order to address SDG 7 (*Affordable and Clean Energy for All*) it is, therefore, imperative to develop energy system solutions which are compatible with a migratory lifestyle, as opposed to the "static" approach used in energy systems planning thus far [12].

1.2. Need for Plug and Play based flexible and sustainable energy systems

1.2.1. Overview

Given the current high costs for national grid extensions, mini-grid and micro-grid solutions are expected to be the main driver of electrification for almost half of all households currently lacking energy access [13]. The rapid cost decline for renewable energy systems (RES) over the past few years has already increased their rate of deployment, providing electricity to 140 million people as of 2020 [9]. Similarly, there are ongoing trends towards more distributed generators (DG) as RES are increasingly installed closer to their respective loads, helping reduce transmission power losses. The uncertainties and intermittence of RES generation, however, lead to several technical challenges for their integration into the main grid. At higher levels of RES penetration in particular, maintaining operational parameters such as voltage and frequency within their allowable limits can become increasingly problematic [14].

Nomadism represents a more extreme example of the above issues, as both loads, and generation are not only distributed but also changing over different temporal and spatial scales. Simultaneously, remote working policies implemented as a result of the COVID-19 pandemic have led to a global increase in "digital nomads" [15]. As a lot of these individuals move around in campers or recreational vehicles, they contribute to increasingly decentralized and spatially unpredictable electricity generation and demand. Based on the above and the trends mentioned in Section 1.1, there is a growing need to develop flexible and mobile energy systems - which are able to switch operation between a centralized and decentralized manner to accommodate both rural and urban migration.

Recent advancements in distributed control algorithms have enabled the consideration of mobile and decentralized energy system solutions in the context with a migratory lifestyle [16,17]. In most regions currently practicing nomadism, in fact, family clusters tend to be relatively close to each other, following similar seasonal migration patterns [3]. This creates the possibility for temporary interconnection among them to form a mobile herder household energy supply system (i.e. *micro-grid*), or allowing flexible plug-in operations with the main grid. This sort of plug and play (PP) operation can be highly advantageous as it allows the smoothing out of load/generation imbalances among various households, with potentially different electricity generation and storage capacities. Plug-in operations with the main grid, on the other hand, could be beneficial by providing an additional energy supply to nomadic communities when available, while helping improve main grid stability through a more distributed storage capacity [18]. Additionally, market engagement opportunities may be generated, potentially helping to reduce the total electrification costs for nomadic communities.

Moreover, PP holds potential to accelerate the decarbonization of heating supply systems for nomadic communities living in colder regions. Their traditional heating systems rely on inefficient combustion of fossil fuels and could be replaced by electric-heating (EH) technologies if a reliable and sufficient level of sustainable electrification can be achieved. This may be particularly attractive if implemented at a community rather than single household level, as a higher degree of modularity can be integrated in system capacity evolution strategies [7]. Currently, however, all PP studies have focused on fine timestep resolution control operations to guarantee network reliability. Furthermore, none have analysed it through a multi-energy vector perspective to accelerate the transition to sustainable heating. Given the potential long-term implications of PP as climate induced migration increases, a novel modelling and planning framework which integrates mobility considerations is needed to quantify its long-term impact, and develop policy recommendations for both rural and urban migrants.

1.2.2. Plug and Play (PP) operation

PP operation can be allowed without control law changes in the power-electronic-level devices, guaranteeing system stability and synchronisation (in terms of frequency and voltage), as detailed in Section 2.1.3. As such, it is compatible with operational control strategies including predictive Machine Learning (ML) methods, advanced control algorithms, or heuristic-based approaches [14,19–23]. Nonetheless, recent studies suggest that for smaller scale systems with no dispatchable generator, a simple Load-Following Strategy (LFS) is likely to result in similar conclusions on initial design capacities and total costs as more complex methods [23–25]. Its logic prioritises renewable generation, followed by battery energy storage and grid interactions, if available, to meet any remaining load/demand – adopting simple heuristics to operate the system. Accordingly, it is well suited to the technological and communication capacity limitations associated with a nomadic mobile energy supply system. Based on the more limited control opportunities available, a flexible capacity planning strategy under uncertainty is, therefore, more likely to impact the distribution of total system costs (as well as other performance metrics) obtained through long-term PP operation.

1.2.3. Energy system planning and role of flexibility

General micro-grid planning over a long horizon is a complex problem that has been extensively investigated in the literature, and is closely related to rural electrification objectives [26–29]. The aim is typically minimisation of Net Present Cost (NPC) although there are a few works which have more closely looked into sustainability [30,31] and resilience [17,32,33] considerations too. Traditionally, to help reduce the computational expense of the optimisation, planning problems can be simplified to single-year approaches, where the performance of the system over the entire planning horizon is determined based on the simulation of only one "representative" year [34]. While the computational simplifications are substantial, this kind of approach does not account for load growth or corresponding energy system evolution. These are key elements when considering rural electrification objectives, where load growth can be particularly uncertain, thus multi-year approaches are more applicable in the targeted context [35,36].

The majority of these approaches are based on either Mixed-Integer-Linear Programming (MILP), Linear Programming (LP), Multi-Stage Stochastic Programming (MSSP), or evolutionary methods such as the genetic algorithm (GA). In most works, however, to address the nonlinearities of certain model components and problem dimensionality, significant simplifications are made – particularly in terms of uncertainty recognition and quantification. These may also include considering only a specific component of the system (i.e., battery in [37] for example), static energy system capacities through project duration or only using a few representative days for each year [28,29]. The manuscript in [38], for example, implemented a multi-year approach for the planning of a microgrid with no dispatchable generator, but did not account for any stochastic elements nor a complete planning horizon. Similarly, other works focused on rural electrification, while modelling load growth stochastically, do not account for uncertainty on energy generation over the project duration [39]. Typically, only one or two sources of uncertainty are considered, and their realizations limited to a predefined set of scenarios. Furthermore, the majority of these approaches limit decision-making flexibility options through the project lifetime, primarily to maintain computational tractability. Generally, in fact, only one level of expansion is considered for each technology, and expansion decisions are only made based on scenario realizations, with limited generalizability [24,25,39,40].

The planning and design of a nomadic energy supply system, however, must balance delicate trade-offs between mobility, security, efficiency, sustainability, and economic feasibility in light of the extensive associated operational uncertainties. Flexibility is a potential value enhancing paradigm to help address some of the above-mentioned limitations of common energy system planning approaches [41,42].

Flexibility in Design explicitly considers decision-making possibilities over a systems lifetime, allowing for more cost-effective adaptability, sustainability, and resilience under uncertainty. This is highly applicable for evolvability in complex energy systems projects. This field aims to produce computational tools to support early conceptual design phases as part of a holistic framework, with primary applications across uncertainty modelling, concept generation, and design space exploration [41]. These methods help produce modularised designs with decreased exposure to downside risks while capitalizing on upside opportunities, shifting the distribution of performance as compared to standard approaches. Improvements typically range from 10 to 30 %, although several energy system application studies suggest even greater expected performance enhancements may be achievable, depending on the problem at hand [43–46]. Planning a PP based mobile energy system through a *Flexibility in Design* approach can, therefore, likely help mitigate investment risk and fiscal burdens incurred by the targeted low-income nomadic communities under uncertainty.

Given the degree of uncertainty these rural isolated energy system planning problems normally face, it is essential to consider a greater level of decision-making flexibility to enable adaptability to different encountered conditions. Deep Reinforcement Learning (DRL), while widely investigated and proven for applications in energy systems management and control [47], has yet to be directly applied to multi-year microgrid planning and design under several sources of uncertainty. The closest work would be that by [48], however DRL was only used by the authors for the dispatch problem, and components were sized through evolutionary methods. Additionally, energy system capacity planning or decision-making flexibility under uncertainty were not considered. An important motivation for this work is, therefore, that a data-driven approach may help address some of the bottlenecks and simplifications required by current state of the art methods such as MILP or MSSP. A DRL based approach, it follows, is likely to enhance the planning and optimisation process of future energy systems by allowing an expanded design, strategy, and uncertainty space - which is more realistic of the actual conditions encountered, while maintaining reasonable computational expense.

1.3. Data-Driven energy systems design for flexibility

Standard methods for *Flexibility in Design*, however, present many of the same computational limitations as the planning approaches mentioned in the former section. Typically, they are also based on an MSSP or decision tree formulation over a long time horizon [50]. At each stage, the possible actions represent system adaptability options, and the optimal solution quickly becomes intractable if several uncertainty sources or flexibility strategies are considered [51]. Path independence must be assumed, which is unrealistic in the context of a nomadic energy system. The solution procedure has also been criticized, as one that is difficult to understand and implement in practice for decision makers [52]. Decision rules help address some of these issues, by mapping system evolution strategies to physical design variables in a way which is intuitive for decision makers [41]. Although they have been shown to estimate similar values of flexibility as more complex alternatives, strategies are still limited to generic real options or domain specific ones.

A data-driven approach to *Flexibility in Design* may help address some of these limitations, by more systematically evaluating the potential decision rule space, as shown recently by Caputo and Cardin [49]. Their results on an example energy systems design and planning case study suggest that Deep Reinforcement Learning (DRL) can enhance solution expected value under uncertainty, leading to several potentially interesting applications. DRL is a subset of ML defined by a sequential decision making process, and is best known for its state of the art results on game-like environments [53,54]. In recent years, however, it has been successfully implemented in several engineering domains with significant variations in action and state spaces, as well as objectives [53,55–59]. In the energy systems field, it has been extensively

investigated for management and control [47], but it has yet to be directly implemented on micro-grid design and capacity planning for rural electrification.

Generally, the aim would be to iteratively find the best adaptability strategies under a wide range of scenarios by balancing exploration and exploitation in the learning process. Deep Neural Networks (DNNs) can be used to accurately approximate a complex decision-making policy and corresponding value of each flexibility strategy over time. This allows for the tractable integration of several sources of uncertainties and mobility considerations associated with a mobile energy supply system. The design and uncertainty space relevant to an energy system planning problem, it follows, can be significantly expanded as opposed to analytical methods such as MSSP [49]. Consequently, a data-driven approach may also help identify novel, unintuitive and more dynamic adaptation strategies outside of the standard generic real options framework. This is particularly valuable in the case of a unique, complex, and highly uncertain context such as the one tackled in this work, where there is limited domain expertise. As a result, energy system designs and flexibility strategies developed through a data-driven approach are likely more compatible with (and representative of) the range of real-world conditions potentially faced by mobile energy systems in the future.

1.4. Contributions

In light of the pressing issues discussed above, this paper presents the design and planning of a mobile multi-energy supply system which is compatible with a nomadic lifestyle, seeking to integrate both electrification and heating supply decarbonisation objectives. To address deficiencies of common "static" energy planning approaches in this context, a new modelling framework is proposed where the spatial dispersion (and network configuration) among nomadic dwellings varies based on seasonal requirements, as well as main grid availability. In order to tackle the resulting highly complex and dimensional stochastic optimisation problem, a novel DRL based data-driven approach is proposed to avoid the simplifications and compromises associated with typical planning methods mentioned in Section 1.2.3. The primary contributions are summarized below.

1. Development of a novel modelling framework for the design and planning of mobile multi-energy supply systems, focused on achieving SDG7 for global nomadic communities, but applicable to general trends in increasing decentralization of energy systems. This represents the first study integrating considerations of seasonal centralized to decentralized operation over a long planning horizon, enabled by PP operation. To these authors' knowledge, the novel approach proposed is the first allowing for changing network topologies and/or spatial dispersion over time, while capturing the fiscal and mobility constraints faced by a typical targeted nomadic community. The framework is developed to be generic and easily scalable to the design and capacity planning optimisation of different communities wishing to operate a mobile energy supply system.

2. Implementation of a new DRL based approach optimising long-term flexibility strategies for nomadic communities, marking the first time DRL is used in an energy access planning context. Several sources of uncertainty are, as a result, able to be tractably integrated, including on energy generation and demand, spatial dispersion, and main grid availability. This approach enables for the first time the consideration of monthly adaptability strategies, capturing the interactions between mobility and seasonality on optimal system design. Resulting policies can be highly dynamic, and are more generalizable than those obtained using only a few select scenarios. Results are benchmarked against baseline rigid designs with better economies of scale to estimate the values of decision-making flexibility in the context of a migratory lifestyle across economic, sustainability and resilience metrics.

3. Multi-objective assessment of value of PP operation and "breakeven" spatial dispersion point to determine investment guidelines and policy insights to accelerate nomadic energy access. This is performed using a variation of real-options theory and represents the first study focusing on long term planning of mobile PP enabled energy systems enabled. As such, it yields important quantitative results on its practical feasibility and global impact potential.

The novel modelling framework and methodology is exemplified through a case study based on Mongolian herder communities, developed in collaboration with local energy sector stakeholders. The general approach, however, may be easily extended to other nomadic communities worldwide, or to the overall needs to better understand the feasibility of mobile energy systems which switch periodically between centralised and decentralised operation to increase network reliability.

The rest of the paper is organized as follows. Section 2 details the novel modelling framework developed to account for mobility. Section 3 introduces the novel proposed DRL based design and planning optimisation methodology, as well as the benchmark rigid stochastic capacity planning model, including the different optimisation parameters implemented. Section 4 gives a more in-depth overview of the application case study which initially motivated this work. Finally, Section 5 presents the results of the study, accompanied by discussions and conclusions for future work in Sections 6 and 7, respectively.

2. Mobile Micro-Grid model formulation

2.1. Problem statement

2.1.1. Introduction

This section describes the formulation of the model for the planning of mobile-multi energy networks for nomadic communities over a long-term horizon. The integrated model accounts for changing and uncertain movement patterns over time, micro-grid distribution losses, variation in renewable energy output, and different heating generation sources. Furthermore, it considers the possibility of connection/ market engagement with the main distribution grid when available, allowing switching between centralized and decentralized operation based on seasonality and user requirements. Hourly resolution is used for power flow calculations to tractably approximate generation, distribution, and losses at the system level. On the other hand, monthly resolution is implemented for cost calculations within the flexible capacity planning model. This allows for adaptation strategies based on changing climatic conditions and intra-year seasonality, while considering the impact of long-term realisations of uncertainty on mobile micro-grid performance.

Two different mobility scenarios (MS) are thus developed to reflect the diversity in movement pattern. MS1 depicts operation of the micro-grid in the rural or decentralized setting, typically in the summer months for nomadic communities as they migrate to maintain their livestock. MS2, on the other hand, represents operation in a centralized setting, or when connection to the main grid is available, typically in winter months for nomads - when pasture productivity is lowest.

2.1.2. Objectives

Two primary alternatives are compared based on the model presented in the following sections for a community operated mobile micro-grid. Firstly, an inflexible baseline is considered where all capacity is installed at project start and no adaptation strategies in response to uncertainty are possible. Secondly, a flexible design exercising modular deployment of different technologies over time based on realised operating conditions is developed. The baseline rigid design presents lower unit costs due to economies of scale and reduced transport charges, while the flexible design allows for the deployment of resources only if and when needed. The objective for both designs is minimization of Expected Net Present Costs ($ENPC_{es}$) for the energy system (es) over 30 years (T).

Nonetheless, this work is also motivated by the UN SDGs and associated sustainability (i.e., Scope 1 and 2 emissions) and reliability energy access objectives. Emissions are thus found from the direct combustion of coal for heating and environmental footprint of the electricity purchased from the main distribution grid, respectively [60]. Additionally, 30-year horizon total unmet load (TUL) is used as a proxy for estimating system resilience and energy security, considering at a higher level the impact of demand and generation uncertainty. It should be noted that these metrics are not explicitly optimised, but are used for the qualitative evaluation of design alternatives in 5.3.

2.1.3. PP control and operation

A core enabler of this kind of analysis and system design methodology is the novel potential for PP operation, one of the key functionalities benefitting from the high-penetrated power electronic devices, e. g., inverter. PP operation enables the mobility of energy sources by implementing intelligent algorithms to guarantee operational stability, which is essential for compatibility with a nomadic or migratory lifestyle [17,61]. This creates an opportunity for economic, sustainability and energy security benefits, even in the rigid baseline design, as further discussed in Section 5.4. The concept of PP operations only requires real-time power-electronic level control design that has been widely investigated in [62,63]. The control objective within the micro-grid system is to guarantee that frequency and voltage converge to their reference values (i.e., frequency of 50 Hz and voltage of $220\sqrt{2}$ V) to maintain network stability. This is achieved through active power sharing across DGs connected to the microgrid energy supply system. The images on the left of Fig. 1 demonstrate the changing microgrid configurations and respective estimated time resolutions for a typical operational scenario encountered. Starting from normal conditions (1), one of the DGs is disconnected in (2) as a family decides to migrate, and it eventually reconnects to the microgrid in (3). The right sub-plots (a)-(d) in Fig. 1 demonstrate that PP operation is able to maintain appropriate values for frequency $f(Hz)$, active power ratio mP, active power $P(kW)$, and voltage $V_{mag}(V)$, respectively, across scenarios (1)-(3). Over the simulated 5 s horizon, it should be noted, some minor transients are still recorded at a fine resolution due to the changing system loads. Nonetheless, the demonstration results suggest PP operation should be an effective and reliable solution to allow temporary interconnection among nomadic community members.

2.2. Nanogrid and network configuration

The architecture of the energy system proposed in this paper is a distributed form of mobile micro-grid. A single herder family *nanogrid (*i. e. *sub-microgrid) unit* (N_n) which moves and settles very close to each other is used as a building block for the simulation, allowing for improved scalability and compatibility with PP operation [64]. The term nanogrid is justified as each herder family unit is normally expected to be able to produce, store and distribute electricity in fully self-sustaining or islanded manner- as sometimes necessary during migration. Through interconnection with other community members, however, the aim is to reduce the total system cost by more efficiently managing supply and demand. Fig. 2 graphically captures the network configuration, both at the level of an individual herder family nanogrid unit (N_n-left) and the PP mobile micro-grid (right).

