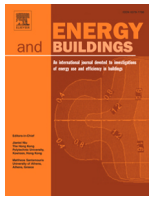
Energy & Buildings 346 (2025) 116144



Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild



Robust co-design framework for buildings operated by predictive control ⋆

Paola Falugi [a,*], Edward O'Dwyer [b], Marta Zagorowska [c], Eric Kerrigan [d,e], Yuanbo Nie [f], Goran Strbac [e], Nilay Shah [b]

[a] Engineering and Construction Department, University of East London, London, UK
[b] Department of Chemical Engineering, Imperial College London, London, UK
[c] DCSC, Delft University of Technology, Delft, the Netherlands
[d] Department of Aeronautics, Imperial College London, London, UK
[e] Department of Electrical and Electronic Engineering, Imperial College London, London, UK
[f] Department of Automatic Control and Systems Engineering, University of Sheffield, UK

ARTICLE INFO

Keywords:
Integrated optimisation of building sizing & operation
Energy efficiency & economic optimisation
Uncertain operating conditions
Closed-loop predictive control
Decomposition approach
Computational tractability

ABSTRACT

Cost-effective decarbonisation of the built environment is a stepping stone to achieving net-zero carbon emissions since buildings are globally responsible for more than a quarter of global energy-related CO_2 emissions. Improving energy utilisation and decreasing costs requires considering multiple domain-specific performance criteria. The resulting problem is often computationally infeasible.

The paper proposes an approach based on decomposition and selection of significant operating conditions to achieve a formulation with reduced computational complexity. We present a robust framework to optimise the physical design, the controller, and the operation of residential buildings in an integrated fashion, considering external weather conditions and time-varying electricity prices. The framework explicitly includes operational constraints and increases the utilisation of the energy generated by intermittent resources.

A case study illustrates the potential of co-design in enhancing the reliability, flexibility and self-sufficiency of a system operating under different conditions. Specifically, numerical results demonstrate reductions in costs up to 30% compared to a deterministic formulation. Furthermore, the proposed approach achieves a computational time reduction of at least 10 times lower compared to the original problem with a deterioration in the performance of only 0.6 %.

1. Introduction

1.1. Motivation

Achieving net-zero carbon emissions necessitates the cost-effective design and operation of energy-efficient buildings capable of integrating with complex energy systems characterized by renewable sources, storage technologies, and dynamic market conditions. The most critical difficulties in addressing the decarbonization challenge stem from the exponential growth of energy demand, the complexity of the involved systems, and multiple uncertainties affecting energy production and usage. Active participation of buildings in the whole energy system to achieve emission targets is extremely valuable since commercial and residential buildings are responsible for about 38 % of global CO_2 emissions [1]. Active support of buildings for the whole energy system necessitates adopting renewable resources, various storage technologies, and smart devices. It also requires consumers' participation in dynamic electricity markets characterised by time-varying prices. The design of flexible and cost-efficient buildings requires optimising the technologies' size considering the dweller needs and how the system operates under different plausible conditions. Frequently, the effect of how the system is operated is neglected in the design process, usually because including the optimal closed-loop operation of the system in the sizing problem and considering the uncertainty in the operating conditions leads to a computationally challenging problem. Nevertheless, neglecting operational uncertainty and the behaviour of closed-loop systems may lead to a suboptimal design.

Motivated by the requirement for buildings to increase their energy efficiency at affordable costs, we propose a framework for optimally de-

⋆ This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) [grant EP/V012053/1].
* Corresponding author.
E-mail address: p.falugi@uel.ac.uk (P. Falugi).

https://doi.org/10.1016/j.enbuild.2025.116144
Received 12 December 2024; Received in revised form 29 May 2025; Accepted 12 July 2025
Available online 14 July 2025
0378-7788/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

signing the system and the controller parameters considering uncertain operating conditions.

1.2. Existing works

Addressing the decarbonization challenge requires buildings to operate near their performance limits and explicitly account for constraints. Model Predictive Control (MPC) is recognized as a natural choice for operating buildings [2]. Furthermore, the availability of near-term weather forecasts and the prospective participation of consumers in dynamic electricity markets characterised by time-varying prices makes Economic Model Predictive Control (EMPC) [3,4] the natural choice for a dweller to improve performance while optimising the electricity cost. However, the performance of predictive controllers depends on tuning their parameters, such as prediction horizon, sampling time and discretisation step. Traditionally, the success of the tuning process relied on experience. In Khusainov et al. [5], an automatic tuning of the MPC is proposed in a co-design framework to achieve an optimal trade-off between performance and computational resources in nominal conditions. Their approach automates the MPC and hardware co-design using the Bi-objective Mesh Adaptive Direct Search algorithm (BiMADS) to handle discrete variables.

However, optimising the building design in isolation from its operation can be suboptimal. Such a limitation motivates the control co-design concept, which considers the simultaneous design of the systems and the controller to push a system's performance to its achievable limits. Recently, Garcia-Sanz [6] recognized the importance of adopting a control co-design concept to push the system's performance to its achievable limits. The control co-design framework considers multidisciplinary subsystem interactions in a unified manner, enabling the opportunity to improve performance substantially. In the 1980s, the co-design idea was integrated with optimisation schemes to identify control and parameters [7]. Frameworks co-optimising closed-loop software implementation and hardware performance appear in Suardi et al. [8], Kircher and Zhang [9] where the co-design consists of a multi-objective formulation since the decision variables span different time scales and belong to different areas. A survey on control co-design applications is provided in Diangelakis et al. [10]. The optimal equipment selection considering the system operation is critical to increasing the energy efficiency of large heating systems, as pointed out in Henze et al. [11], Powell et al. [12].