The typical nanogrid unit consists of energy generating units – in this case solar PV and wind – as well as a storage system (BESS). While both DC and AC loads may be present in the system (more details in Section 4.4), interactions among nodes connected through the PP configuration is designed to be 3-phase AC, thus converters and inverters are included where appropriate. Heating demand may be present during the winter months which must be fully met using either a traditional coal stove (left) or electric heater (right). Moreover, it is assumed for it to be possible for a group of nomads, perhaps from the same extended family or community, to operate as a "cluster". The cluster radius variable $r_{cluster}$ is thus defined as the smallest possible circle which could encompass them all – as shown in Fig. 2. It is estimated that this is not likely to vary significantly *within* a single season, although there may be important *year-to-year* differences based on the impacts of climate change. Stochastic realizations for a single month m are thus obtained proportionally to average seasonal values at project start, directly impacting the distance $L_{cb}{}^t$ required to connect the mobile energy system, as shown in (1).
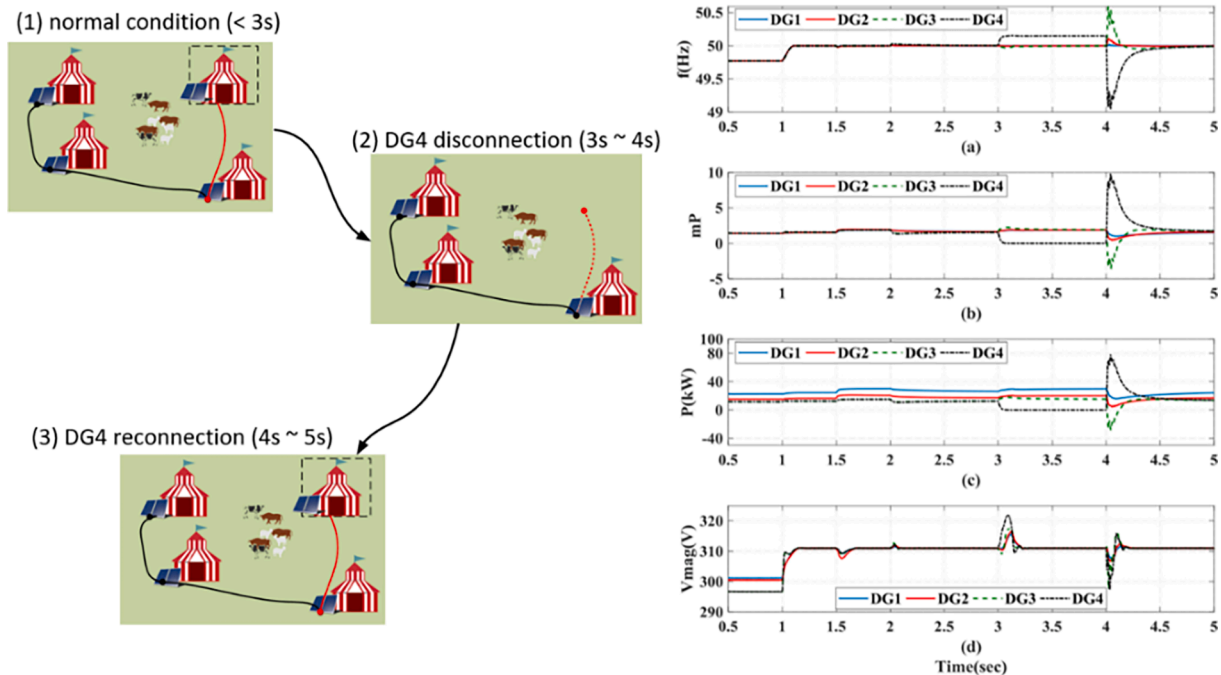

Fig. 1. PP operation demonstration including typical (1) ~ (3) operational scenarios (left) and performance evaluation of frequency, active power ratio, active power, and voltage, in (a)-(d), respectively, over the same time horizon (right).
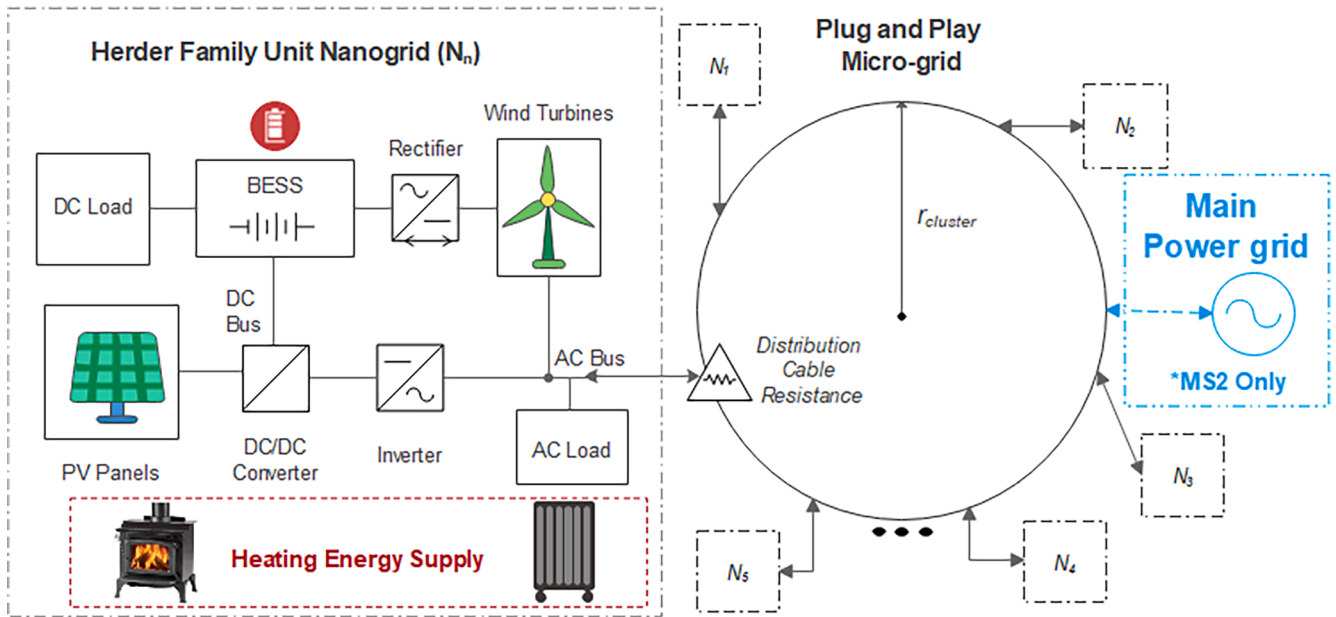
Fig. 2. Multi energy vector block diagram of herder nanogrid (left) and PP microgrid (right).

$$L_{cb}{}^t = 2\pi r_{cluster}{}^t \tag{1}$$

The different nodes are assumed to be configured in a fixed, ring-like interconnection, evenly spaced across the outer circumference obtained from $r_{cluster}$. This configuration was previously reported to optimise distribution efficiency under a similar decentralised setting as depicted above [14,19,21,65,66]. The proposed approach can, therefore, account for changing migration distances over time, while also integrating social considerations on spacing among family units. Conditional connection to the main grid during MS2 is approximated as an additional node in the network (blue in Fig. 2).

2.3. Renewable energy generation

The renewables.ninja platform [67] is used for the simulation of wind ($P_{wind}^{n,t}$) and solar PV ($P_{pv}^{n,t}$) intermittent generation for each nanogrid n at an hourly resolution t, based on locations representative of each MS. The closest matches in terms of technical performance parameters were used since the small-scale systems considered are not available for direct simulation within the platform. More specifically, Solar PV generation is found through the Global Solar Energy Estimator, with default tilt and azimuth ($\beta = 35°$, $\phi = 180°$), an assumed additional system loss fraction of 10 %, and no tracking mechanism implemented [68]. Small scale wind generation is computed through the Virtual Wind Farm model and historical wind speeds obtained from NASA's MERRA-2 dataset, implementing bias-corrected reanalysis to increase the accuracy of prediction, at the lowest recorded height [69]. Moreover, to investigate the impact of RES intermittency, random Gaussian noise is added to the baseline deterministic profiles. Respective stochastic parameters for wind and PV during MS1/2 are found from their historical hourly generation potential [70].

2.4. Power flow calculations

2.4.1. Nanogrid operation

The LFS is implemented for operational control simulation of the design alternatives evaluated. For any hourly timestep t, the net electricity demand load $P_{NEL}^{n,t}$ at node n is calculated from the hourly electricity demand $ED^{n,t}$ and on-site renewables generation $P_{res}^{n,t}$ (accounting for inverter and converter efficiencies). EH loads are prioritised over BESS charging, limited by installed capacity $\theta_{eh}^{n,t}$, hourly maximal

output $P_{eh,max}^{n,t}$, and power available from total generation $P_{eh,res}^{n,t}$. The remaining power $P_{res,PL}^{n,t}$ and heating demand $P_{NHL}^{n,t}$, if any, dictate BESS operation at each node, which is prioritised over PP distribution to minimise distribution losses. The energy flows into the BESS capacity θ_{ba}^t are then determined from their technical parameters, namely charging C_{in} and discharging C_{out} rates as well as maximum nominal input $P_{ba,in,max}^{n,t}$ and output $P_{ba,outmax}^{n,t}$ power. They are also bounded by the minimum $E_{ba,min}^t$ and maximum $E_{ba,max}^t$ storage capacities for the selected BESS. The battery storage level E_{ba}^t is thus computed as the sum of the stored energy from the previous timestep E_{ba}^{t-1}, storage energy flow P_{ba}^t and any potential additional installed storage capacity during that period $E_{ba,new}^t$. Note that newly purchased batteries are assumed to arrive *fully charged,* and conversion losses are accounted for via the battery roundtrip efficiency parameter η_{ba}. Remaining excess energy $P_{eg}^{n,t}$ after nanogrid operation is then used to engage in PP operation.

2.4.2. PP Micro-grid distribution

Distribution requirements are calculated at the entire system level from each node, which are classified as either $N_{deficit}$ or $N_{surplus}$, based on the obtained energy balance. The system wide unmet load $P_{ul,es}^t$ and available extra generation $P_{eg,es}^t$ are thus used to roughly estimate microgrid distribution requirements $P_{dist,es}^t$ at an hourly resolution. The microgrid power distribution potential is limited by the capacity of the cables connecting the different units θ_{PPmg}^t, and is capped at no more than 30 % of the system wide electricity load $P_{dist,es,max}^t$, aiming to balance network reliability with increasing infrastructure cost, as shown in (2).

$$P_{dist,\,es}^t = \min\left(P_{ul,es}^t, P_{eg,es}^t, \theta_{PPmg}^t, P_{dist,\,es,\,max}^t\right) \tag{2}$$

Furthermore, to tractably approximate performance, it is assumed that the distribution of power is spread *homogenously* across the microgrid network, as both loads, and generation should be well distributed spatially in light of migration compatibility objectives. More specifically, representative power ($P_{dist}^{t,n} = \frac{P_{dist,es}^t}{N}$) and length ($L_{dist}^{t,n} = \frac{L_{cb}^t}{N}$) requirements for PP distribution per each node n are developed for preliminary loss calculations. Dwellings are assumed to be connected via 3-phase transmission rated at 220/380 V through low-cost copper cables [17]. Interconnection resistance per node $R_{dist}^{t,n}$ is thus found from the

manufacturer reported reference value and tuned to the actual operating conditions through the material temperature coefficient α_{cb}. Cable resistance is used to estimate energy system losses, $P_{dl,es}^t$ according to (3), neglecting reactive power and only focusing on thermal losses associated with the induced current $I_{dist}^{t,n}$ in each line. Finally, distribution loss proportion $DL\%$ is found from (4), seeking to capture overall network efficiency. The balance of power flows in the network after PP operation is ensured by computing the final excess electricity generation $P_{res,exc.pp}^t$, unmet electricity load $P_{uel,es,ppp}^t$ and/or unmet heating load $P_{uhl,es,ppp}^t$. These values then dictate utility grid interactions, as detailed in the following section.

$$P_{dl,\,es}^t = \sum_{n \in N} 3 R_{dist}^{t,\,n} I_{dist}^{t,\,n\,2} \tag{3}$$

$$DL\,(\%) = 100\,\frac{P_{dl,\,es}^t}{P_{dist,\,es}^t} \tag{4}$$

2.4.3. Utility grid interaction and coal requirements

The potential to buy electricity from the distribution grid during MS2 is considered in this section. Interconnection is simulated at a system level through simple heuristics (i.e. market engagement is *not explicitly* considered) to focus on flexibility strategies. Interactions with the main distribution grid, limited by line capacity θ_{Grid}^t, unmet heating load $P_{uhl,es,ppp}^t$, and availability ρ_{grid}, are calculated based on system energy balance, as shown by (5) below. Note that ρ_{grid} is a binary variable representing grid availability for that timestep, sampled from a *beta* distribution. Case (i) then represents timesteps when blackouts are happening so no interactions happen. Case (ii) shows situations when demand has not been fully met, thus electricity is purchased from the grid. For case (iii), on the other hand, the excess generation after meeting micro-grid loads $P_{res,exc.pp}^t$ is sold back to the main grid. Note that excess generation P_{ex}^t which cannot be captured by grid interactions is assumed wasted at this stage. Please refer to Appendix I for the complete power flow model equations.

$$P_{grid}^t = \begin{cases} 0 \ , \ (i) \\ \min\left(\theta_{Grid}^t, P_{eh,max}^{\,t}, P_{uhl,es,ppp}^t\right) \ , \ (\text{ii}) \\ -\min\left(\theta_{Grid}^t, P_{res,exc.pp}^t\right) \ (\text{iii}) \end{cases}$$

$$\rho_{grid} = 0 (i)$$

$$P_{uhl,es,ppp}^t \geq 0, \rho_{grid} = 1 \ (ii)$$

$$P_{uhl,\,es,\,ppp}^t \leq 0\,; \rho_{grid} = 1 \ (iii) \tag{5}$$

2.5. Emissions generation

Based on the above, the heating demand not met via EH, and therefore mass of coal required for the *entire nomadic community*, is calculated as shown in (6)-(7). $P_{coal,es}^t$ captures the thermal energy requirements after PP distribution $P_{uhl,es,ppp}^t$ and grid interactions P_{grid}^t. The associated coal mass M_{coal}^t is estimated from its average energy density $\mu_{localcoal}$ and thermal efficiency of the traditional stove used for combustion $\eta_{coalstove}$. Energy system carbon footprint over time is estimated from the mass of raw coal burned and grid purchased electricity according to (8), using their respective weighted average emissions factors (EF_{grid} and EF_{coal}).

$$P_{coal,es}^t = \max\left(P_{uhl,es,ppp}^t - P_{grid}^t, 0\right) \tag{6}$$

$$M_{coal}^{\,t} = \frac{P_{coal,es}^t}{\eta_{coalstove}\mu_{localcoal}} \tag{7}$$

$$CO2_{eq.emissions,es}^t = EF_{grid}P_{grid}^t + EF_{coal}M_{coal}^{\,t} \tag{8}$$

2.6. System costs

The model presented thus far ultimately seeks to calculate system costs at a monthly m resolution C_{es}^m in each uncertainty scenario. As shown in (9) below, the formulation accounts for investment costs C_{inv}^m, operational costs C_{opex}^m, any expenses (or revenues) incurred through interactions with the distribution grid C_{grid}^m, as well any potential penalty (or credit) associated with generated emissions C_{CO2}^m. Representing liquidity and economic limitations faced by these communities, a constraint is formulated to cap monthly household energy system expenditure at the cluster level, as shown by (10). The modelling of each of these cost components is discussed in more details below.

$$C_{es}^m = C_{inv}^m + C_{opex}^m + C_{grid}^m + C_{CO2}^m \tag{9}$$

$$C_{es,max}^m \leq C_{es,max}^m N_n \tag{10}$$

2.6.1. Investment costs

Investment costs in each timestep, as shown in (11), are calculated from the communication $CC_{n,i}^{\,m}$, infrastructure IC_{es}^m, expansion $EC_{n,i}$, and replacement costs $RC_{n,i}$, net the residual value (if any) $SV_{n,i}$ recorded across all nodes and technologies evaluated. Communication infrastructure costs CC_{es}^m required for the safe operation and control of the PP microgrid are preliminarily estimated through a simplified model based on system distribution requirements $P_{dist,system}^m$ and a cost constant factor c_{cc}, as shown in (12). Assuming a 2G/3G/4G base station is already present, the communication infrastructure needed is thus simply an add-on support enabling peer-to-peer communication among nanogrid units. They are considered as a one-off investment incurred only in timesteps where $P_{dist,es}^m$ is higher than monthly distributed electricity in all previous periods $P_{dist,es}^{m,max,past}$, representing the need for infrastructure upgrade - neglecting the effect of EoS. Similarly, IC_{es}^m – which represent the need for additional cabling to connect the various nanogrid units- are recorded for timesteps when the current cluster radius is higher than that recorded in any of the previous simulation months $r_{cluster}^{max,past}$. It is thus assumed that cabling infrastructure purchased at any point during project life will remain available for connection if needed again. The cost of required extra cabling is computed according to (13), using a unit length cost factor c_{ic} obtained from regional manufacturers [71], invariant with time or scale.

The formulation of $EC_{n,i}$ in (14), on the other hand, accounts for both economies of scale (EoS) and cost of transportation in the targeted settings. The estimations for the investment constant K_i and EoS factors α_i are, therefore, slightly reduced and increased respectively in MS2 as compared to the MS1 [72], and based on specifications from regional manufacturers per component [73–79]. Coal stove capacity is an exception to this, where it is assumed that no starting stage investment is required, although replacement costs are incurred every 7-years [80]. As such, stove capacity is not explicitly optimised, and it is simply assumed one will be needed per family at a fixed (and nearly negligible) cost over time. Any other given capacity decision $\theta_{n,i}^{m,new}$ is expected to yield additional power electronics requirements $\theta_{n,pe}^{m,new}$, assumed to be proportional to expansion magnitude. This is based on reported ratios of nominal power electronic to system component i capacity, $\varepsilon_{pe,i}$, averaged over several micro-grids from the literature [14,19–21,81,82]. A safety factor sf_{pe} of 1.3 is included to further ensure no undersized components, as shown in (15), yielding total power electronics costs $PEC_{n,i}^{\,m}$ through a constant cost factor $c_{pe,i}$.