The co-design problem requires identifying several significant operating conditions to achieve a reliable system design. The description of such operating conditions necessitates several years of exogenous data to obtain a set of annual scenarios and a useful uncertainty model of the time-varying exogenous variables. The amount of data often leads to challenging numerical problems, so a common approach is considering a limited data set. However, a naive choice of subsets of data in most cases results in a solution far from optimal [13,14]. Many selection methods and aggregation techniques select and aggregate typical subsamples without considering what might be relevant to the problem under consideration, and they easily neglect extreme events. The importance subsampling is an approach that extracts fewer observations (subsamples) from long time series through systematic identification of timesteps carrying essential information for the problem under consideration by assigning to each subsample a measure of its importance in realising the problem output. Recent importance subsample techniques [14,15] identify extreme events that are significant for the problem of interest, looking at their effect on the problem output, achieving a design choice capable of good performance on a more extensive set of operating conditions at a reduced computational burden.

1.3. Novelty

Existing co-design approaches often suffer from computational intractability, particularly when accounting for uncertainty and controller tuning. As a result, their practical application in building design has remained limited. We propose a computationally tractable co-design framework that jointly optimises technology sizing and EMPC controller parameters, while ensuring robustness to uncertain operating conditions through scenario-based evaluation. The work builds on the approaches [5,14,16], considering performance, uncertainty and control tuning at the same time. The proposed framework explicitly addresses discontinuities arising from the discrete nature of sizing parameters and accommodates time-varying, piecewise-constant electricity prices within the EMPC cost function, leveraging techniques from Khusainov et al. [5]. Importance sub-sampling techniques are used to decompose the problem to ensure computational tractability without significant performance impact. Furthermore, this work extends the deterministic formulation from Falugi et al. [16] and proposes a multi-objective co-design framework that automates technology selection while incorporating uncertainty and control parameter tuning. In contrast to previous works, the present formulation introduces robustness to uncertainty and automated control tuning as integral components of the co-design framework. Incorporating uncertainty necessitates alternative problem formulations and solution strategies compared to deterministic approaches. To address such challenges, the framework adopts a scenario-based optimisation strategy inspired by the subsample selection methodology introduced in Hilbers et al. [14]. This concept is further extended here to consider the impact of highly uncertain, time-varying electricity prices and dynamic correlation of variables at different time instants through the predictive feedback controller capability.

The rest of the paper is structured as follows. Section 2 introduces the problem formulation and the numerical challenges due to uncertainty and possible discontinuities. Section 3 presents the decomposition techniques adopted in this paper to improve the numerical properties of the algorithm and describes the proposed co-design framework. Section 4 presents numerical results of the proposed framework applied to a residential building case study. The paper ends with conclusions in Section 5.

2. Problem formulation

The co-design framework must use building models to consider the building operation managed by the MPC. To obtain such models, we can follow a systematic workflow starting from a detailed physics-based model followed by low-order control-oriented modelling and heating system component modelling. The model complexity is reduced using model order reduction methods [17]. The models used in the design phase have structures analogous to electrical circuits composed of resistors and capacitors [18], with heat flows and temperature differences represented as currents and voltages. Such structures are commonly used to represent the thermal behaviour of a building in both simulation [19] and control [20] applications. Reviews of such methods can be found in Harish and Kumar [21]. Where data are available, parameterisation can be achieved through system identification methods. At the design stage, or more generally, in the absence of data, model parameters can be chosen based on the building fabric dimensions and materials. For our design-oriented framework, the latter case is most relevant. The last section illustrates the co-design framework on a three-bedroom dwelling with a high insulation level, adopting a single-zone lumped-capacitance method with the model described as ordinary differential equations [22]. Future work could include calibration against monitoring data when transitioning from the design to the implementation phase.

2.1. System dynamics

The system is described by

$$\dot{x}(t) = f(x(t), u(t), w(t), p, t) \tag{1}$$

where $f(\cdot)$ is continuous, $x(t) \in \mathbb{R}^{n_x}$ is the system state, $u(t) \in \mathbb{R}^{n_u}$ the control input, $p \in \mathbb{R}^{n_p}$ a vector of constant design parameters, and $w(t) \in$

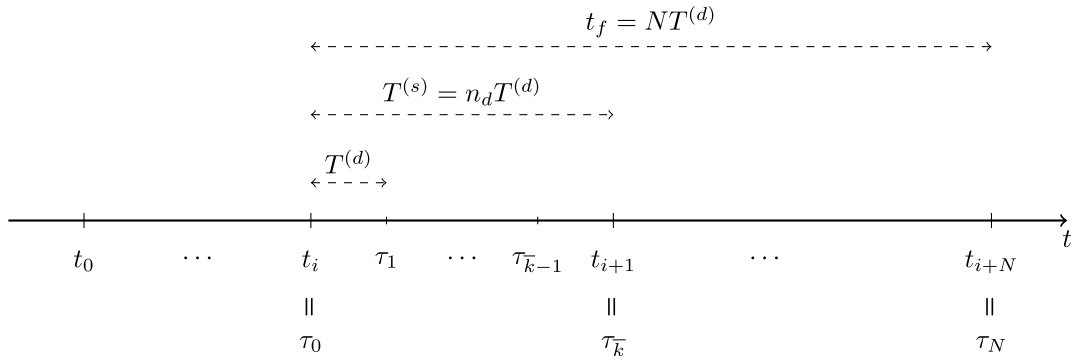


Fig. 1. Notation of time variables.

$\mathbb{R}^{n_w}$ a vector of uncertain time-varying exogenous inputs belonging to a time-varying uncertainty set $\mathbb{W}(t)$. The exogenous inputs $w(t)$ define the system's operating conditions. The aim is to design a system and a controller to achieve high performance under plausible operating conditions described by $w(\cdot) \in \mathcal{W} := \{\text{measurable bounded functions } w : [0,\infty) \to \mathbb{R}^{n_w} \ : \ w(t) \in \mathbb{W}(t)\}$. The uncertainty on the short time scale in real-time forecasting is assumed to be implicitly counteracted by the MPC controller operated in closed-loop. We are primarily interested in large uncertainties captured by scenarios characterising significant operational differences.

Given the sizing parameter p and the prediction horizon t_f, the closed-loop controller at time t minimizes the performance cost with Eq. (2):

$$\int_t^{t+t_f} \ell(x(\tau), u(\tau), w(\tau), p, \tau)\mathrm{d}\tau \tag{2}$$

while satisfying the constraints Eqs. (3)–(6):

$$\dot{x}(\tau) = f(x(\tau), u(\tau), w(\tau), p, \tau), \qquad \forall \tau \in \mathcal{T} \text{ a.e.}, \tag{3}$$

$$g(\dot{x}(\tau), x(\tau), \dot{u}(\tau), u(\tau), w(\tau), p, \tau) \le 0, \qquad \forall \tau \in \mathcal{T} \text{ a.e.}, \tag{4}$$

$$c(\dot{x}(\tau), x(\tau), \dot{u}(\tau), u(\tau), p, \tau) = 0, \qquad \forall \tau \in \mathbb{T}, \tag{5}$$

$$\psi(x(t), x(t+t_f), p, t, t_f) \le 0, \tag{6}$$

where 'a.e.' stands for 'almost everywhere' in the Lebesgue sense on the interval $\mathcal{T} := [t, t+t_f] \subset \mathbb{R}$, where t and t_f denote, respectively, the initial time and the prediction horizon. The set $\mathbb{T} \subset \mathcal{T}$ is a finite subset of $\mathcal{T}$. The inequality constraints Eq. (4), where $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_p} \times \mathbb{R} \to \mathbb{R}^{n_g}$ and n_g indicate the number of constraints, describe general constraints on the variables. General boundary equality and inequality constraints can be imposed through Eq. (6), where $\psi : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^{n_I}$ and n_I is the number of terminal constraints.

It is assumed that the states and inputs lie in compact sets $\mathbb{X}$ and $\mathbb{U}$, respectively, which are included in Eq. (4). The optimal closed-loop policy at time t is obtained by minimising Eq. (2) subject to the constraints Eqs. (3)–(6). Such a closed-loop policy, and consequently its performance, is determined by the system design parameters p, the realisation of the uncertainty $w(\cdot)$ in the interval $[t,\ t+t_f]$ and the implemented algorithm computing the optimal controller, which necessitates the selection of several parameters such as prediction horizon, discretisation step, and sampling time. The systematic choice of these parameters is the objective of the algorithm proposed in this work.

2.2. Risk measures

The system is required to provide good performance under a large variety of operating conditions that are unknown at the design stage. The cost Eq. (31) and in consequence Eq. (2) changes for different realisations of the uncertainty, and as a result, a measure of risk providing a surrogate for the overall cost is adopted [23]. In particular, the measure of risk assigning a single value to an uncertain variable Z is a functional $\mathcal{R} : Z \to (-\infty, \infty)$, which is required to be a coherent measure of risk to be a good risk quantifier [23]. Examples of coherent risk measures are the expectation $E[Z]$ and the worst-case realisation $\max(Z)$. The choice of risk measures in this work will be discussed in Section 3.

2.3. Numerical solution

An algorithmic implementation of the optimisation problem Eqs. (2)–(6), given the information on the system state and its environment, computes a control law at time instants $t = t_i$ for $i = 0, 1, \dots$, and imposes the conditions Eqs. (3)–(6) on a discrete set $\mathcal{T}^d := \{\tau_0, \tau_1, \dots, \tau_N\}$ of time instances satisfying $\tau_0 = t_i < \tau_1 < \dots < \tau_N = t_i + t_f$ (Fig. 1). Denote $T^{(s)} := t_{i+1} - t_i$ as the sampling time for all i and $T_k^{(d)} := \tau_{k+1} - \tau_k$ the discretization steps where $\tau_k \in \mathcal{T}^d$, $k = 0, 1, \dots, N$. For the sake of simplicity, we assume that there exists a positive integer n_d such that $T^{(s)} = n_d T_0^{(d)}$.