Replacement costs $RC_{n,i}$ are incurred when a system component i reaches the end of its predicted operational life L_i, using the same parameters as (14). As such, $RC_{n,i}$ may be experienced for multiple

technologies in the same month- in contrast to $EC_{n,i}$, which is limited to a single system component i per timestep. The asset salvage value $SV_{n,i}$ is included for cases where capacity abandonment decisions are made, or for NPC calculations at the end of each simulated episode. As shown in (16), it is a function of remaining estimated functional hours L_i^{θ} for a component with an original maximal lifetime L_i^{max}, and initial investment cost based on the salvaged capacity amount $\theta_{sv,n,i}^m$, corrected through a depreciation factor k_i^{salv} which is smaller than 1.

$$C_{inv}^m = CC_{es}^m + IC_{es}^m + \sum_{i \in I} \sum_{n \in N} EC_{n,i}^m + RC_{n,i}^m - SV_{n,i}^m \tag{11}$$

$$CC_{es}^m = \begin{cases} c_{cc}\left(P_{dist,es}^m - P_{dist,es}^{m,max,past}\right), & P_{dist,es}^m > P_{dist,es}^{m,max,past} \\ 0, & P_{dist,es}^m \leq P_{dist,es}^{m,max,past} \end{cases} \tag{12}$$

$$IC_{es}^m = \begin{cases} c_{ic}\left(r_{cluster}^m - r_{cluster}^{max,past}\right), & r_{cluster}^m > r_{cluster}^{max,past} \\ 0, & r_{cluster}^m \leq r_{cluster}^{max,past} \end{cases} \tag{13}$$

$$EC_{n,i}^m = K_i\left(\theta_{n,i}^m - \theta_{n,i}^{m-1}\right)^{\alpha_i} + PEC_{n,i}^m = K_i\left(\theta_{n,i}^{m,new}\right)^{\alpha_i} + PEC_{n,i}^m \tag{14}$$

$$PEC_{n,i}^m = c_{pe,i}\theta_{n,pe}^{m,new} = c_{pe,i}sf_{pe}\varepsilon_{pe,i}\theta_{n,i}^{m,new} \tag{15}$$

$$SV_{n,i}^m = k_i^{salv}\frac{L_i^{\theta}}{L_i^{max}}K_i\left(\theta_{sv,n,i}^m\right)^{\alpha_i} \tag{16}$$

2.6.2. Operational costs

The operational costs C_{opex}^m detailed in (17)-(21) below account for expenses due to heating via a traditional coal stove C_{coal}^m, maintenance of fees associated with each energy system component C_{om}^m and the economic value of unmet electricity load C_{ul}^m. Expenses associated with space heating are found through the required monthly coal mass M_{coal}^m and the local average price by weight c_{coal} – assumed invariant with time [80]. O&M costs are found as a function of total energy system installed capacity $\theta_{i,es}^m$ and respective periodic estimated maintenance fee $c_{om,i}$ for each component i, using recently reported average values for small scale systems [83]. To compute the economic value of energy not served, the total monthly unmet load $ED_{ul,es,final}^m$ for the entire system is multiplied by the curtailment fee c_{ul}, as shown in (20). It is assumed that a *5 % monthly shortage is allowed*, thus no charges are associated below that level. The value for c_{ul} is estimated through the Willingness to Pay (WTP) approach, which relates perceived marginal benefits to the maximal price a customer would be willing to pay for a service [84]. It is, therefore, assumed equivalent to the local cost of diesel which would be needed to supply the unmet load – neglecting generator investment cost and assuming standard conversion efficiencies. Cash flows resulting from grid interactions can be either positive or negative, based on the monthly energy balance P_{grid}^m, and are computed using either the reported discounted tariff c_{grid} or subsidized price received p_{grid}.

$$C_{opex}^m = C_{coal}^m + C_{om}^m + C_{ul}^m \tag{17}$$

$$C_{coal}^m = c_{coal}M_{coal}^m \tag{18}$$

$$C_{om}^m = \sum_{i \in I} c_{om,i}\theta_{i,es}^m \tag{19}$$

$$C_{ul}^m = c_{ul}ED_{ul,es,final}^m \tag{20}$$

$$C_{grid}^m = \begin{cases} c_{grid}P_{grid}^m, & P_{grid}^m \geq 0 \\ p_{grid}P_{grid}^m, & P_{grid}^m < 0 \end{cases} \tag{21}$$

2.7. System mass

The total system mass M_{es}^m can be found from (22) for any month m as

the sum of the weight of the installed renewables M_{res}^m, batteries M_{batt}^m, power electronics M_{pe}^m, and cabling length required for connection M_{cb}^m. To reflect mobility considerations associated with a migratory lifestyle, an upper limit on energy system weight $M_{es,max}^m$ may be imposed, captured by (23).

$$M_{es}^m = M_{res}^m + M_{batt}^m + M_{pe}^m + M_{cb}^m \tag{22}$$

$$M_{es,max}^m \leq M_{es,max}^m N_n \tag{23}$$

2.8. Uncertainty simulation

Geometric Brownian Motion (GBM) models are developed to estimate stochastic evolutions of the most important sources of uncertainty integrated in the energy system planning problem. This form of stochastic model is chosen as it can represent, for any variable x, both overall long-term trends (if any), and the potential for tail event or low probability scenarios. This is deemed highly compatible with most of the uncertainties considered in this work. It should be noted, however, that if available data were to suggest otherwise, alternative, and more complex stochastic model formulations should be explored outside this proof-of-concept work. Mathematically, the model involves drift μ_x and volatility σ_x variables, respectively. dW_t models the random Wiener process, or stochastic deviations from the mean as sampled from a normal distribution, and dt the time resolution implemented. These values, typically based on historical data, allow one to fine-tune the stochasticity of the model, while not limiting the potential total evolution scenarios [85–87]. They can then be used to simulate uncertainty for any variable x as shown by (24)–(25) below.

$$dX^t = \mu_x X^t dt + \sigma_x X^t dW_t \tag{24}$$

$$X^t = X^0\exp\left(\left(\mu_x - \sigma_x^2/2\right)t + \sigma_x W_t\right) \tag{25}$$

2.9. Energy system model overview

Fig. 3 captures the interaction among different model components detailed in the former sections, leading to calculations of total monthly net system cash flow ($NCF^t = C_{es}^m$) and emissions $CO2_{eq,emissions}^m$ across different simulations. Changes or upgrades to the energy system state are determined by the DRL agent policy captured by the green box, as detailed in Section 3.2, impacting both investment and operational costs incurred in the remaining simulation steps. For the baseline rigid systems, on the other hand, cost calculations are based purely on uncertainty realizations.

3. Methodology

3.1. Rigid stochastic capacity planning model

3.1.1. Mathematical formulation

This section introduces the generic formulation of the rigid capacity planning model under uncertainty. It is later used to benchmark the performance of the proposed DRL approach on an example mobile energy supply system. Considering a discrete finite time horizon planning period of T, with discount rate λ, let $\xi = (\xi^1, \xi^2, \cdots \xi^T)$ be a scenario of uncertainty, where ξ^t is a vector capturing the uncertainty observed in period t, thus able to account for multiple uncertainty sources and relevant variables. S defines the set of all possible uncertainty scenarios s used for evaluation, assumed to be finite at S_s for simplicity, with equal probability of occurrence. The baseline single objective rigid stochastic planning model for the mobile energy system considered can thus be formulated as shown by (26)-(28). θ_i^t represents the installed nominal capacity for various system components i, invariant with time t. $\mathcal{R}^t\left(\theta_i^t, \xi_s^t\right)$ is the revenue function for the installed technologies and $\mathcal{C}^t\left(\theta_i^t, \xi_s^t\right)$
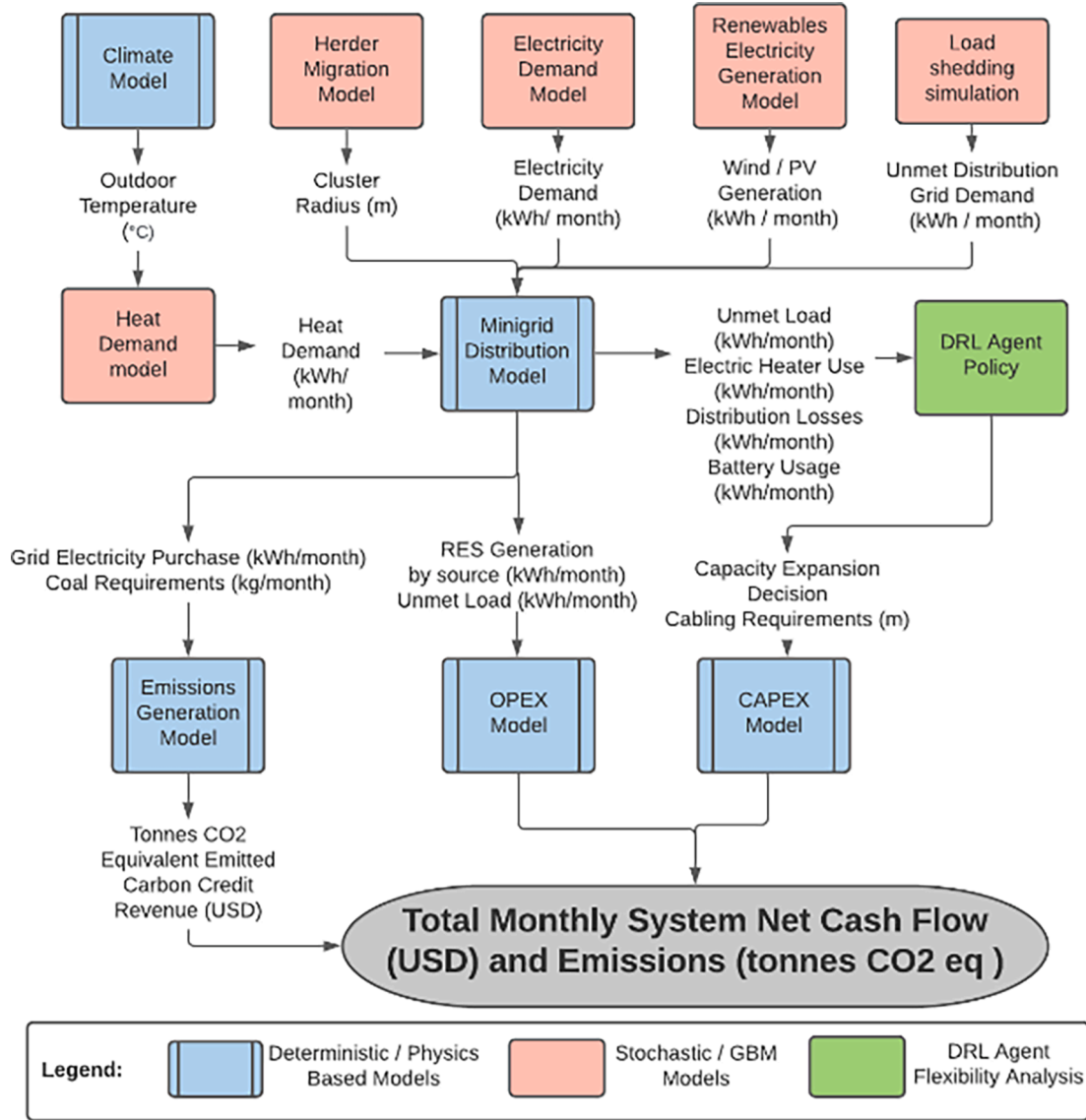
Fig. 3. Model flowchart showing complex interactions among system components and simulations.

the cost function for the same technologies in period *t*. Initial investment costs for each of these components *CAPEX$_i$* are included as a zero-stage decision, while periodic energy system costs under uncertainty $C_{s,es}^t$, detailed in Section 2.6, are captured by (27).

$$\min ENPC = \frac{1}{S_s} \sum_{s=1}^{S} \left(CAPEX_i + \sum_{t=1}^{T} \left(\frac{1}{1+\lambda}\right)^t \left[\mathscr{R}^t\left(\theta_i^t, \xi_s^t\right) \right. \right.$$

$$\left. \left. - \mathscr{C}^t\left(\theta_i^t, \xi_s^t\right) \right] \right) = \frac{1}{S_s} \sum_{s=1}^{S} NPC_s \qquad (26)$$

$$C_{s,es}^t = \mathscr{R}^t\left(\theta_i^t, \xi_s^t\right) - \mathscr{C}^t\left(\theta_i^t, \xi_s^t\right) \qquad (27)$$

$$\forall t \in T, s \in S, i \in I \qquad (28)$$

3.1.2. Optimisation

The optimal sizing of the baseline inflexible systems is found through the genetic algorithm (GA) implementation in Python [88]. GA is a well-known and widely used approach to global optimisation compatible with mixed continuous/discrete variables, inspired from the selection process found in natural biological systems [89,90]. Solutions are iteratively improved within the design space through operations such as

crossover, mutation, and fitness selection to try and imitate the natural evolution process. Initially, candidate solutions for different baseline mobile energy systems (or "chromosomes") are generated probabilistically with a random component across the design space. This helps explore a wide range of capacities for the different technological option baselines considered, as clarified in Section 5.1. These candidates are then refined and or/discarded sequentially through the genetic operations mentioned. The fitness function this process seeks to optimise is given by (26), or the minimisation of ENPC over 2,000 (fixed) simulated scenarios. A penalty function is implemented to ensure constraint-feasible solutions, as presented in [91].

3.2. DRL flexible capacity planning model

3.2.1. Background and notation

DRL is designed to make decisions in Markov Decision Process (MDP) and is intuitive in nature. It involves an agent interacting with a system (or environment) over time through a large number of iterations as part of a sequential decision-making problem, with the aim of reward maximization [92]. Generally, at each time step *t*, the agent finds itself at state s^t, and selects an action a^t following a decision-making policy π. As a result of taking the action, the agent transitions to a successive state

s^{t+1} receiving a corresponding scalar reward R^t, which are governed by the environment's dynamics and reward function $\mathscr{R}(s^t, a^t)$, respectively [93]. This repeated interaction, has the objective to improve the accuracy of expected value estimations for either each state $V_\pi(s^t)$ or state-action pair $Q_\pi(s^t, a^t)$. They refer to the total expected value of being in a particular state or the *decomposed* expected value by each action available to the agent in that state, as shown in (29) and (30) respectively. Based on the above, the agent can map and rank state-action pairs through their relative value. Ultimately, the aim is to develop an optimal decision-making policy π^*, as captured in (31), which maximizes *discounted* (through γ) accumulated rewards per episode. In a flexibility planning application, an episode is equivalent to a simulation of the full project horizon.

$$V_\pi(s^t) = \mathbb{E}[Q_\pi(s^t, a^t)\,|s^t] \tag{29}$$

$$Q_\pi(s^t, a^t) = \mathbb{E}[R_\pi(s^t, a^t)] \tag{30}$$

$$\pi^* = \mathrm{argmax}_\pi \mathbb{E}_{\tau \sim \pi(\tau)}\left[\sum_t \mathscr{R}(s^t, a^t)\right] = \mathrm{argmax}_\pi \mathbb{E}_{\tau \sim \pi(\tau)}\sum_t \gamma R^t \tag{31}$$

The optimal policy (or flexibility strategy) is then typically iteratively approximated inside a DNN through either value based or policy-based methods. The DNN often consists of several layers to improve the granularity of value (or policy) function representation over the training process (as shown inside the agent in Fig. 4), with optimal architectures varying based on the problem at hand [94]. Approaches belonging to the former, such as Q-learning, normally follow the trajectory of the highest value estimates to formulate the optimal policy. This allows for off-policy learning, as each update can use data collected during all training interactions, although potentially leading to estimation bias and improper credit assignment [55,95]. Policy based methods, on the other hand, parametrize the policy directly by creating an underlying action selection probability distribution, making them generally more effective in high-dimensional stochastic environments [96]. Most modern DRL algorithms, nonetheless, implement hybrid actor-critic formulations to combine aspects of both for state-of-the-art results on benchmark tasks [53,93,97,98], motivating their selection for this work. Generally, they are defined by an actor which determines which action to take under a policy, and a critic which approximates how good the actor's selection was based on value function estimates. The actor-critic architecture implemented in this study is further detailed in Section 3.2.4.

3.2.2. DRL problem formulation

A DRL based approach is used to evaluate and optimize an implicit form of *decision rules* to act as a triggering signal for making a particular flexibility decision during system operation, as captured by a policy π within a DNN. The major difference compared to alternative methods
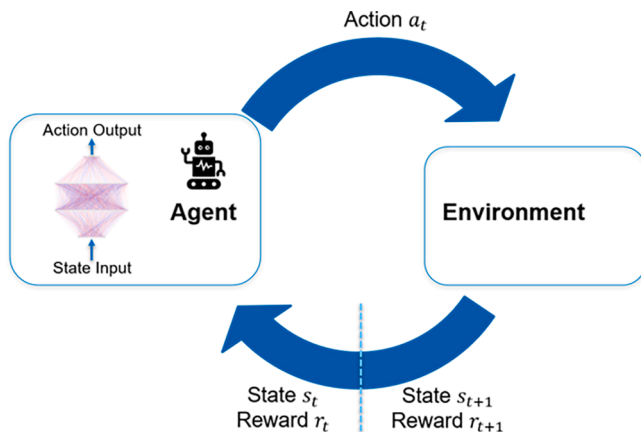
presented in Section 1.2.3 is that π can be characterized by a much larger set of parameters, and thus actions need not be restricted to a static pre-determined form. The primary steps required to formulate an energy system design and planning optimisation through the DRL Flexibility Analysis framework are problem formulation, uncertainty recognition, and action space creation.

More generally, it is assumed that the condition of the energy system for any $t \in T$ can be defined by a state observation, $s^t \in \mathbb{X} \subseteq \mathbb{R}^{n_x}$, where s^t captures the installed capacities for each component i, in addition to other factors. Based on the information given to stakeholders at each evaluation time step, they can sequentially select a system evolution decision, $a^t \in \mathbb{A} \subseteq \mathbb{Z}^{n_a}$, changing the capacity of a component if deemed necessary. Considering the targeted objective, a reward function, $R : \mathbb{X} \times \mathbb{A} \times \mathbb{X} \to \mathbb{R}$, can be defined to capture the impact of different available actions over time. This is based on a series of periodic revenue $\mathscr{R}^t(s^t, \xi^t)$ and cost $\mathscr{C}^t(s^t, \xi^t)$ functions, approximated through environment interactions, as well as a *deterministic* expansion cost function $\mathscr{H}^t(a^t, s^t)$, including a zero-stage decision on starting system configuration.