The parameter $T^{(s)}$ describes how often new measurements are retrieved, and an optimisation problem is solved to compute the control law. The finite parametrisation of the problem involves a finite parametrisation of all the trajectories that we highlight, adding the tilde symbol ˜ to their definition. The control algorithm, given the state x_i at time t_i and the sequence $\tilde{\mathbf{w}}_{[\tau_0,\tau_N]} = (\tilde{w}(\tau_0), \tilde{w}(\tau_1), \dots, \tilde{w}(\tau_N))$ over the horizon N, minimizes a discretized formulation of the problem Eqs. (2)–(6) with a cost

$$J^{(D)}(\tilde{\mathbf{u}}, \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, t, t_f, p, x_i) = \sum_{k=0}^{N} \ell^D(\tilde{x}(\tau_k), \tilde{u}(\tau_k), \tilde{w}(\tau_k), p, t) \tag{7}$$

that depends on the discretisation method [24]. We denote as

$$\begin{aligned}\tilde{\mathbf{u}}^*(x_i, t_i; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, p, p_c) = \ & (\tilde{u}^*(\tau_0; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, x_i, p, p_c), \\ & \tilde{u}^*(\tau_1; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, x_i, p, p_c), \dots, \tilde{u}^*(\tau_N; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, x_i, p, p_c))\end{aligned} \tag{8}$$

and

$$\begin{aligned}\tilde{\mathbf{x}}^*(x_i, t_i; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, p, p_c) = \ & (\tilde{x}^*(\tau_0; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, x_i, p, p_c), \\ & \tilde{x}^*(\tau_1; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, x_i, p, p_c), \dots, \tilde{x}^*(\tau_N; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, x_i, p, p_c))\end{aligned} \tag{9}$$

the optimal input and state sequences, respectively, returned by the control algorithm. The optimal sequences Eqs. (8) and (9) depend on the vector p_c of the controller parameters of interest, such as the prediction horizon, the sampling time, and the discretisation step.

The control law $\kappa_N(t, x_i, t_i; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, p, p_c)$, applied to the system is determined by the first $\bar{k}$ elements of the sequence $\tilde{\mathbf{u}}^*(x_i, t_i; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, p, p_c)$ in the time interval $t \in [t_i, t_i + T^{(s)}]$ as follows

$$\begin{aligned}&\kappa_N(t, x_i, t_i; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, p, p_c) := \\ &\quad \phi(t, [\tilde{u}^*(\tau_0; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, x_i, p, p_c), \dots, \tilde{u}^*(\tau_{\bar{k}}; \tilde{\mathbf{w}}_{[\tau_0,\tau_N]}, x_i, p, p_c)])\end{aligned} \tag{10}$$

where $\bar{k}$ satisfies $\tau_{\bar{k}} = t_i + T^{(s)}$ and the interpolation function $\phi(\cdot)$ is determined by the discretization method that was used. The achieved

closed-loop cost $V_{cl}(\cdot)$ in the interval $[t_i, t_i + T^{(s)}]$ is

$$V_{cl}(x_i, t_i, \mathbf{w}_{[t_i,t_{i+N}]}, p, p_c) = \int_{t_i}^{t_i+T^{(s)}} \ell(x(\tau), \kappa_N(\tau, x_i, t_i; \tilde{\mathbf{w}}_{[t_i,t_{i+N}]}, p, p_c), w(\tau), p)\mathrm{d}\tau \tag{11}$$

where $x(t)$ refers to the closed-loop trajectory obtained by applying the control law Eq. (10) and $\mathbf{w}_{[a,b]}$ is the vector of exogenous signals in the interval $[a, b]$.

2.4. Multi-objective co-design problem

The optimisation of a building's performance requires considering multiple objectives.

2.4.1. Multiple objectives

An essential performance criterion in designing and operating a building is its economic cost, which consists of investment and operating costs given by time-varying electricity prices. Thus, the first objective is the minimisation of the system present value cost [25, Appendix A] given by

$$J^{(1)}(p, p_c) := \mathcal{R}\left(\sum_{i=0}^{N_y} V_{cl}(x_i, t_i, \mathbf{w}_{[t_i,t_{i+N}]}, p, p_c) + V_I(p)\right) \tag{12}$$

with a chosen risk measure $\mathcal{R}$, where $V_I(p)$ considers investment costs and N_y is the number of samples required to cover a whole year. The initial condition x_i can also be considered as an uncertain parameter. Since the performance accounts for a time-varying economic criterion, the choice of the controller parameters requires the inclusion of additional objectives considering the performance of the closed-loop system with respect to a reference performance [4].

Even though extensive research has been performed on EMPC, little has been said about closed-loop performance for the time-varying case [3,26]. In general, it is not straightforward to identify feasible trajectories and, consequently, a reference performance $V_{cl}(x_j^r, t_j, \mathbf{w}_{[t_j,t_{j+N}]}, p, p_c^r)$ for $j = 0, \dots, N_y^r$ for time-varying formulations. The superscript r denotes a reference term. For this reason, the reference performance is determined by physically meaningful controller parameters from which we expect to achieve the best closed-loop performance defined by Eq. (2). The achieved asymptotic average cost is no worse than the one provided by the reference trajectory. The sensitivity analysis of the optimal cost with respect to these parameters can also indicate whether their choice is adequate. Note that a reference is deemed suitable if the associated performance is not too sensitive to small parameter changes, especially including, as an additional design parameter, a degree of tightening in the terminal constraints to enforce a robustness margin at the operational level. Let

$$\mathbf{V}_{\mathbf{M}}(x_0, t_0, \mathbf{w}_{[t_0,t_{M+N-1}]}, p, p_c) := [V_{cl}(x_0, t_0, \mathbf{w}_{[t_0,t_N]}, p, p_c), \dots, V_{cl}(x_{M-1}, t_{M-1}, \mathbf{w}_{[t_{M-1},t_{M+N-1}]}, p, p_c)]' \tag{13}$$

where $'$ indicates the transpose of a vector. The cost accounting for the controller closed-loop performance over a year is defined as

$$J^{(2)}(p, p_c) := \mathcal{R}(D(\mathbf{V}_{\mathbf{N_y}}(x_0, t_0, \mathbf{w}_{[t_0,t_{N_y+N-1}]}, p, p_c), \mathbf{V}_{\mathbf{N_y}}(x_0, t_0, \mathbf{w}_{[t_0,t_{N_y+N-1}]}, p, p_c^r))) \tag{14}$$

where $D(\cdot)$ denotes the norm of choice. Other choices could be relevant to perform the auto-tuning of the controller [27], but for simplicity, we consider that the discretisation time can only assume values for which the approximating error of the system dynamics is acceptable.

Another important tuning criterion is the computational resources required by the controller. The MPC tuning requires a trade-off between closed-loop performance improvements and the additional computational effort needed to provide such performance improvements. Long prediction horizons and small sampling and discretisation steps might be unnecessary to achieve good performance while still determining the complexity of the problem to be solved. Consequently, the third cost accounts for the preference for short horizons and large sampling and discretisation steps

$$J^{(3)}(p, p_c) := \mathcal{R}(Q(p_c; \mathbf{w}_{[t_0,t_{N_y}]})) \tag{15}$$

where $Q(\cdot)$ indicates a cost of choice depending on the controller parameters. For example, a possible choice consists of the computational time required to compute the objective $\mathbf{V}_{\mathbf{N_y}}(x_0, t_0, \mathbf{w}_{[t_0,t_{N_y+N-1}]}, p, p_c)$ that highlights the potential dependence of the cost $Q(\cdot)$ on the specific realization of $\mathbf{w}_{[t_0,t_{N_y}]}$.

2.4.2. Co-design problem

The co-design problem is formulated as the following multi-objective problem

$$\min_{p,p_c}(J^{(1)}(p, p_c), J^{(2)}(p, p_c), J^{(3)}(p, p_c))$$

subject to

$$p \in \mathcal{P}, p_c \in \mathcal{P}_c \tag{16}$$

where $\mathcal{P}$ and $\mathcal{P}_c$ describe the feasible spaces for the sizing and controller parameters, respectively. Note that the costs $J^{(1)}(p, p_c)$, $J^{(2)}(p, p_c)$ and $J^{(3)}(p, p_c)$ depend on the closed-loop system performance. The multi-objective problem Eq. (16) with conflicting objectives does not have a single solution that simultaneously optimises each objective, but a set of possible optimal solutions known as the Pareto frontier [5,28]. A solution is *Pareto optimal* if the improvement in one objective's value degrades some other objective values. All Pareto optimal solutions are equally good, and a single choice relies on preference or additional criteria. In this work, preferences and the proposed approach determine the final optimal solution.

Problem Eq. (16) is computationally complex since it requires numerous function evaluations consisting of time-consuming simulations to evaluate the closed-loop behaviour under different operating conditions. Problem Eq. (16) is a bi-level optimisation task, where the objective functions are evaluated via black-box simulations, hence gradient information isunavailable. Their values are determined by performing black-box simulations [29]. The derivative information on the objective functions is not available since some of the design parameters included in p and p_c can only assume a finite number of values, and the time-varying electricity prices are piece-wise constant. Moreover, the chosen risk measure can also induce discontinuities in the objective function, for example, if the chosen $\mathcal{R}(\cdot)$ is the max operator.

3. Problem decomposition and importance subsamples

To mitigate the computational burden of solving problem Eq. (16), we propose a decomposition-based approach that enables efficient identification of near-optimal design parameters through control tuning and scenario subsampling. In the present framework, we extend the idea of *importance subsampling*, proposed in Hilbers et al. [14], to problems presenting dynamics correlating variables at different time instants. Since time-varying electricity prices induce a large variability in the costs and the optimal design choices, we determine the importance of a subsample based on the associated optimal cost and the design parameter p.

3.1. MPC tuning with subsample selection

Even if the multi-objective problem Eq. (16) operates on reduced data sets, such a problem is still computationally demanding and requires trade-off choices. Therefore, we propose decomposing the problem Eq. (16) into two problems. In particular, we note that the constraints $\mathcal{P}_c$ and $\mathcal{P}$ on the controller and sizing parameters, respectively, are independent. The proposed co-design framework aims to mitigate the computational complexity by decoupling multiple objectives and solving multiple optimisation problems of reduced size.