The energy system state is then updated as a function of the previous state, action selected, and realisations of stochastic variables. As such, the system is identified to hold the Markov Property, and can be formulated within the MDP framework. The objective, as for the baseline design, is minimisation of ENPC (or maximization of negative reward) through the development of a system evolution policy, $\pi : \mathbb{X} \to \mathbb{A}$. Note that a^t is then obtained from π by using s^t as an input, as shown in Fig. 4. The generic model for DRL based energy system flexible capacity planning is defined according to (32), and can be implemented on the same scenarios S_s as in (26) above. Furthermore, as $\pi(\emptyset, \cdot)$ is approximately captured by a DNN with parameters $\emptyset \in \mathbb{R}^{n_\emptyset}$, the capacity planning problem simplifies to obtaining optimal network parameters, or $\emptyset^*$.

$$\mathscr{P}(\pi) := \begin{cases} \max_\pi \mathbb{E}_\pi[-NPC] \\[4pt] s.t. \\[4pt] NPC_s = \sum_{t=0}^{T}\left(\frac{1}{1+\lambda}\right)^t R^t_s \\[4pt] R^t_s = C^t_{s,es} = \mathscr{R}^t(s^t, \xi^t_s) - \mathscr{C}^t(s^t, \xi^t_s) - \mathscr{H}^t(a^t, s^t) \\[4pt] s^{t+1} = f(s^t, a^t, \xi^t_s) \\[4pt] a^t = \pi(s^t) \\[4pt] \forall t \in T, s \in S \end{cases} \tag{32}$$

3.2.3. Environment creation

Central to the applicability of DRL flexibility analysis is the development of an agent-environment interaction framework which is representative of the targeted objective. The Open AI gym format environment [99] is implemented for the optimization of the mobile system given by (32). In this context, the system evolution strategy is primarily dependent on the periodic demand for different energy services, installed generating capacities and migratory patterns, among other factors. With the aim of reflecting this, the states for the agent are defined through a box observation set as:

$$s^t = \left[ED^t, HD^t, r^t_{cluster}, L_{cb}{}^t, \theta_{pv}{}^t, \theta_{wind}{}^t, \theta_{eh}{}^t, \theta_{bess}{}^t, ED^t_{ul,es,final}, M^t_{es}\right] \tag{33}$$

where ED^t and HD^t represent the electricity and heating stochastic demand realizations in timestep t, respectively. $r^t_{cluster}$ and $L_{cb}{}^t$ measure the uncertain spatial dispersion of the mobile system investigated and are used in cabling investment and distribution loss calculations. As in the former section, $\theta_i{}^t$ represents the installed *nominal* capacity for each system component i considered. $ED^t_{ul,es,final}$ is included to stimulate the explicit consideration of unmet load, outside of charges incurred, while M^t_{es} is used to account for increases in system mass as it evolves over

Fig. 4. Graphical overview of DRL agent-environment interaction.

time. Budget constraints are implicitly evaluated based on recorded rewards. All state variables are normalized based on the maximum and minimum values obtained for them over 10,000 simulated scenarios, independent of the environment. Furthermore, although not shown in (33), time limit awareness is included to improve end of episode policy performance and training stability [100].

The decision-making possibilities for the agent are also selected with the aim of reflecting the ones available to a ger community, through a multi-discrete action set ($\in A_{i,j}$). Through this formulation, the DRL decision space is such that, at any given time step, the agent may decide to leave the energy system configuration unchanged. On the other hand, based on estimated value function and policy, a decision enabled by flexibility may be made to alter the state of the microgrid. In that case, the multi-discrete set means a decision is made on which technology i to act on, and whether to pursue capacity abandonment, expand by 500 W or expand by the (rounded) maximum number of modules θ^t_{Budget} possible under the budgetary constraints C^t_{Budget} for that period. As actions are selected at the entire system level, a simple heuristic is implemented to determine which node to implement the capacity adaptability decision on. For expansion decisions on solar PV, wind or BESS, additional component capacity is assumed to be installed at the node with the largest unmet load in the previous month $ED^{n,t-1}_{uel,max}$. Abandonment actions or EH expansions, on the contrary, are allocated to the node with the largest *excess* generation $P^{n,t-1}_{res,excess}$ prior to implementing the flexibility option.

Fig. 5 below captures the internal evaluation process, going from state observation (left) to action selection (right) using a DNN parametrized agent policy, where components are evaluated within the environment technological boundaries. The iterative interaction feedback loop guiding policy updates during training, is visualized through the changes in system state, reward and policy driven by a capacity expansion decision made at time t on $t + 1$.

3.2.4. Algorithmic approach

Considering the dimensionality and computational expense of the problem tackled, the ACER (*Sample Efficient Actor Critic with Prioritized Experience Replay)* algorithm is implemented [101]. Each component is captured by a separate neural network, allowing different strategies to be implemented in policy as compared to value function updates. Experience replay allows agents to focus on the interactions most

significant to experienced reward, improving convergence rate during training [102]. Sample efficiency was motivated by the complexity of the model, evaluated with different timestep resolutions for operation and decision making, meaning each environment episode simulation is relatively expensive computationally. This is achieved in ACER by combining several ideas from previous works with novel developments.

More specifically, multi-step impacts on state-action value function $Q_{ret}(s^t, a^t)$ approximations for each generated trajectory are found recursively using the Retrace algorithm [103], as shown in (34). Multi-step returns used allow for significant reductions in policy gradient bias and learning iterations needed for the Critic component of the algorithm. A duelling architecture is implemented, through which the vectorized value function estimates $Q_{\varnothing}(s^{t+1}, a^{t+1})$ as well as the policy $\pi_{\varnothing}(a^t|s^t)$, are outputs of the DNN with parameters $\varnothing$. As shown in (29), the estimate $V_{\varnothing}(s^{t+1})$ is then easily found as the expectation of $Q_{\varnothing}$ under $\pi_{\varnothing}$. Additionally, note that γ is mathematically equivalent to the financial discount factor $\frac{1}{1+\lambda}$ shown in the problem formulated in (32). It is used to in DRL to account for the different value of immediate rewards compared to long term ones. The truncated importance weight $\overline{\rho}^t$ under current behaviour policy μ, is thus computed as shown in (35) [104].

$$Q_{ret}(s^t, a^t) = R^t + \gamma\overline{\rho}^{t+1}\left[Q_{ret}(s^{t+1}, a^{t+1}) - Q_{\varnothing}(s^{t+1}, a^{t+1})\right] + \gamma V_{\varnothing}(s^{t+1}) \quad (34)$$

$$\overline{\rho}^t = \frac{\pi_{\varnothing}(a^t|s^t)}{\mu(a^t|s^t)} \quad (35)$$

Policy updates are then conducted in a very similar manner to Trust-Region Policy Optimization [105], although using the running average policy network as a baseline instead. The policy is thus decomposed into a probability distribution f, and network capturing the properties of the policy distribution $\tau_{\varnothing}$, with the update split into two stages. In the first stage, a linearized Kullback–Leibler (KL) divergence constraint is combined with a standard loss minimization optimization. In the second stage, backpropagation is used to compute the derivatives with respect to the policy parameters, adopting a trust region based on the distribution f and advantage function. The gradient of ACER is thus defined as shown by (36) below, where c is the correction term coefficient for cases with very high variance. Please refer to the original publication for further details and definitions [106].
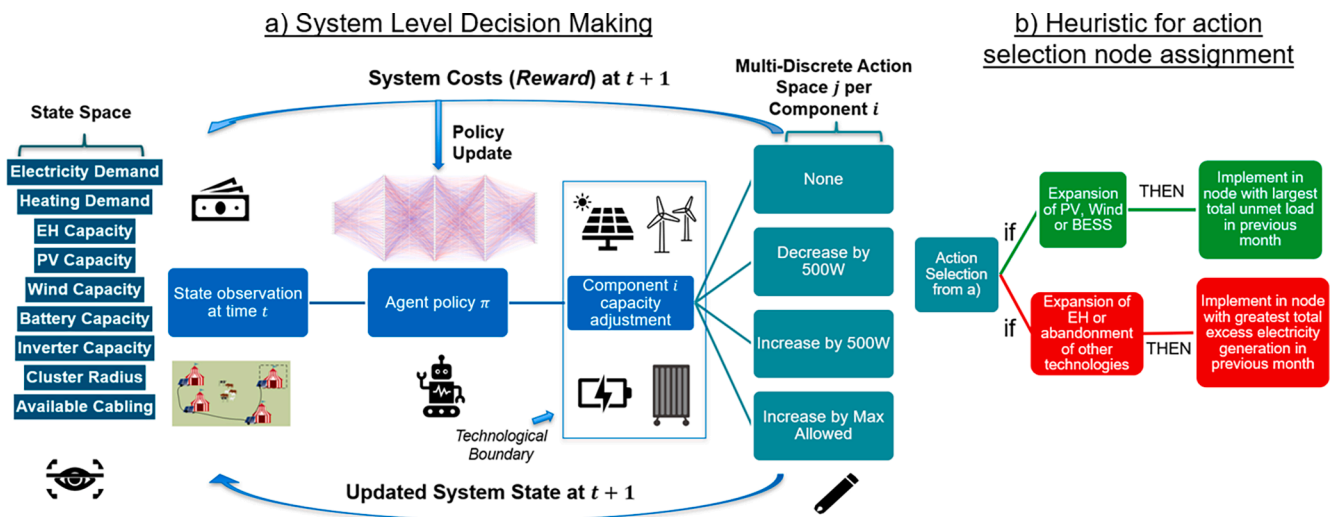

Fig. 5. a) Graphical overview of DRL system level decision making process going from microgrid state space input (left) to action selection output (right) b) Simple heuristic used for allocation of system level DRL actions to individual microgrid nodes, either the ones with either largest unmet load in previous month (top) or greatest total excess electricity generation (bottom).

$$\widehat{g}_{ACER}{}^t = \overline{\rho}^t \nabla_{\tau_{\varnothing(s^t)}} \log f(a^t | \tau_\varnothing(s^t)[Q_{ret}(s^t, a^t)$$

$$- V_\varnothing(s^t)] + \mathbb{E}_{a \sim \pi} \left(\left[\frac{\overline{\rho}^t(a) - c}{\overline{\rho}^t(a)} \right] \nabla_{\tau_{\varnothing(s^t)}} \log f(a^t | \tau_\varnothing(s^t)[Q_\varnothing(s^t, a^t)$$

$$- V_\varnothing(s^t)] \right) \tag{36}$$

3.2.5. DRL agent training and testing

The DRL agent described in the former section is trained using a budget of 6 million monthly decision timesteps, equivalent to 500,000 simulated full 30-year episodes. As performed on a standard laptop Intel (R) Core(TM) i7-8665U CPU @ 1.90 GHz, the complete training process takes approximately 7–8 h. All training scenarios are stochastically generated to be *i.i.d* from each other, thus leading to an *infinite* amount of potential scenario evolutions and state-action pairs. This was motivated by the need to develop a more generalizable adaptability strategy under different realizations of uncertainty. No exactly equal trajectories are thus possible inside the environment, and training process is fully non-anticipative, with all operational decisions based purely on recorded observations at a particular timestep. Hyperparameters are found through Optuna [107], with final training conducted using 50 environment steps per policy update, parallelized through 4 processes, a correction term of 12.3 and a replay ratio of 5. The replay ratio and denotes the average relative replay learning per policy learning – which only kicks in after 20,000 replay starts in this case.

The DNN architectures implemented for both the actor and critic component of the algorithm are fully connected, with input layers matching the problem feature space described above. The output layer for the actor is composed of twelve neurons, representing the multi-discrete action selection process depicted in Fig. 5 above, whereby four independent SoftMax distributional operators are then used to characterize decisions for each component. Additionally, six hidden layers with 64 neurons each are implemented to approximate the complex stochastic functions in each. Crucially, it should be noted that this approximation derived from the DNN captured by (32) does not always guarantee the optimality of the results obtained - especially when considering the non-convexity of the function evaluated. Nonetheless, several studies suggest strong convergence properties for DRL to low lying local minima when using gradient-based techniques, which employ a number of heuristics to reduce the likelihood of obtaining sub-optimal solutions [53,108–114]. The Stochastic Gradient Descent Momentum method [115] is thus implemented for DNN training parameter updates, using the exponentially moving average of the trust-region adjusted values from (36).

Two versions of the environment are developed to separate agent training and evaluation. In the training environment, a negative reward penalty (-USD 1,000) is included for timesteps where either the budgetary or mobility constraints presented in Section 4.2 are violated.

This allows the DRL agent to implicitly learn their form, looking for system solutions which are able to minimise the occurrence of constraint violation, and is the most commonly used method. While more advanced approaches are available to ensure constraints are robustly met, particularly in safety critical applications [116–118], they are typically associated with a high computational expense and are complicated to implement. As such, given in this setting the constraints represent socio-cultural considerations rather than safety issues, the simpler constraint violation penalty method is deemed justified.

This penalty is then removed in the testing environment, allowing evaluation of the true ENPC rather than the one obtained through the modified training reward function. Furthermore, to evaluate flexible system performance as compared to the baseline designs, all results are reported using the same 2,000 simulated *out of sample* episodes. This means that while trained on the same stochastic model *parameters*, no complete scenario used in testing would have been encountered during training. The different time step resolutions implemented are summarised by Fig. 6 below.

3.3. Value of flexibility and Plug and Play operation

The total value of flexibility for the energy system considered VoF^z_{Total}, on any metric z, is captured by (37). In essence, it comes from the combination of tactical system planning flexibility $VoF^z_{Strategic}$, and the added value enabled by the PP control developments $VoF^z_{Operational}$. As shown in (38), it is thus estimated as the difference between the expected value for the flexible DRL PP solution $EV_{Flexible}{}^z$ and the best performing inflexible benchmark $EV_{BestBaseline}{}^z$. For $z = ENPC$, as formulated in 3.1 and 3.1.2, VoF^{ENPC}_{Total} represents the maximum that should be paid to embed said flexibility in the energy system design. Furthermore, (39) is used to decompose the operational flexibility $VoPP^z$ generated by PP control possibilities from the DRL based decision-making flexibility proposed in this work. The value of operational flexibility is thus approximated from the difference in performance for baseline designs EV^z_{NoPP} as compared to ones integrating PP operation EV^z_{WithPP}. While it is recognized the values for these flexibility sources may not be purely additive, they are considered separately in this work to determine whether or not the infrastructure investment required to enable PP operation is worthwhile in the targeted context. Decomposing relative performance improvements by source may also be beneficial for various policy-making objectives. It gives relevant decision makers important quantitative insights across different metrics Z. These can then be used to help develop market support instruments or incentive programmes, with the aim of maximizing added value from flexible PP operation for nomadic communities. Please note these estimates are calculated over 2000 *out of sample* scenarios, as further detailed in 5.1.

$$VoF^z_{Total} = VoF^z_{Strategic} + VoF^z_{Operational} \tag{37}$$

- Equivalent to one 30-year simulation of system lifecycle
- 360 decision timesteps per episode (12 per year)
- Batch simulation of microgrid operation for entire month using hourly resolution

Environment Episode

- 200,000 environment episodes
- 72,000,000 decision timesteps
- ~ 5.2 x 10^{10} operational hours
- Reward penalty recorded for budgetary/ mobility constraint violation at the end of each month

DRL Agent Training

- 2,000 environment episodes
- 720,000 decision timesteps
- ~ 5.2 x 10^8 operational hours
- No reward penalties
- Same 30-year i.i.d uncertainty scenarios as benchmarked approaches

DRL Agent Testing

Fig. 6. Different timestep resolutions implemented for the proposed DRL approach.

$$VoF^z_{Total} = EV_{Flexible}{}^z - EV_{BestBaseline}{}^z \tag{38}$$

$$\forall z \in Z \tag{39}$$

4. Application case study

4.1. Background and previous work

Nomadic energy access and mobile energy systems are highly relevant in Mongolia, where it is estimated that 25–40 % of the national population continues to live a migratory lifestyle, normally moving between 4 and 30 times a year [119]. Changing climatic conditions, however, have increased the frequency of "zhuds" – extremely cold winters followed by extremely dry summers – making the traditional way of life more difficult for nomadic communities. This has resulted in an explosion of population migration to UB, where out of the approximately 1.4 million residents around 60 % live in "ger areas" [119]. A "ger" refers to the traditional Mongolian herder dwelling, comprised of a basic wooden frame and a number of insulating layers to allow for mobility. On the other hand, "ger areas" represent the urban regions surrounding UB where the majority of these low-income communities now settle in the winter [120], and normally migrate to rural surroundings during the summer months, with annual movement ranges of around 7–8 km [119]. Previous studies estimate nomadic community clusters are typically separated by 3.8 km in the spring, 2.0 km in the summer and 2.3 km away in the autumn [121]. Fig. 7 below graphically captures the migration patterns investigated, where MS1 represents operation in the area between UB and Hently in the summer, while MS2 is defined by migration to ger areas in UB during winter months, as in 2.1.

4.1.1. Electrification

The national network infrastructure is largely unsuitable to provide reliable and sustainable electrification across the potential MS encountered. It currently relies primarily on coal (71 %) and oil (25 %) for centralised generation, with a large associated carbon footprint [122]. More specifically, during MS1 the Northern Energy System is the primary option for herder electricity connection [119]. Highly underdeveloped and unstable it mainly relies on imports from Russia [123]. It also only covers a fraction of the potential migration range, with 45 % of rural gers unable to access electricity services consistently [124]. During MS2, on the other hand, gers are normally able to connect to the distribution grid in UB. This network, however, also experiences very frequent load shedding and blackouts, leading to limited energy security for ger communities [125].

In light of the above, household or community-based decentralised power supply systems are likely the most cost-effective option for achieving SDG7 in Mongolia [126]. In fact, more than 100,000 Small Home Solar (SHS) systems and 5,000 small wind systems (SWS) have already been distributed through the Renewable Energy and Rural Electricity Access Project (REAP), with capacities typically ranging from 20 to 100 W per ger [123,126]. REAP, however, was primarily based on capital investment subsidies and very low interest financing, thus not considering load evolution, mobility, or local resource assessment in capacity planning. Most of the systems currently installed are, as a result, either no longer in use or significantly undersized [126].

4.1.2. Heating supply decarbonisation

Electrification of nomadic communities has a very strong link with Mongolian government heating supply decarbonisation objectives, as residents of ger areas normally are unable to connect to the district heating network [80]. With the majority of the country in the extreme cold zone, there is substantial demand for space heating over a long winter season. As a result, it is estimated that ger areas contribute about 45 % to 75 % of UB's annual average fine particulate matter ($PM_{2.5}$) emissions [80,127]. This is driven by the inefficient combustion of raw coal, biomass, or household waste, the dominant heating supply sources for low-income gers, making UB one of the worlds most polluted cities in winter months, with significant associated health issues [128]. Based on a median income of only around 350 USD ger^{-1} month^{-1}, costs for space heating also contribute around 15–20 % of annual ger expenditures [80,129]. During winter, up to half of their limited monthly income may be spent on meeting heating demand alone, prioritising the cheapest source available.

Given the severity of and urgency of the air pollution issue, most previous works on ger energy systems seek to address decarbonisation of the heating supply under the extreme climatic conditions typically experienced [80,120,130]. Traditional coal-stoves are used in most gers, thus these studies normally focus on either benefits from using improved stove designs [131,132] and/or better thermal insulation [80,130]. Partly, this is because these works are centred on ger areas in the outskirts of UB where the grid capacity is a limiting factor to alternative electricity-based heating solutions, and the pollution is particularly severe. Most recently, however, [133] presented the experimental investigation and simulation of a solar PV generator combined with an EH for a single ger in UB, as compared to one using a traditional coal stove. The results from the study over a single year suggest a transition to electricity-based heating supply system can lead to improved indoor environmental quality, cost reductions, and significantly lower heat energy consumption.

4.2. Local context constraints

Given the mean monthly coal expenditure reported by [80], the total energy systems max expenditure $C^m_{es,max,ger}$ is limited to USD 110 month^{-1} ger^{-1}. This results in approximately USD 20–60 month^{-1} ger^{-1} available for capacity expansion, depending on the respective coal expenditure, formulated at the entire 18-ger energy system level.
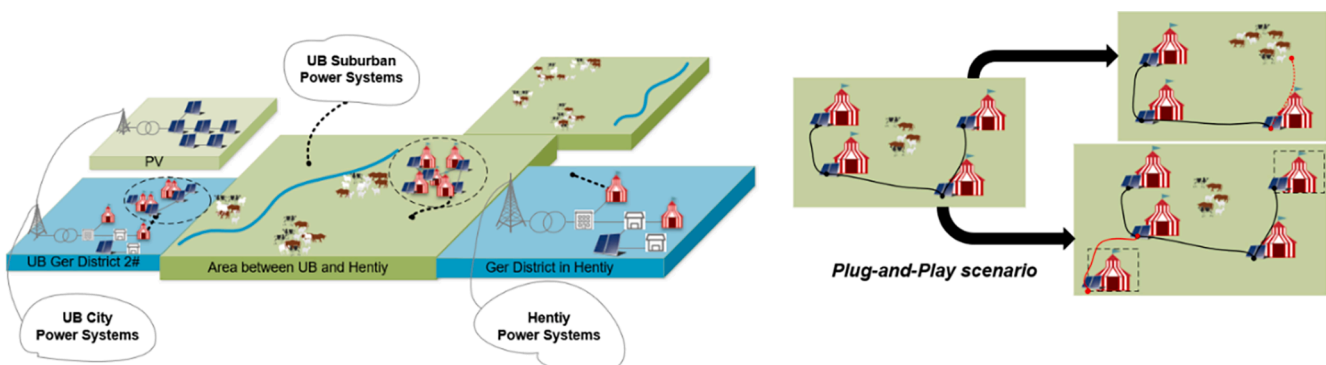

Fig. 7. Spatial boundaries for typical migration pattern (left) and example multi-ger PP configurations (right).

Furthermore, it is well established that a typical 5-wall Mongolian ger weighs 250–400 kg, with an additional 75–100 kg for a standard coal stove and at least one 20 kg bag of raw coal normally used during migration [120,130]. Considering the above, an upper limit on total energy system weight $M_{es,max,ger}^{m}$ of 120 kg ger^{-1} is preliminarily estimated.

4.3. Herder migration

The assumptions used to model ger migration patterns and spatial dispersion uncertainty were developed and validated through direct engagement with a team of local consultants [72] and an interactive workshop with regional energy sector stakeholders [134]. Consequently, it was identified that 18 gers is representative of a typical ger nomadic community. Recent studies suggest that the average area occupied by such a community over the migration season is around 147 km^2, leading to an average seasonal cluster radius of **6.85 km** at project start [72,121]. Furthermore, previous works suggest climate change is likely to lead to growth in overall migration distance and spatial dispersion over the 30-year horizon [119]. With the aim to capture how this uncertainty may influence optimal energy system configurations, mean annual growth rate and volatility are preliminarily estimated based on available data. The annual cluster radius $r_{cluster}^{y}$ evolutions are thus simulated according to (24)–(25). The resulting potentially large difference between projected (dashed line) and realised (solid lines) cluster spatial dispersion is shown in Fig. 8, capturing the uncertainty space considered.

4.4. Electricity demand (ED)

The baseline daily electricity load profile for a typical ger household was estimated through interactions with a local consulting team [72], given the limited literature data available. These values suggest most nomadic households continue to follow a more traditional way of life for now, typically only consuming enough electricity to operate a 20 W lamp and 80 W refrigerator. In winter months, however, the local climatic conditions as well as economic constraints mean that the refrigerator is not needed nor used. This assumption is maintained throughout project horizon, with starting electricity demand estimates of roughly 63 and 6 kWh month^{-1} ger^{-1} during the summer and winter months, respectively. Nonetheless, this load is likely to grow significantly over the next 30 years, with recent ones experiencing a 5–7 % annual increase

in electricity demand, driven by a combination of more productive uses of electricity integrated into nomadic lifestyles, higher indoor environmental standards, and the adoption of low-cost electric appliances [120,123,135].

Baseline load profile projections are thus formulated based on the integration of a 213 W TV with a 30 W antenna, more light bulbs, and a number of additional electric appliances (i.e., radios, cell-phone) by years 15 and 30, respectively [72]. The stochastic electricity demand model is defined from (24)–(25). The baseline daily load profiles for summer days are shown on the left in Fig. 9 below at the starting point (Year 0), midpoint (Year 15) and end (Year 30) of the project. Long term evolutions and variation under uncertainty for different seasons, on the other hand, are captured on the right in Fig. 9. For power flow calculations in Appendix I, it is assumed that these same hourly load profile curves are maintained, the magnitude of which is scaled proportionally to the actual realized monthly demand as compared to the one projected at time of writing.

4.5. Heating demand (HD)

As a baseline for estimating heating demand (HD), it is assumed that all 18 gers in the systems evaluated have 5 walls with standard insulation. While the benefits of improved insulation for gers have been discussed in several other studies [129,130,133], the standard insulation case defined in these works is still representative of the majority of ger dwellings. Seasonal demand calculations are obtained through consultation with a local team [72] and are aligned with values found in the literature [80,129,130,133]. Respective demand-temperature correlations are then used to extrapolate monthly profiles based on the assumptions given in Appendix II, resulting in starting demand of 11,500 kWh/ year per ger. Volatility for heating demand is assumed to be approximately equivalent to that on year-to-year climatic conditions, found via standard regression analysis of the average *winter* temperatures between 1901 and 2020 [136]. The stochastic HD model is thus formulated yearly according to (24)–(25), including potential climate change induced growth [137–139]. The impact of this evolution and differences in potential Year 30 HD profiles is shown for 3 stochastic scenarios in Fig. 10. As for ED above, the seasonal profile *distribution* is maintained across years simulated, and the value in each timestep scaled proportionally to the one at project start [72].
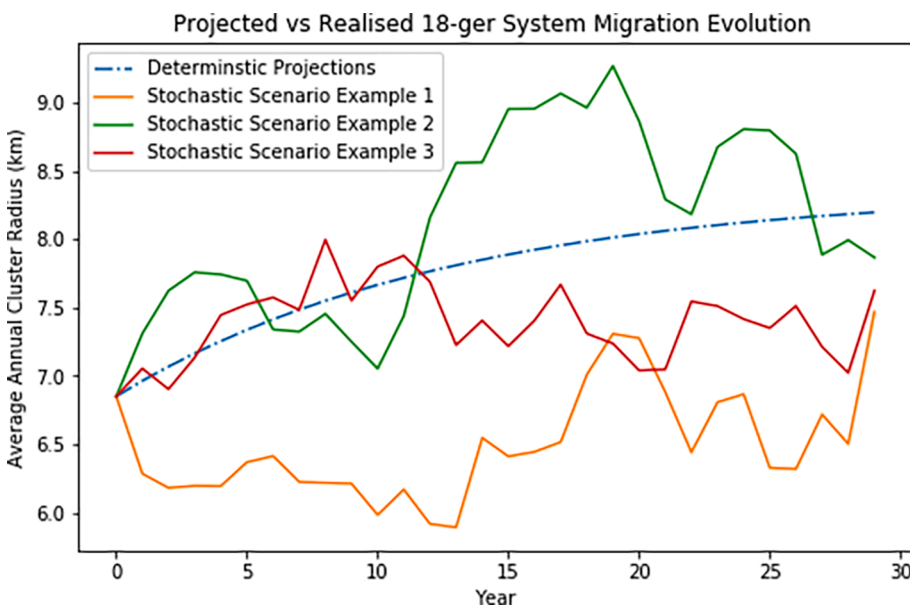

Fig. 8. Deterministic projections (**dashed line**) and 3 randomly selected example stochastic simulated scenarios (**solid lines**), showing the average annual (over the migration season) 18-ger cluster radius in km (y-axis), over the 30-year project horizon (x-axis). The projected migration growth is assumed to be driven by the climate change effects, as shown by the deterministic projections(**dashed line**), although it is recognized the magnitude of this impact is highly uncertain both in its long-term evolution and intra-year variability, as captured by the different stochastic example scenarios(**solid lines**).
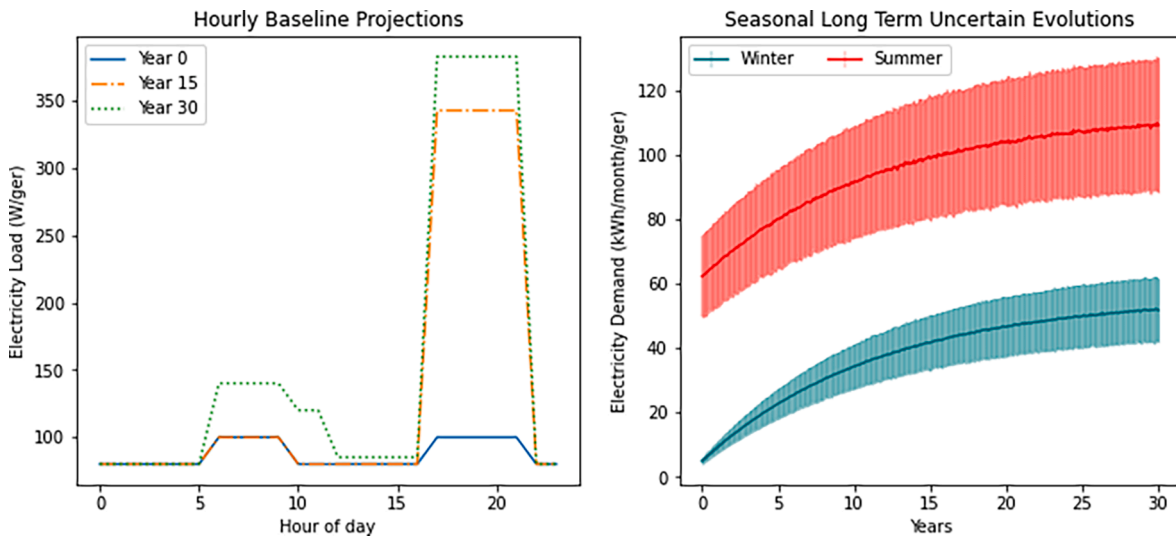
Fig. 9. Hourly projected load (left) and seasonal long term stochastic evolutions of electricity demand (right).
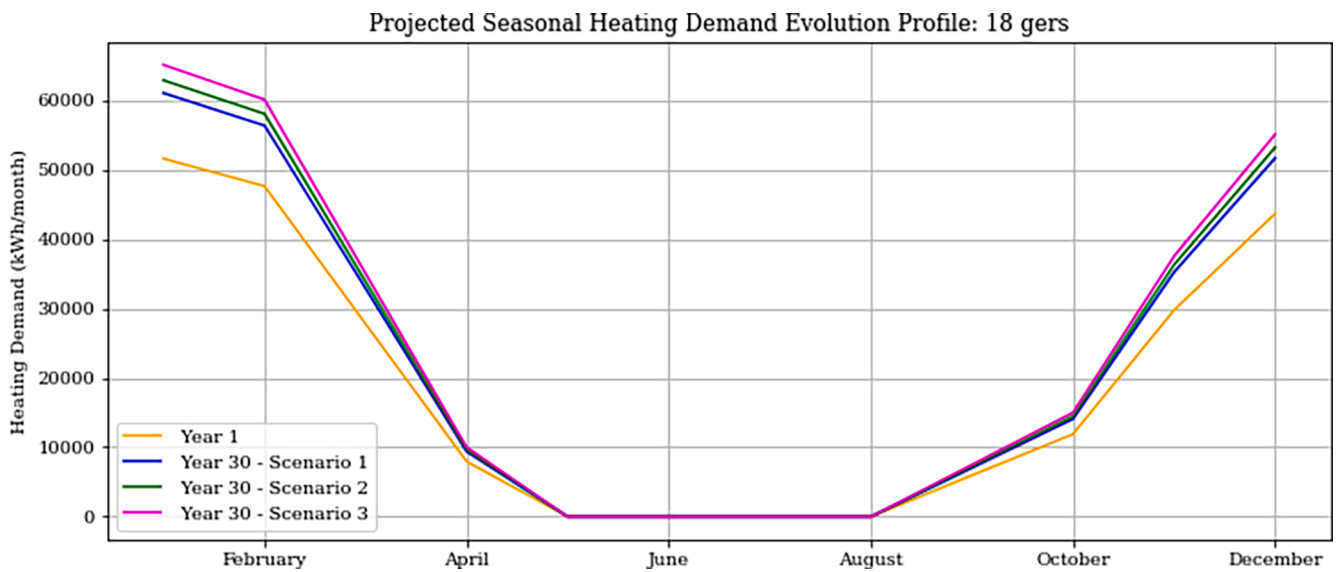

Fig. 10. Projected starting and end of project life heating demand monthly profiles for 3 scenarios.

4.6. Renewable energy generation

For the system sizes and representative locations considered in this manuscript, a mean annual capacity factor (CF) of 19.8 % for solar PV is obtained, similar as reported in [133]. The initial estimate for annual wind capacity factor is slightly higher at 22.2 %. This is lower than values from other wind generation sites in Mongolia, as the small-scale systems selected are operating at reduced wind speeds and conversion efficiencies compared to utility level ones [126,135]. Still, both CF estimates are consistent with those which could be extrapolated from alternative data sources [140,141] or with previous studies [122,135,142].

Please refer to Appendix II for the complete list, values and ranges of the parameters used in this study.

5. Results

5.1. Summary

The energy system design alternatives discussed in the remainder of

Table 1
Summary of ger energy system design alternatives evaluated in this study.

Ger Energy System Alternative	Description
Flexible Design (FD)	18-ger multi-energy PP *microgrid* design and adaptability strategy
Flexible Nanogrid Design (FND)	DRL based single nanogrid unit design and adaptability strategy
Baseline Design 1 (BD1)	Single nanogrid unit stochastically optimal rigid design with no EH
Baseline Design 2 (BD2)	Single nanogrid unit stochastically optimal rigid design with EH option
Baseline Design 3 (BD3)	18-ger microgrid stochastically optimal rigid design with no EH
Baseline Design 4 (BD4)	18-ger microgrid stochastically optimal rigid design with EH option

this section are summarised in Table 1 below, aiming to capture the benefits of flexibility and a systems approach to planning for ger communities. The stochastically optimal rigid baselines results suggest the deployment of a SHS system for BD1 ($\theta_{pv}^{BD1} = 525$ W/ger). This is approximately 5–10 times with what is currently available to herders, or roughly twice that obtained using commercial software such as HOMERPRO, which neglects load growth. Interestingly, the optimal PV capacity in BD2 is equal to that found in BD1, although with a roughly matching EH capacity (625 W/ger) for thermal energy supply. In BD2, therefore, the EH units are run almost exclusively with power purchased from the UB grid. BD3 and BD4 present very similar initial design capacities to their counterparts which are optimised for no PP connection (BD1 and BD2), albeit with reductions in BESS requirements of roughly 8 % and 10 %, respectively. System evolution for FD, on the other hand, varies based on scenarios, although with final capacities generally much greater (and diversified) than those obtained via rigid baselines.

The combined relative *out of sample* performance of the 18-ger energy system design alternatives optimised using the model from Section 2 are summarised by Table 2 below. They show that the FD yields improvements across all metrics estimating economic feasibility, sustainability, and energy security of alternatives, highlighting the importance of embedding capacity expansion flexibility under uncertainty. Note that, as presented in Section 3.3, *VoF* refers to improvements achieved by FD for each metric as compared to the *best performing benchmark design on the same metric*. Furthermore, to capture decision makers with differing risk-profiles, Value at Risk (VaR) and Value at Gain (VaG) are evaluated for each solution, tailored to risk-averse and risk-seeking operators, respectively. These results suggest that FD may be highly valuable for ger communities' electrification along different risk and utility metrics, as discussed in more detail below.

5.2. DRL flexible system evolution strategy

The FD produced via the proposed DRL based approach suggests the development of a highly dynamic decision-making policy, with capacity evolution based on uncertainty realisations in each scenario. The modular nature of the adaptability strategy is graphically captured in Fig. 11 below, showing installed nominal capacities over time for different technological components. The plot suggests that in early years the DRL policy presents limited variability, and largely mirrors the findings from the inflexible baselines, with deployment of PV and BESS. Over time, however, it grows increasingly dynamic as more RES capacity is installed within the energy system. This is primarily driven by expansion of wind technologies after about year 7, which then also allows for more EH capacity to be sustainably and reliably operated. Furthermore, in line with near term projections likely to be more accurate, the policy grows more variable over the project life, adapting to the magnitude of uncertainty encountered.

The planning strategy obtained can be further decomposed for improved interpretability by evaluating actions selected under different conditions. Fig. 11, for instance, could be broken down into the expansion timing distribution for each technology, and used to recommend options for exercising flexibility at various stages of the project. Other metrics can be developed for this purpose, through a mix of intuition, domain expertise, and trial and error. The ratio of installed renewables to EH nominal capacity provides such an example, as a

pattern can be observed where this ratio increases at first, then starts to decrease around year 12 and stabilizes at **around 4.2 to 5.1** in the final periods. The behaviour of the DRL agent under different shortage levels also provides some interesting insights, as shown in Fig. 12. It indicates that it is worthwhile to prioritise RES expansion (and wind in particular) under high shortage levels, although important to achieve a balanced generation mix over time. The results also imply that while the majority of EH expansion should happen when there is no unmet load, some capacity should be deployed even in early high shortage years. Most likely, this is because there could still be periods of high EH utilization rate within those years, which can lead to significant cost savings. Energy storage, on the other hand, seems to be prioritized in years with low unmet load levels. This suggests the DRL agent estimates BESS to be particularly helpful at smoothing out those smaller power imbalances, created by intermittent RES generation or stochasticity introduced into the load profiles.