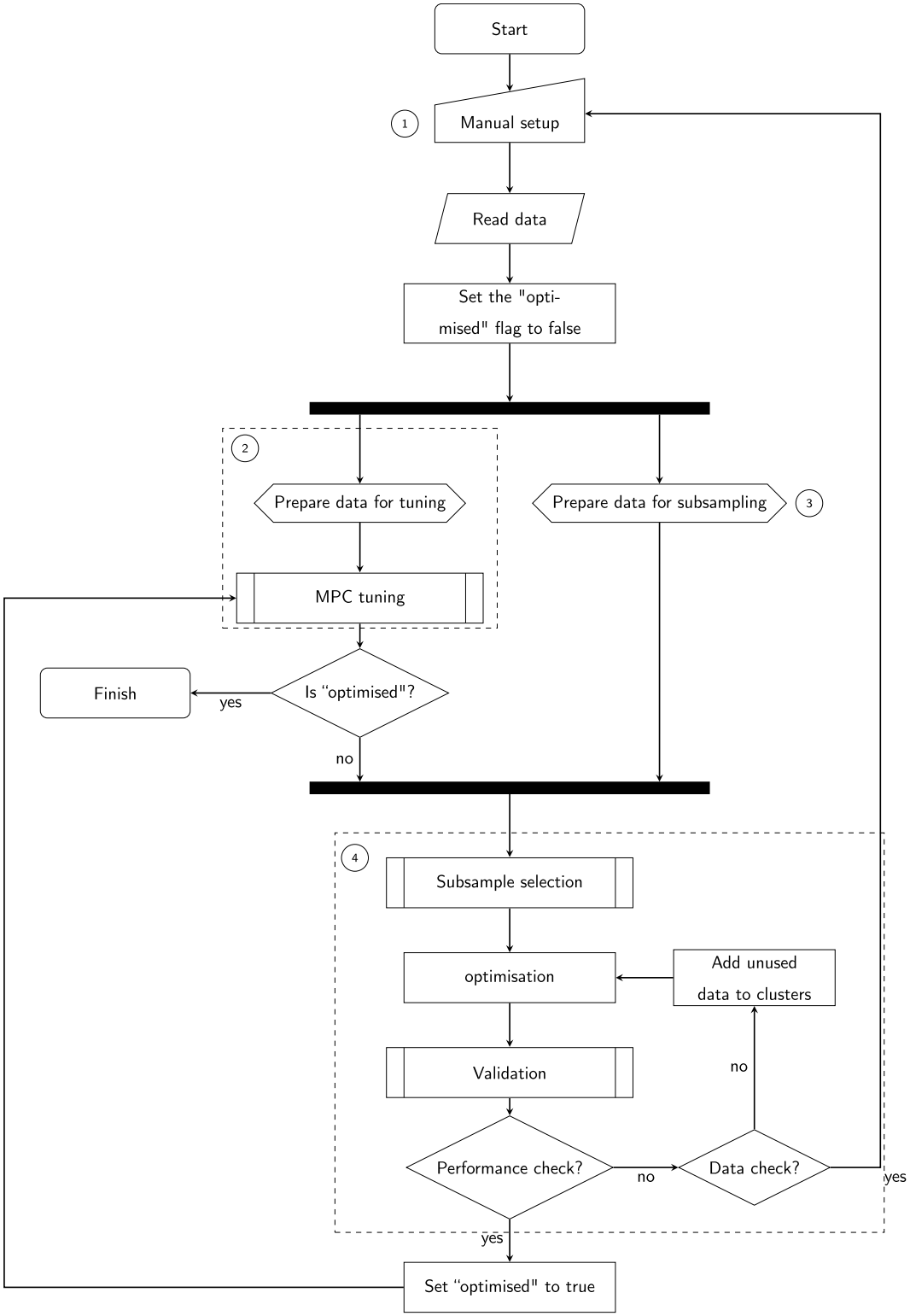


Fig. 2. Iterative co-design flowchart.

In particular, the design parameter p^* is a solution to the following problem $\mathbb{P}^{CD}$, which uses the information obtained by running MPC tuning p_c^* and subsample selection $\mathcal{H}_c$:

$$\min_{p\in\mathcal{P}} \mathcal{R}\left(\sum_{i\in\mathcal{H}_c} \nu_i \mathbf{1}'_{\mathbf{N_i}} \mathbf{V}_{\mathbf{N_i}}(\hat{x}_i, t^{(i)}, \mathbf{S^{(i)}}, p, p_c^*)\right) + V_I(p) \tag{17}$$

The problem $\mathbb{P}^{CD}$ provides a design parameter p^* that has a degree of robustness that is the result of a compromise between computational complexity and modelling accuracy.