5.3. Relative performance evaluation

5.3.1. Expected energy system cost

The low-income communities targeted by these studies motivate prioritising cost as an initial objective. FD achieves significant improvements in this front particularly as compared to BD1 and BD2, reducing ENPC by 25.3 % and 20.7 %, respectively. The PP enabled inflexible solutions BD3 and BD4 perform slightly better due to the benefits of interconnection, although FD still results in an ENPC 23.3 % lower than BD3 and a 15.1 % reduction from BD4. Expected cost reductions compared to HOMER deterministic design are even more striking at around 41 %, although highly dependent on the economic value of energy not served c_{ul} estimated earlier. Reduction in energy system cost in any given scenario are achieved by FD through an improved adaptability strategy under uncertainty. More practically, this results in decreased expenditure on coal over time, reduced charges due to unmet load, and discounting of capital costs through modular deployment of components, as shown in Fig. 11. The greatest economic VoF is recorded for VaR, with FD achieving over a 30 % reduction in downside costs as compared to alternatives investigated, although significant improvements are also found in terms of upside potential (VaG). Furthermore, as the capacity expansion strategy developed is highly dynamic and dependant on different realisations of uncertain variables, FD performs much more consistently and is stochastically dominant across *nearly* all scenarios evaluated. The VoF estimated for each economic metric is reported in Table 2.

5.3.2. Lifetime emissions

Even when formulated for purely cost minimisation objectives, the proposed approach is found to yield significant improvements in system design sustainability, as captured by Fig. 13. Reductions in estimated lifetime emissions of 54.6 %, 67.5 %, 52.9 % and 56.6 % are obtained compared to BD1 and BD2, BD3 and BD4, respectively. Interestingly, the static designs evaluating EH (BD2 and BD4) yield higher net emissions than the others, driven by the UB grid carbon footprint, in line with previous work [80]. This suggests a sustainable ger energy system solution must not rely on the national infrastructure available, but rather on the development of a more distributed and adaptable network, particularly to integrate the multi-energy boundaries encountered. The

Table 2
Design decision making table for alternatives evaluated in this study over 2000 out of sample scenarios.

| Metric	Objective	BD1	BD2	BD3	BD4	FND	FD	VoF
ENPC (USD)	*Economic Feasibility*	439,212	413,771	428,088	386,459	373,503	328,178	**85,593**
VaR, 5 % (USD)	*Economic Risk*	492,314	516,332	480,058	473,062	419,512	367,421	**124,893**
VaG, 95 % (USD)	*Economic Upside*	381,456	370,591	365,221	352,669	338,229	293,404	**77,187**
Lifetime Emissions (t. CO$_2$ Eq.)	*Sustainability*	10,133	14,169	9,808	13,375	4,937	4,601	**5,532**
Expected TUL (kWh)	*Energy Security*	29,992	29,992	28,102	27,171	8,274	7,188	**22,804**
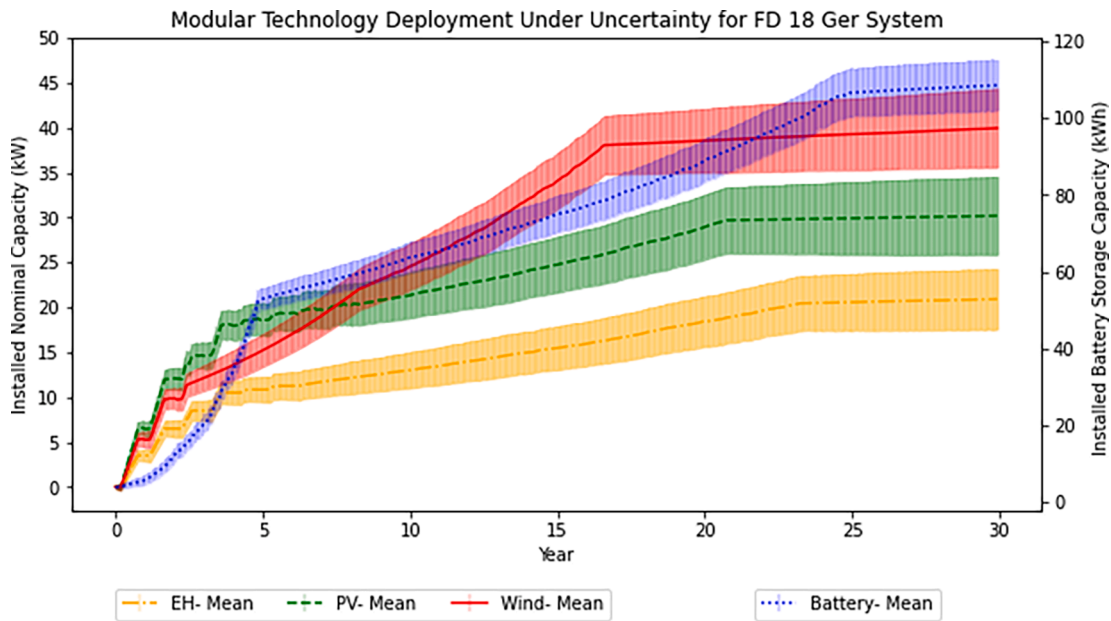
Fig. 11. Cumulative System Capacity Expansion averaged over 2000 episodes showing 95% confidence interval.
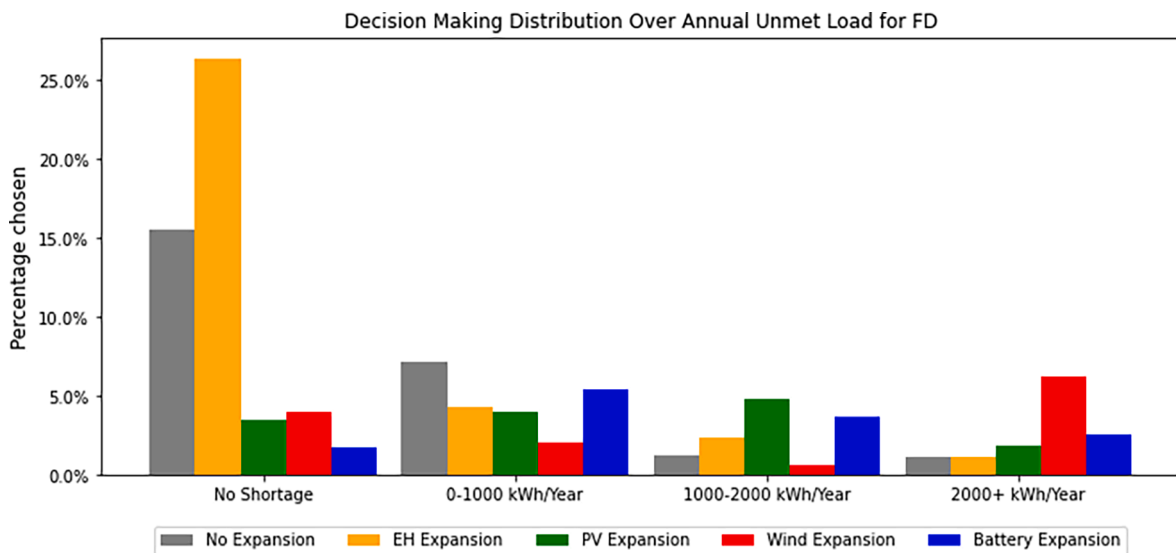

Fig. 12. Distribution of capacity expansion decisions vs shortage levels over 2000 episodes for the DRL agent.

respective contributions of coal and EH to meeting thermal energy demand for FD over time, shown in Fig. 13, suggest that most annual carbon footprint reductions can be achieved around year 18. Nonetheless, some emissions are maintained for the entire project duration, as supplying heat with coal or electricity from the UB grid in certain timesteps can reduce costs. The VoF in the context of sustainability is, in essence, the expected avoided 30-year emissions through the FD as compared to the any of rigid alternatives investigated, estimated at 5,532 tonnes CO_2 eq. for this system.

5.3.3. Energy security and resilience

Outside of improved economic performance and sustainability, FD also achieves substantially enhanced energy security, with reductions in expected TUL of 76.1 % compared to BD1 and BD2, and 74.4 % and 73.5 % from the TUL values obtained in BD3 and BD4. This is driven by the DRL based flexibility strategy, which can closely match capacity and

demand over time for the required provision of energy services. The increased generation capacity for FD also reduces reliance by gers on the UB grid, which can be highly unstable due to frequent load shedding. The PP interconnection allows for the smoothing out of load/generation imbalances as compared to the independent nanogrid units, helping increase system resilience for all stakeholders involved, while also reducing BESS requirements. The VoF for energy security is estimated at 22,804 kWh on average for the duration of the project. Intuitively, this implies that for a simulated 30-year horizon, 22,804 kWh of unmet load could be expected to be avoided by using the FD as opposed any of the alternatives considered. Noting that while some unmet load is still recorded for FD, this results from the allowable capacity shortage presented in 2.4 and the intrinsic stochasticity of the environment investigated, as it would require substantial oversizing of components to consistently meet all loads.
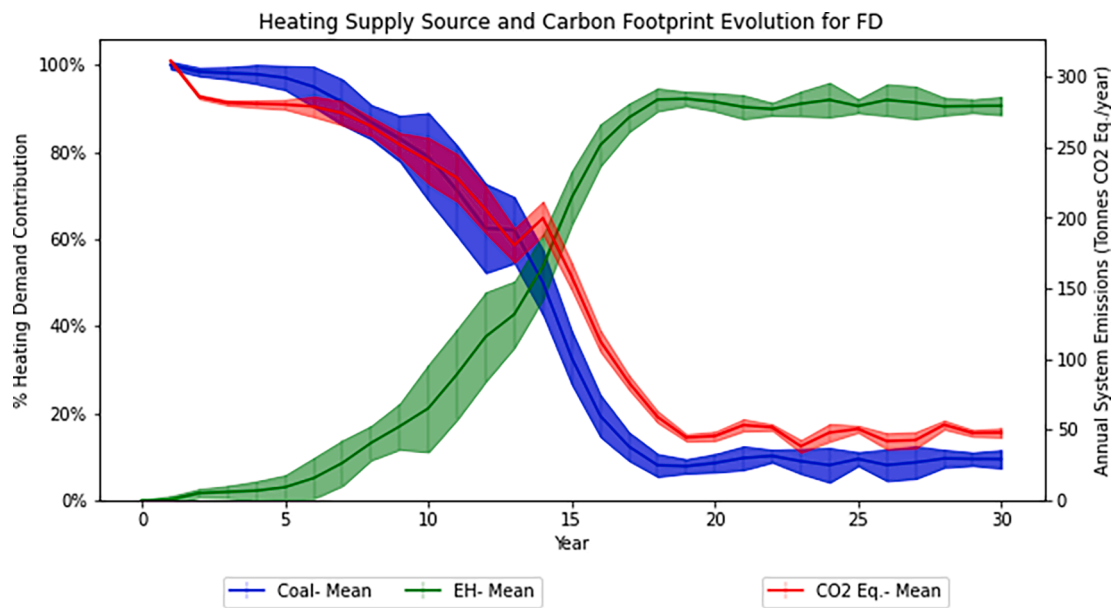
Fig. 13. Heating supply source (left-axis) vs carbon footprint evolution (right-axis) for FD system over 2000 simulated scenarios showing the 95% confidence interval.

5.3.4. Local context constraints

Average energy system expenses (excluding unmet load charges) are around 83 USD month^{-1} ger^{-1} for FD, and substantially lower for BD1 and BD2 at around 52 and 46 USD month^{-1} ger^{-1}, although with significant variation across uncertainty scenarios. The number of timesteps where the budgetary constraints are violated, however, are significantly lower on average for FD. On one hand, this is driven by distributing required investments over project life as compared to significant upfront expenses (and associated replacement costs) recorded for baseline designs. As coal expenses are reduced over time, the budget for expansion decisions for FD increases, which presents nice synergies with the projected load growth, allowing for the deployment of generating technologies only *if and when* needed. Moreover, the PP enabled FD reduces reliance on BESS, which is both the costliest and heaviest technology considered, helping to limit total system mass.

Accordingly, energy system weight for FD is found to be on average around 15 % and 19 % lower than BD1 and BD2 for the first half of the project, respectively. The much higher final generating capacities installed for FD and extra mass associated with cabling, however, lead to a net increase in end of project system weight of roughly 28 % and 22 % as compared to BD1 and BD2. BD3/4 usually result in slightly higher weights than BD1/2, but still within the acceptable bounds. Notably, energy system evolution and expansion possibilities over time are a result of reduced coal mass requirements via EH substitution, similar as what is observed for budget. Given the very large amounts of coal consumed by a typical ger, and the availability of regional lightweight systems, mobility constraint violation is thus very unlikely in FD. The few recorded instances can be attributed to energy generation and demand uncertainty as well as potential load shedding, which can result in additional coal mass requirements from that originally estimated.

5.4. Value of Plug and Play (VoPP) operation

In this section, the VoPP is decomposed from the VoF reported in Table 2 to help assess its economic feasibility, according to (39). The results shown in Table 3 suggest it is a worthwhile investment in all cases. These values may be intuitively thought of as the difference in total cost between the same energy system configurations if integrating

PP operation or not. FD benefits the most from PP, followed by BD2/4, as the explicit consideration of EH increases system load substantially, and thus motivation for interconnection compared to BD1/3. Nonetheless, for the targeted low-income communities, VoPP for BD1/3 is still significant and can likely be further enhanced. These findings highlight the crucial role of PP in enabling economically feasible mobile energy system solutions, with cost reductions driven by a combination of reduced storage requirements, unmet load penalty and electricity purchased from grid. At the same time, they suggest that a systems approach and flexibility strategy under uncertainty are essential to unlock its full potential, with significantly greater VoPP for FD recorded across all objectives evaluated.

It is important to note, however, that this estimated value is highly dependent on the assumptions used, particularly in terms of herder migration modelling. Significantly greater spatial dispersions than the ones investigated could lead to unsustainable system losses from low voltage distribution, the costs of which may outweigh the benefits generated by PP. To further investigate the effect of this, average distribution losses are plotted against the range of potential cluster radiuses encountered in Fig. 14. The "break-even" *starting* cluster radius (where VoPP = 0) is estimated at roughly **12.1 km** for the 18-ger system FD through the economic value of unmet load c_{ul}. This implies that in regions with greater spatial dispersions among herders, higher distribution voltages should be considered to maintain an acceptable network efficiency. Alternatively, government policy programs or decision support tools could be developed to incentivize nomadic households to settle closer to each other, helping to capture the full benefits enabled by PP.

Table 3
Estimated VoPP and externalities on energy security and sustainability for different system design alternatives.

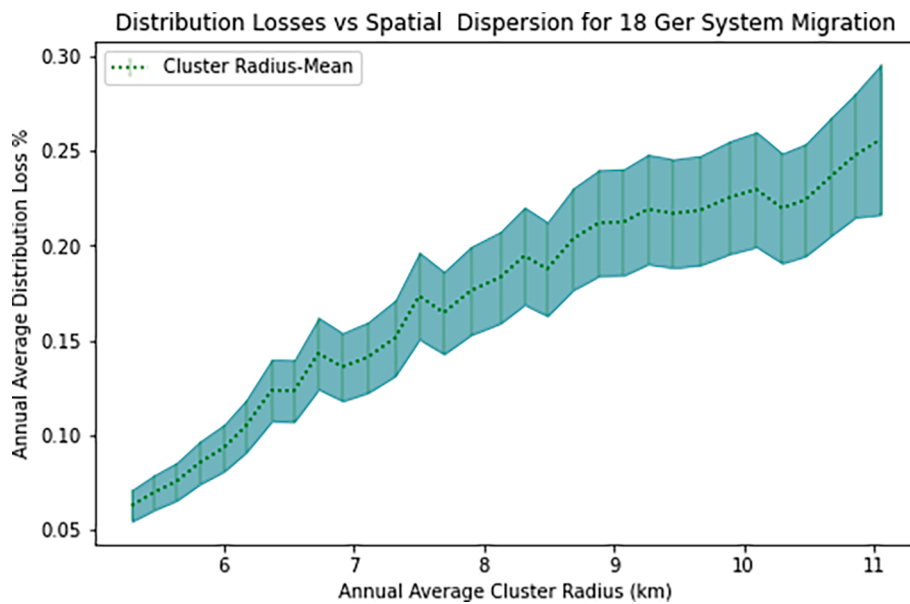| Design(s)	VoPP (USD)	% Δ TUL from PP	% Δ CO_2 from PP
FD/FDN	**45,313**	−15.1	−7.3
BD1/3	**11,134**	−6.3	−3.2
BD2/4	**27,312**	−9.4	−5.6

Fig. 14. Annual distribution loss % as a function of herder PP microgrid spatial dispersion.

6. Discussion

6.1. Changing economies of scale

Flexibility in Design is generally motivated by tensions between EoS and modular adaptable deployment, where each can be advantageous depending on the setting. The results presented in the former section, however, suggest that the flexible solution is stochastically dominant and outperforms the baseline designs in nearly all scenarios evaluated. This may be due to the relatively low magnitude of EoS assumed in the cost calculations. Coal expenses, even if significantly reduced, remain the strongest driver of ENPC for nomadic energy systems in Mongolia. While evolving costs for other technologies may influence this finding, an important advantage of the proposed DRL approach is precisely the simplicity to retrain the agent with updated assumptions as they are verified over time. As such, if governmental programs are developed which influence the optimal solution (i.e. subsidy schemes, carbon credits, etc.), they could be quickly integrated into the model to give up to date recommendations. Finally, it is worth noting that removing the budgetary constraints formulated could lead to the discovery of better performing solutions which take advantage of both EoS and flexibility. While this constraint was relaxed in the zero-stage capital expense decision, there are likely to be some practical, logistical and/or economic benefits to ramping up and concentrating investment in certain time periods.

6.2. Decarbonisation pathways

The trade-offs between costs and emissions for inflexible ger energy systems not considering PP were further evaluated through the Non-dominated Sorting Genetic Algorithm (NSGA-II) to the optimisation of BD1 and BD2 [143]. The pareto frontier obtained suggests that high EH capacities should be deployed even by inflexible nanogrid units seeking to minimise ENPC, as they help reduce coal expenses over time - instead purchasing heavily discounted (particularly at night) electricity from the UB grid. The associated carbon footprint, however, means those solutions are estimated to lead to a net increase in expected system emissions over the project horizon. The results presented in the former sections, in fact, suggest that an economically feasible decarbonisation of ger community energy systems is possible even when accounting for the associated uncertainties, due to several important synergies. Ger communities present substantially higher demand for electricity in the

summer months compared to the winter ones, arising from increased refrigerator energy requirements. This allows RES generation to be utilized almost exclusively for EH in the winter months. The relatively high capital expense of RES and EH for low-income communities, as well cultural value associated with the coal stove, nonetheless, remain important limiting factors to the rate of future feasible decarbonisation in Mongolia.

6.3. PP microgrid feasibility

While this work was focused on the planning rather than operation side, the new control algorithm developments remain largely untested on the field. This is an important issue to address, particularly in terms of overall resulting network efficiency and feasibility at different levels of spatial dispersion. In this study, a ring like interconnection was fixed, however this is likely to vary significantly during actual herder migration, which may strongly impact the performance of a design and flexibility strategy. Additionally, even though only active power was considered here, there are substantial safety and reliability issues associated with low-voltage AC distribution over long distances. Nonetheless, the results from this work imply that there is a significant economic value attached to PP operation, thus motivating the need for more in-depth feasibility studies.

6.4. DRL environment insights for community Decision-Making

The analysis presented in Section 5 already provides some interpretable guidelines that could be used by herder communities to determine their energy system expansion strategy. To capture the full value and dynamicity of the proposed approach, however, a decision support tool could be implemented. This would be targeted at improving data visualisation, understanding of the policy optimisation performed, and ultimately enhancing acceptability. A mobile app could be developed which helps community members track energy system state, explore locations of other nearby PP clusters if migration is required, and dynamically recommend actions over time. Given the communication infrastructure is already present, this could be achieved at a low cost. Furthermore, users with different budgetary or mobility requirements from those assumed in this work could enter that data directly in the app, helping generate solutions better tailored to each customer.

6.5. Limitations and future work opportunities

While this work presents an important starting point to evaluate the planning of future nomadic and mobile energy systems, there are some other limitations which should be noted. Data on load profiles and projections were difficult to find, and several assumptions had to be made to develop the model from Section 2. A higher resolution heat load profile model, perhaps based on outside temperatures in each hour per day, could allow for a more detailed evaluation of different climate change scenarios. The mobility aspect within the energy system simulation is completely novel and the modelling approach implemented hard to validate. Continuing interactions with local energy sector stakeholders and herder communities will be crucial in this regard. Especially valuable insights may be on the flexibility of migration patterns assumed, and willingness to deviate from them to engage in PP operation in different regions. Additionally, the results are highly dependent on realisations of different cost factors over the project life. These were mostly assumed to be invariant in time for simplicity, and for lack of better information, however, are highly likely to differ over the next 30-years. In order to address data limitations at the time of writing, future work ought to investigate reproducibility of solutions when using different stochastic model parameters. Considering the possibility of new bulk procurement programs, it may be interesting to policy makers to perform a more in-depth sensitivity analysis on the impact of EoS in Mongolia. This can suggest what the break-even values may be, or where the rigid designs outperform the flexible ones.

The action space for the DRL agent could be significantly expanded from this implementation to reflect the relevant decision-making possibilities. Multiple technologies could be considered for expansion at once, for instance. Moreover, connection to the UB or a nearby grid could be explicitly modelled, and respective recommendations produced. Several other potential actions were not evaluated here, particularly higher efficiency stoves or improved ger thermal insulation, which could be highly valuable for decarbonisation objectives. Inverter capacity was also not explicitly analysed, thus likely leading to sub-optimal operation at different points of the project. Closer integration of the control and planning aspects of this problem should also be investigated in future work. While the computational expense required may make the problem intractable, it would be valuable to actually simulate PP operation at a finer time resolution over the entire project horizon. Alternatively, a separate DRL agent could be used to develop more advanced operational strategies than the LFS implemented here. Flexible demand response and prioritised load shedding, for instance, are likely to help reduce ENPC incurred by mobile energy systems.

From an algorithmic point of view, there are several more interesting avenues for future research. More specifically, a multi-agent formulation could be developed to better evaluate the costs and benefits to individual herder households, and the resulting optimal flexibility strategies. Recent advancements in distributional DRL allow the formulation of a policy based on the entire distribution of rewards rather than purely expected value [144]. Recommended strategies could thus be tailored to risk-seeking or risk averse decision makers, improving acceptability, and understanding. Finally, exploring alternative DNN architectures which are better tailored to the tackled problem could lead to further improvements. CNNs could be used to map relationship among features which may not be intuitive at first, for instance, while RNNs could help capture the seasonal nature of decision making required in order to better adapt agent policy between winter and summer months.

7. Conclusions

In this paper, a DRL-based approach was developed for the design and planning of a mobile energy system supply system, and illustrated through a case study on Mongolian herder communities. Motivated by the potential for increased climate induced migration and increasingly decentralized generation, a novel modelling framework to integrate mobility considerations was developed. This analysis was enabled by recent "Plug and Play" control developments, allowing temporary interconnection among nomadic households and/or the main grid. Contrary to popular energy system planning methods, the data-driven approach implemented allowed for the tractable integration of several sources of uncertainty as well as mobility within the evaluation model. Both heat and electricity were considered, leading to a more holistic evaluation of different solution alternatives than previously available. The design and planning strategy for a highly flexible and modular energy system was thus optimised through an actor-critic algorithm, with a reasonable computational expense. Benchmark inflexible designs with economies of scale were compared across key economic, sustainability and resilience indicators such as Cost, Equivalent Emissions and Total Unmet Load. The results suggest that the proposed approach can lead to economically feasible mobile energy systems, even in the case of budgetary and mobility constraints. They also show that a DRL-based flexible capacity expansion strategy can offer a highly dynamic and adaptable energy system design, with multi-objective improvements compared to rigid baselines available from institutional programs. Additionally, the value generated by "Plug and Play" operation was also estimated using a variation of real-options theory, suggesting it is a worthwhile investment in all cases. The spatial dispersion among herders within a cluster, however, can be an important limiting factor due to the low voltage of distribution. Finally, while this paper provides an important advancement to enabling nomadic energy access, there are several limitations and opportunities for future work, which could enhance the applicability and overall impact of the methods presented.

CRediT authorship contribution statement

Cesare Caputo: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft, Visualization. **Michel-Alexandre Cardin:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Pudong Ge:** Conceptualization, Software, Visualization, Writing – review & editing. **Fei Teng:** Conceptualization, Investigation, Writing – review & editing, Supervision. **Anna Korre:** Conceptualization, Validation, Investigation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Ehecatl Antonio del Rio Chanona:** Conceptualization, Methodology, Validation, Investigation, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

Appendix

Appendix I: Mathematical formulation of power flow model

$$P_{NEL}^{n,t} = ED^{n,t} - P_{res}^{n,t} \tag{41}$$

$$P_{res}^{n,t} = P_{wind,adj}^{n,t} + P_{pv,adj}^{n,t} \tag{42}$$

$$P_{Load,Net}^{n,t} = HD^{n,t} + ED^{n,t} - P_{res}^{n,t} \tag{43}$$

$$P_{HD,eh,res}^{n,t} = \min\left(HD^{n,t}, P_{eh,max}^{n,t}, -P_{ED,Load,Net}^{n,t}\right), P_{NEL}^{n,t} \leq 0 \tag{44}$$

$$P_{NHL}^{n,t} = HD^{n,t} - P_{eh,res}^{n,t} \tag{45}$$

$$P_{res,PL}^{n,t} = P_{res}^{n,t} - P_{HD,eh,res}^{n,t} - P_{NEL}^{n,t}, P_{ED,Load,Net}^{n,t} \leq 0 \; and \; P_{HD,Load,Net}^{n,t} \leq 0 \tag{46}$$

$$n \in [N_1, N_2, N_3, ..N_N], \; \forall n, t \tag{47}$$

$$P_{ba,outmax}^{n,t} = \theta_{ba}^{n,t} C_{out} \tag{48}$$

$$P_{ba,in,max}^{n,t} = \theta_{ba}^{n,t} C_{in} \tag{49}$$

$$P_{ba}^{n,t} = \begin{cases} -\min\left(P_{ba,\,out,max}^{n,t}, P_{NEL}^{n,t}\right), \; P_{NEL}^{n,t} \geq 0 \; (a) \\ -\min\left(P_{ba,\,out,\,max}^{n,t}, P_{HD,Load,Net}^{n,t}\right), \; P_{res,PL}^{n,t}, P_{ED,Load,\,Net}^{n,t} \leq 0 \; ; \; P_{HD,\,Load,\,Net}^{n,t} \geq 0 \; (b) \\ \min\left(P_{ba,\,in,\,max}^{n,t}, P_{res,PL}^{n,t}\right), \; P_{res,PL}^{n,t} \geq 0; P_{HD,\,Load,\,Net}^{n,t} \; ; \; P_{ED,Load,\,Net}^{n,t} \leq 0 \; (c) \end{cases} \tag{50}$$

$$E_{ba}^{n,t} = E_{ba}^{n,t-1} + P_{ba}^{n,t} \sqrt{\eta_{ba}} + E_{ba,new}^{n,t} \tag{51}$$

$$E_{ba}^{n,t} = E_{ba}^{n,t-1} + P_{ba}^{n,t} \sqrt{\eta_{ba}} + E_{ba,new}^{n,t} \tag{52}$$

$$P_{ul,PB}^{n,t} = \begin{cases} P_{Load,Net}^{n,t} - (E_{ba}^{n,t} - E_{ba,min}^{n,t}), \; P_{Load,Net}^{n,t} \geq 0 \\ 0 \;, \; otherwise \end{cases} \tag{53}$$

$$P_{HD,Load,Net,PB}^{n,t} = HD^{n,t} - P_{eh,res}^{n,t} - P_{eh,batt}^{n,t} \tag{54}$$

$$P_{HD,Load,Net,PB,ES}^{t} = \sum_{n \in N} P_{HD,Load,Net,PB}^{n,t} \tag{55}$$

$$P_{eg}^{n,t} = \min(P_{res,PL}^{n,t} - P_{ba}^{n,t}, 0) \;, \; P_{ba}^{n,t} \geq 0 \tag{56}$$

$$n \in N_{deficit} : P_{eg}^{n,t} \leq 0 \tag{57}$$

$$n \in N_{surplus} : P_{eg}^{n,t} > 0 \tag{58}$$

$$P_{ul,es}^{t} = \sum_{n \in N_{deficit}} P_{ul,PB}^{n,t} + P_{HD,Load,Net,PB}^{n,t} \tag{59}$$

$$P_{uel,es,bpp}^{t} = \sum_{n \in N_{deficit}} P_{ul,PB}^{n,t} \tag{60}$$

$$P_{eg,es}^{t} = \sum_{n \in N_{surplus}} P_{eg}^{n,t} \tag{61}$$

$$P_{dist,es}^{t} = \min\left(P_{ul.es}^{t}, P_{eg,es}^{t}, \theta_{PPmg}^{t}, P_{dist,es,max}^{t}\right) \tag{62}$$

$$s.t. \, 0 \leq P_{dist,es}^{t} \leq 0.3\left(ED_{es}^{t}\right) \tag{63}$$

$$L_{dist}^{t,n} = \frac{L_{cb}^{t}}{N} \tag{64}$$

$$P_{dist}^{t,n} = \frac{P_{dist,es}^{t}}{N} \tag{65}$$

$$R_{dist,ref}^{t,n} = \frac{\rho_{cb} L_{dist}^{t,n}}{A_{cs}^{cb}} \tag{66}$$

$$R_{dist,actual}^{t,n} = R_{dist,ref}^{t,n} \left[1 + \alpha_{copper}\left(T_{actual} - T_{ref}\right)\right] \tag{67}$$

$$P_{dl,es}^t = \sum_{n \in N} 3R_{dist,actual}^{t,n} I_{dist}^{t,n}{}^2 \tag{68}$$

$$I_{dist}^{t,n} = \frac{P_{dist}^{t,n}}{V_{dist}^{t,n}} \tag{69}$$

$$P_{res,exc,pp}^t = P_{eg,es}^t - P_{dist,es}^t \tag{70}$$

$$P_{uel,es,ppp}^t = P_{uel,es,bpp}^t - \left(P_{dist,es}^t - P_{dl,es}^t\right) \tag{71}$$

$$DL(\%) = 100\frac{P_{dl,es}^t}{P_{dist,es}^t} \tag{72}$$

Appendix II: Model parameters and variables values

Please refer to the nomenclature table presented at the beginning of the manuscript for the definition and respective reference for each of these variables/ parameters.

Technical Parameters and Values/Ranges			
T	30 years / 360 months	K_{eh}	1.52 USD/W
λ	5 %	K_{ba}	0.7 USD/ Wh
η_{inv}	90 %	α_{MS1}	0.95
η_{batt}	75 %	α_{MS2}	0.85
$\eta_{coalstove}$	25 %	$\varepsilon_{pe,avg}$	0.769
μ_{coal}	14.6 MJ/kg	k^{salv}	0.7
$A_{cs,cb}$	1.5 mm^2	$L_{pv}{}^{max}$	25 years
ρ_{cb}	1.678x10^{-8} Ωm	$L_{wind}{}^{max}$	20 years
T_{ref}	20 °C	$L_{pe}{}^{max}$	15 years
$\alpha_{cb,adj}$	0.393 % / °C	$L_{eh}{}^{max}$	13 years
EF_{grid}	0.711 tCO$_2$ eq./ MWh	c_{coal}	40 USD/ t
EF_{coal}	1.37 tCO$_2$ eq./ 1000 kg	$c_{om,pv}$	0.0048 \$/kW/month
θ_{UBGrid}	2 kW	$c_{om,wind}$	0.0032 \$/kW/month
A_{grid}	1	$c_{om,pe}$	0.0001 \$/kW/month
B_{grid}	9	c_{ul}	0.3417 USD/kWh
c_{cc}	0.04 USD/ kWh	c_{grid}	0.041 USD/kWh
c_{ic}	5.2 USD/m	p_{grid}	0.17 USD/kWh
$c_{pe,avg}$	0.352 USD/W		
K_{pv}	2.64 USD/W		
K_{wind}	1.91 USD/W		

GBM Model(s) Parameters	
$\mu_{migration}$	0.5 %
$\sigma_{migration}$	5 %
μ_{ED}	3.14 %
σ_{ED}	15 %
μ_{HD}	1 %
σ_{HD}	6.5 %
σ_{PV}	19.3 %
σ_{wind}	14.2 %

References

[1] O, Edenhofer, et al., *Climate Change 2014 Mitigation of Climate Change Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change Edited by.* 2014.

[2] Johnson DL. The nature of nomadism. The nat nomadism 1969.

[3] Ferris E. Research on climate change and migration where are we and where are we going? Migration Studies 2020;8(4):612–25.

[4] Piguet E, Pécoud A, De Guchteneire P. Migration and climate change: an overview. Refug Surv Q 2011;30(3):1–23.

[5] T.W, HOUSE, *Report on the Impact of Climate Change on Migration*, W. House, Editor. 2021: Online.

[6] K.K, Rigaud, et al., *Groundswell.* 2018.

[7] A, Subbiah, et al., *Addressing developmental needs through energy access in informal settlements.* Field Actions Science Reports. The journal of field actions, 2016 (Special Issue 15): p. 80-91.

[8] Iea. WEO-2017 Special Report: Energy Access. Outlook 2017.

[9] World Bank, I., *SE4ALL.* 2020, DataBank.

[10] Hoffmann R, et al. A meta-analysis of country-level studies on environmental change and migration. Nat Clim Chang 2020;10(10):904–12.

[11] Melde S, Laczko F, Gemenne F. Making mobility work for adaptation to environmental changes: results from the MECLEP global research. Geneva, Switzerland: International Organization for Migration; 2017.

[12] Nations, U., *The Millennium Development Goals Report 2012.* 2012, UN.

[13] Sayar F. State of the global mini-grids market report 2020. J Chem Inform Model 2019;53(9):1689–99.

[14] Mohammed A, et al. Ac microgrid control and management strategies: evaluation and review. IEEE Power Electron Mag 2019;6(2):18–31.

[15] K, Ehn, A. Jorge, and M. Marques-Pita, *Digital Nomads and the Covid-19 Pandemic: Narratives About Relocation in a Time of Lockdowns and Reduced Mobility.* Social Media+ Society, 2022. **8**(1): p. 20563051221084958.

[16] Ge P, Chen B, Teng F. Event-triggered distributed model predictive control for resilient voltage control of an islanded microgrid. Int J Robust Nonlinear Control 2021;31(6):1979–2000.

[17] P, Ge, et al., *A Resilience-Oriented Centralised-to-Decentralised Framework for Networked Microgrids Management.* arXiv preprint arXiv:2109.00245, 2021.

[18] Werth A, Tokoro M, Tanaka K. Bottom-up and recursive interconnection for multi-layer dc microgrids. IEEE; 2016.

[19] Borazjani P, et al. A review on microgrid control techniques. IEEE; 2014.

[20] Gu W, et al. Modeling, planning and optimal energy management of combined cooling, heating and power microgrid: a review. Int J Electr Power Energy Syst 2014;54:26–37.

[21] Jadav KA, Karkar HM, Trivedi I. A Review of microgrid architectures and control strategy. J The Institution of Engineers (India): Series B 2017;98(6):591–8.

[22] Hartono B, Budiyanto Y, Setiabudy R. Review of microgrid technology. IEEE; 2013.

[23] Wang Y, Rousis AO, Strbac G. On microgrids and resilience: a comprehensive review on modeling and operational strategies. Renew Sustain Energy Rev 2020; 134:110313.

[24] Faraji J, Hashemi-Dezaki H, Ketabi A. Multi-year load growth-based optimal planning of grid-connected microgrid considering long-term load demand forecasting: a case study of Tehran. *Iran* Sustainable Energy Technologies and Assessments 2020;42:100827.

[25] Fioriti D, et al. Multi-year stochastic planning of off-grid microgrids subject to significant load growth uncertainty: overcoming single-year methodologies. Electr Pow Syst Res 2021;194:107053.