The flowchart in Fig. 2 shows the full co-design procedure with four main elements:

① Manual setup that includes:
- Choosing the risk measure $\mathcal{R}$;

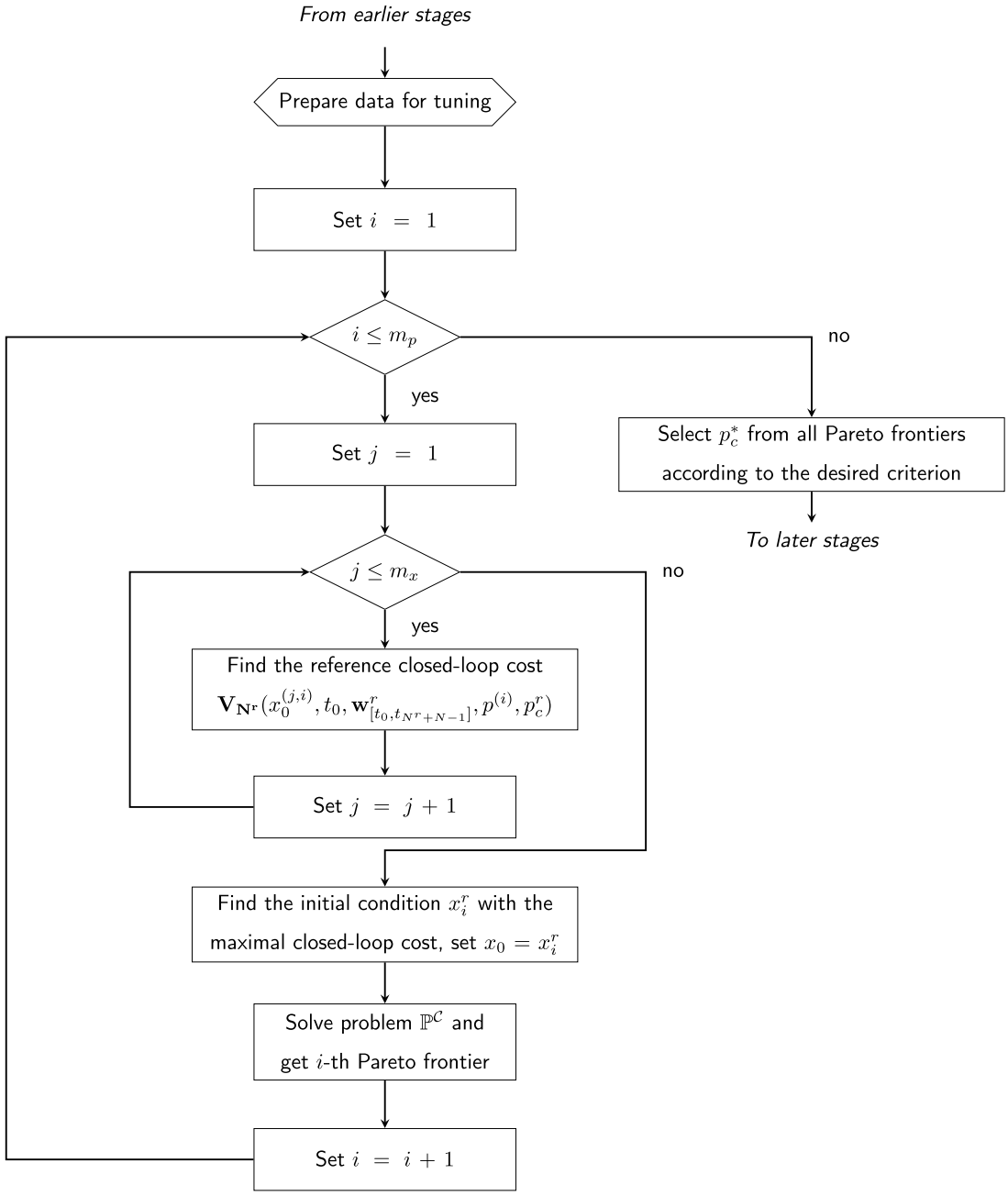


Fig. 3. MPC tuning.

- Choosing the dataset;
- Choosing the preferred criterion to find a trade-off between computational complexity of the controller and its performance [30];
- Choosing clustering method;
- Choosing the maximal number of clusters $k_{\max}$;
- Setting up the numerical methods from Section 2.3.

② Preparation of data for control and MPC tuning. Data preparation for controller tuning includes:

- Choosing a persistently exciting dataset $\mathbf{w}^r_{[t_0,t_{N^r+N-1}]}$;
- Choosing controller reference parameter p_c^r;
- Picking m_p values $p^{(i)} \in \mathcal{P}$, $i = 1, \dots, m_p$;
- Picking m_x initial states $x_0^{(j,i)} \in \mathcal{X}_{p^{(i)}}$, $j = 1, \dots, m_x$, $i = 1, \dots, m_p$.

The details of MPC tuning are in Section 3.2 and the corresponding flowchart is in Fig. 3.

③ Preparation of data for subsampling that includes:

- Choosing the initial subsamples $\mathbf{S^h}$, $h = 1, \dots, m$;
- Setting m initial states $\hat{x}_h$.

④ Subsample selection, optimisation, and validation that are presented in Section 3.3 and shown in the flowchart in Fig. 4. The controller p_c^* obtained from MPC tuning ② and the data preparation for subsampling ③ are used to find the optimal solution, if the selected performance threshold $\mathcal{J}$ is reached (arrow *Performance checks complete*), or to require manual adjustments of the selected criteria, if all data have been used and the performance threshold has not been reached (arrow *Data checks complete*).