[26] Inversin, A.R., *Mini-grid design manual.* 2000.

[27] Ehsan A, Yang Q. Optimal integration and planning of renewable distributed generation in the power distribution networks: a review of analytical techniques. Appl Energy 2018;210:44–59.

[28] Lian J, et al. A review on recent sizing methodologies of hybrid renewable energy systems. Energ Conver Manage 2019;199:112027.

[29] Mandelli S, et al. Off-grid systems for rural electrification in developing countries: definitions, classification and a comprehensive literature review. Renew Sustain Energy Rev 2016;58:1621–46.

[30] Chowdhury SA, et al. Off-grid rural area electrification through solar-diesel hybrid minigrids in Bangladesh: resource-efficient design principles in practice. J Clean Prod 2015;95:194–202.

[31] Boait P, et al. ESCoBox: a set of tools for mini-grid sustainability in the developing world. Sustainability 2017;9(5):738.

[32] Lammers K, Bertheau P, Blechinger P. Exploring requirements for sustainable energy supply planning with regard to climate resilience of Southeast Asian islands. Energy Policy 2020;146:111770.

[33] M, Ozanne, M, Brucoli. *Holistic masterplanning for resilient mini-grids: A case study methodology analysis applied to rural communities in Palestine.* in *2015 IEEE Canada International Humanitarian Technology Conference (IHTC2015).* 2015. IEEE.

[34] Gamarra C, Guerrero JM. Computational optimization techniques applied to microgrids planning: a review. Renew Sustain Energy Rev 2015;48:413–24.

[35] Pecenak ZK, Stadler M, Fahy K. Efficient multi-year economic energy planning in microgrids. Appl Energy 2019;255:113771.

[36] Petrelli M, et al. Multi-year planning of a rural microgrid considering storage degradation. IEEE Trans Power Syst 2020;36(2):1459–69.

[37] Alharbi H, Bhattacharya K. Stochastic optimal planning of battery energy storage systems for isolated microgrids. IEEE Trans Sustainable Energy 2017;9(1): 211–27.

[38] Guerrero RC, Angelo M, Pedrasa A. An milp-based model for hybrid renewable energy system planning considering equipment degradation and battery lifetime. IEEE; 2019.

[39] Zhang Y, et al. Life cycle planning of battery energy storage system in off-grid wind–solar–diesel microgrid. IET Gener Transm Distrib 2018;12(20):4451–61.

[40] Mehrjerdi H. Dynamic and multi-stage capacity expansion planning in microgrid integrated with electric vehicle charging station. J Storage Mater 2020;29: 101351.

[41] Cardin M-A. Enabling Flexibility in Engineering Systems: A Taxonomy of Procedures and a Design Framework. ASME J Mech Des 2014;136(1):1–14. https://doi.org/10.1115/1.4025704.

[42] de Neufville R, Scholtes S.. *Flexibility in engineering design.* MIT Press; 2011.

[43] Martínez-Ceseña EA, Mutale J. Application of an advanced real options approach for renewable energy generation projects planning. Renew Sustain Energy Rev 2011;15(4):2087–94.

[44] Caunhye AM, Cardin M-A, Rahmat M. Flexibility and real options analysis in power system generation expansion planning under uncertainty. IISE Trans 2022. https://doi.org/10.1080/24725854.2021.1965699.

[45] Abdin AF, Caunhye A, Zio E, Cardin M-A. Optimizing generation expansion planning with operational uncertainty: A multistage adaptive robust approach. Appl Energy 2022;306:1–18. https://doi.org/10.1016/j.apenergy.2021.118032.

[46] Fernandes B, Cunha J, Ferreira P. The use of real options approach in energy sector investments. Renew Sustain Energy Rev 2011;15(9):4491–7.

[47] Perera A, Kamalaruban P. Applications of reinforcement learning in energy systems. Renew Sustain Energy Rev 2021;137:110618.

[48] Perera ATD, et al. Introducing reinforcement learning to the energy system design process. Appl Energy Elsevier 2020:114580.

[49] Caputo C, Cardin M-A. Analyzing real options and flexibility in engineering systems design using decision rules and deep reinforcement learning. ASME J Mech Des 2022;144. https://doi.org/10.1115/1.4052299.

[50] Martínez-Costa C, et al. A review of mathematical programming models for strategic capacity planning in manufacturing. Int J Prod Econ 2014;153:66–85.

[51] Copeland T, Antikarov V. Real Options: A Practitioner's Guide. New York. NY, United States.: Thomson Texere; 2003.

[52] Cardin M-A, Xie Q, Ng TS, Wang S, Hu J. An approach for analyzing and managing flexibility in engineering systems design based on decision rules and multistage stochastic programming. IISE Trans 2017;49(1):1–12. https://doi.org/10.1080/0740817X.2016.1189627.

[53] Andriotis C, Papakonstantinou K. Managing engineering systems with large state and action spaces through deep reinforcement learning. Reliab Eng Syst Saf 2019; 191:106483.

[54] Degrave J, et al. Magnetic control of tokamak plasmas through deep reinforcement learning. Nature 2022;602(7897):414–9.

[55] A, Haj-Ali, et al., *A View on Deep Reinforcement Learning in System Optimization.* 2019.

[56] H, Mao, et al. *Resource management with deep reinforcement learning.* in *Proceedings of the 15th ACM Workshop on Hot Topics in Networks.* 2016.

[57] Vázquez-Canteli JR, Nagy Z. Reinforcement learning for demand response: a review of algorithms and modeling techniques. Appl Energy 2019;235:1072–89.

[58] Mnih V, et al. Human-level control through deep reinforcement learning. Nature 2015;518(7540):529–33.

[59] Yonekura K, Hattori H. Framework for design optimization using deep reinforcement learning. Struct Multidiscip Optim 2019;60(4):1709–13.

[60] EPA, Scope 1 and Scope 2 Inventory Guidance, WRI,, editor. Greenhouse Gas Protocol: WASHINGTON. USA: D.C; 2020.

[61] Pogaku N, Prodanovic M, Green TC. Modeling, analysis and testing of autonomous operation of an inverter-based microgrid. IEEE Trans Power Electron 2007;22(2):613–25.

[62] Ge P, et al. Resilient secondary voltage control of islanded microgrids: an eskbf-based distributed fast terminal sliding mode control approach. IEEE Trans Power Syst 2020;36(2):1059–70.

[63] Riverso S, Sarzo F, Ferrari-Trecate G. Plug-and-play voltage and frequency control of islanded microgrids with meshed topology. IEEE Trans Smart Grid 2014;6(3): 1176–84.

[64] Nasir M, et al. Solar PV-based scalable DC microgrid for rural electrification in developing regions. IEEE Trans Sustainable Energy 2017;9(1):390–9.

[65] Che L, et al. Optimal interconnection planning of community microgrids with renewable energy sources. IEEE Trans Smart Grid 2015;8(3):1054–63.

[66] Elsayed AT, Mohamed AA, Mohammed OA. DC microgrids and distribution systems: an overview. Electr Pow Syst Res 2015;119:407–17.

[67] S, Pfenninger, I, Staffell. *Renewables Ninja.* 2019; Available from: https://www.renewables.ninja/about.

[68] Pfenninger S, Staffell I. Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy 2016;114:1251–65.

[69] Staffell I, Pfenninger S. Using bias-corrected reanalysis to simulate current and future wind power output. Energy 2016;114:1224–39.

[70] Gelaro R, et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J Clim 2017;30(14):5419–54.

[71] S.J.W.C, Co., *SAA Approval 3 Core Cable 1.5MM 2.5MM 6MM PVC Insulated Twin & Earth TPS Flat Cable VAF Electrical Cord Wire.* 2021: Alibaba.

[72] NovaTerra, *NovaTerra Response Datasheet*, G.E.I.i.M. Study, Editor. 2020: Submitted to Imperial College London.

[73] Company CBEA. 2000W electric alibaba heaters for portable 220v electric heaters, electric mini fan heater, ceramic PTC electric heater heater. Alibaba; 2021.

[74] Co YJET. L., 300w 500W 800w 1000w 1200w 1500w pure sine wave off grid solar power inverter. Alibaba: Alibaba; 2021.

[75] Alibaba PP. Station 220wh Emergency Backup Lithium Battery 200w Ac Inverter Outlet Solar Generator - Buy Rechargable Batteries. Solar Power Bank: Camping Back Up Power Product on Alibaba.com; 2019.

[76] Ltd., R.S.E.C., *Rosen Monocrystalline 500W Price Mono 500Wp Solar Panel 500Watt.* 2021: Alibaba.

[77] Co ZZPST. L., *small size 50A 12V DC laboratory plating Rectifier /switching mode power supply*. Alibaba: Alibaba; 2021.

[78] Co., S.F.T.,. Solar Power Station 300W 500W 1000W 1500W 2000W Mobile Generator Power Bank Generator Outdoor Portable Power Station Generator. Alibaba: Alibaba; 2021.

[79] Equipment, Q.G.N.E., *Vertical wind generator 500w to 5kW ,vertical wind turbine for home use.* 2021: Alibaba.

[80] Coalition, C.a.C.A., *Financing Househeold Clean Heating Solutions.* 2018, UN Environment Finance Initiative.

[81] ESMAP, W.B., *Mini-Grids-for-Half-a-Billion-People-Market-Outlook-and-Handbook-for-Decision-Makers.* 2019.

[82] ESMAP, *Energy Sector Management Assistance Program (ESMAP) Annual Report* W. B. Group, Editor. 2021: Washington, D.C.

[83] IRENA. Renewable Power Generation Costs in, 2021. International Renewable Energy Agency: Abu Dhabi; 2020.

[84] A.M, Borchers, J.M, Duke, G.R, Parsons, *Does willingness to pay for green energy differ by source?*, in *Energy Policy.* 2007.

[85] Postali FA, Picchetti P. Geometric Brownian motion and structural breaks in oil prices: a quantitative analysis. Energy Econ 2006;28(4):506–22.

[86] Dufresne D. The integral of geometric Brownian motion. Adv Appl Probab 2001; 33(1):223–41.

[87] Marathe RR, Ryan SM. On the validity of the geometric Brownian motion assumption. Eng Econ 2005;50(2):159–92.

[88] Virtanen P, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17(3):261–72.

[89] M, Sakawa, *Genetic algorithms and fuzzy multiobjective optimization*. Vol. 14. 2002: Springer Science & Business Media.

[90] Sivanandam S, Deepa S. Genetic algorithms. In: Introduction to genetic algorithms. Springer; 2008. p. 15–37.

[91] Mohamed FA, Koivo HN. Online management genetic algorithms of microgrid for residential application. Energ Conver Manage 2012;64:562–8.

[92] Markov AA. The theory of algorithms. Trudy Matematicheskogo Instituta Imeni VA Steklova 1954;42:3–375.

[93] R.S, Sutton, A.G, Barto, *Reinforcement learning: An introduction*. 2018: MIT press.

[94] B, Baker, et al., *Designing neural network architectures using reinforcement learning*. arXiv preprint arXiv:1611.02167, 2016.

[95] C, Zhang, et al., *A study on overfitting in deep reinforcement learning*. arXiv preprint arXiv:1804.06893, 2018.

[96] Y, Li, *Deep reinforcement learning: An overview*. arXiv preprint arXiv:1701.07274, 2017.

[97] Arulkumaran K, et al. Deep reinforcement learning: A brief survey. IEEE Signal Process Mag 2017;34(6):26–38.

[98] M, Ghavamzadeh, et al., *Bayesian reinforcement learning: A survey*. arXiv preprint arXiv:1609.04436, 2016.

[99] Brockman G, et al. Openai gym 2016. arXiv preprint arXiv:1606.01540.

[100] F, Pardo, et al. Time limits in reinforcement learning. in *International Conference on Machine Learning*. 2018.

[101] Bhatnagar S, et al. Incremental natural actor-critic algorithms. Adv Neural Inf Proces Syst 2007;20:105–12.

[102] T, Schaul, et al., *Prioritized experience replay*. arXiv preprint arXiv:1511.05952, 2015.

[103] R, Munos, et al., *Safe and efficient off-policy reinforcement learning*. arXiv preprint arXiv:1606.02647, 2016.

[104] V, Mnih, et al. Asynchronous methods for deep reinforcement learning. in *International conference on machine learning*. 2016. PMLR.

[105] Schulman J, et al. Trust region policy optimization. International conference on machine learning. PMLR; 2015.

[106] Z, Wang, et al., *Sample efficient actor-critic with experience replay*. arXiv preprint arXiv:1611.01224, 2016.

[107] T, Akiba, et al. Optuna: A next-generation hyperparameter optimization framework. in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019.

[108] T, Kuremoto, et al., *Training deep neural networks with reinforcement learning for time series forecasting*, in *Time Series Analysis*. 2019, IntechOpen.

[109] N, Tomin, A, Zhukov, A, Domyshev, *Deep Reinforcement Learning for Energy Microgrids Management Considering Flexible Energy Sources*, in *EPJ Web of Conferences*. 2019, EDP Sciences. p. 01016.

[110] Bach F. Breaking the curse of dimensionality with convex neural networks. J Machine Learn Res 2017;18(1):629–81.

[111] J, Bhandari, D, Russo, *Global optimality guarantees for policy gradient methods*. arXiv preprint arXiv:1906.01786, 2019.

[112] DeVore R, Hanin B, Petrova G. Neural network approximation. Acta Numerica 2021;30:327–444.

[113] Bartlett PL, Montanari A, Rakhlin A. Deep learning: a statistical viewpoint. Acta Numerica 2021;30:87–201.

[114] P, Henderson, J Romoff, J, Pineau, *Where did my optimum go?: An empirical analysis of gradient descent optimization in policy gradient methods*. arXiv preprint arXiv:1810.02525, 2018.

[115] Liu Y, Gao Y, Yin W. An improved analysis of stochastic gradient descent with momentum. Adv Neural Inf Proces Syst 2020;33:18261–71.

[116] P, Petsagkourakis, et al., *Constrained reinforcement learning for dynamic optimization under uncertainty*. arXiv preprint arXiv:2006.02750, 2020.

[117] J, Achiam, et al. *Constrained policy optimization*. in *International conference on machine learning*. 2017. PMLR.

[118] Petsagkourakis P, et al. Reinforcement learning for batch bioprocess optimization. Comput Chem Eng 2020;133:106649.

[119] Kakinuma K, et al. Socio-ecological interactions in a changing climate: a review of the Mongolian pastoral system. Sustainability 2019;11(21):5883.

[120] Byambadorj T, Amati M, Ruming KJ. Twenty-first century nomadic city: Ger districts and barriers to the implementation of the Ulaanbaatar City Master Plan. Asia Pac Viewp 2011;52(2):165–77.

[121] Lessons from the Dzud: Baival; 2012.

[122] (IEA), I.E.A. *Mongolia Electricity Market Report*. 2021; Available from: https ://www.iea.org/countries/mongolia.

[123] Sovacool BK, D'Agostino AL, Bambawale MJ. Gers gone wired: lessons from the renewable energy and rural electricity access project (REAP) in Mongolia. Energy Sustain Dev 2011;15(1):32–40.

[124] Chimed G. Energy sector of mongolia. Country Report: Country Report. Energy Sector of Mongolia; 2019.

[125] Acosta MN, et al. Optimal under-frequency load shedding setting at Altai-Uliastai regional power system, Mongolia. Energies 2020;13(20):5390.

[126] Tamir K, Urmee T, Pryor T. Issues of small scale renewable energy systems installed in rural Soum centres in Mongolia. Energy Sustain Dev 2015;27:1–9.

[127] Y, Song, *Disclosable Version of the ISR - Ulaanbaatar Clean Air Project - P122320 - Sequence No : 19 (English)*. 2021, World Bank Group: Washington, D.C.

[128] Allen RW, et al. An assessment of air pollution and its attributable mortality in Ulaanbaatar, Mongolia. Air Qual Atmos Health 2013;6(1):137–50.

[129] Running I, et al. Clean Air with a Mongolian Ger. in *2020 IEEE Conference on Technologies for Sustainability (SusTech)*. IEEE; 2020.

[130] Purev U-U, Hagishima A. A field survey of traditional nomadic dwelling gers used as urban habitats in ulaanbaatar. *Mongolia* Evergreen 2020;7(2):155–71.

[131] Maddalena RL, et al. Quantifying space heating stove emissions related to different use patterns in mongolia. Energy and Environ Res 2014;4(3):147.

[132] S, Cowlin, et al., *Impact of improved stoves on indoor air quality in Ulaanbaatar, Mongolia*. 2005.

[133] Bayandelger B-E, Ueda Y, Adiyabat A. Experimental investigation and energy performance simulation of Mongolian Ger with ETS heater and solar PV in Ulaanbaatar city. Energies 2020;13(21):5840.

[134] GCRF, NovaTerra, and M.o. Energy. *Energy Infrastructure in Mongolia: Local, National and International perspective*. in *Energy for Development – low-carbon energy and industry for economic growth in Mongolia Workshop 1* 2021. Virtual: Global Challenges Research Fund.

[135] A.D, Bank, *Mongolia: Upscaling Renewable Energy Sector Project*. 2018, INTEGRATION environment & energy GmbH: Ulaanbaatar, Mongolia.

[136] W.B, Group, *Climate Change Knowledge Portal: Mongolia Historical Data*, W.B. Group, Editor. 2021.

[137] Yang X, et al. Effects of climate change on cultivation patterns and climate suitability of spring maize in Inner Mongolia. Sustainability 2021;13(14):8072.

[138] X, Dai, et al. *Climate change impact and adaptation in Inner Mongolia*. in *IOP Conference Series. Earth and Environmental Science*. 2009. IOP Publishing.

[139] Dashkhuu D, et al. Long-term trends in daily temperature extremes over Mongolia. Weather Clim Extremes 2015;8:26–33.

[140] World Bank Group E. Technical University of Denmark and Vortex, *Global Wind Atlas version* 2020;3.

[141] World Bank Group E. and SolarGIS Global Solar Atlas version 2020;2:2.

[142] Otsuki T. Costs and benefits of large-scale deployment of wind turbines and solar PV in Mongolia for international power exports. Renew Energy 2017;108:321–35.

[143] Blank J, Deb K. Pymoo: Multi-objective optimization in python. IEEE Access 2020;8:89497–509.

[144] M.G, Bellemare, W, Dabney, R, Munos. *A distributional perspective on reinforcement learning*. in *International Conference on Machine Learning*. 2017. PMLR.