3.2. MPC tuning

The automatic tuning of the MPC controller is given by a trade-off between the objectives $J^{(2)}(p, p_c)$, accounting for the closed loop

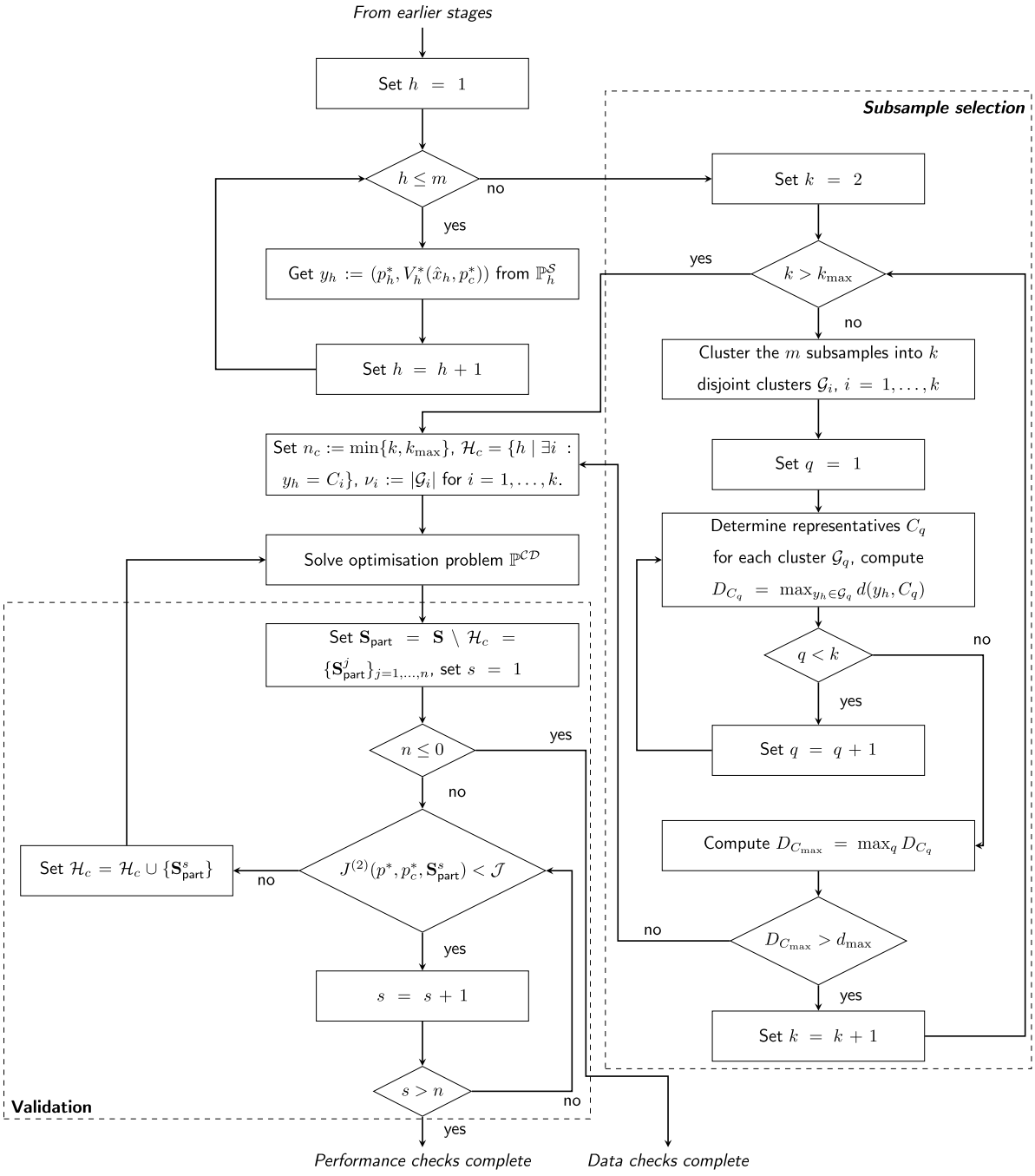


Fig. 4. Subsample selection, optimisation, and validation.

performance, and $J^{(3)}(p, p_c)$, considering the controller computational complexity, where p is a-priori unknown and consequently considered an uncertain parameter.

The tuning problem $\mathbb{P}^{\mathcal{C}}$ is defined as

$$\min_{p_c} \mathcal{R}_p((J^{(2)}(p, p_c), J^{(3)}(p, p_c))$$

subject to

$$p_c \in \mathcal{P}_c,\ p \in \mathcal{P}, \tag{18}$$

where we have introduced the risk measure $\mathcal{R}_p$ with respect to p since the optimal controller parameters need to be a good choice for all $p \in \mathcal{P}$. In particular, in problem Eq. (18), we use as a risk measure the max operator to guarantee the performances on the Pareto frontier for all possible system configurations described by the compact set $\mathcal{P}$. The optimal p_c, denoted as p_c^*, is a preferred choice determined by a compromise between closed-loop performance and computational complexity. The formulation of the problem $\mathbb{P}^{\mathcal{C}}$ can be further simplified since the evaluation of the cost function $J^{(2)}(p, p_c)$ can be reliably performed on persistently exciting [31] training data sets $\mathbf{w}^r_{[t_0, t_{N^r+N-1}]}$ of limited length. The persistently exciting requirement on $\mathbf{w}^r_{[t_0, t_{N^r+N-1}]}$ consists of asking for a sufficiently rich data set that guarantees the system controlled by the MPC visits all the operating conditions of interest. The training data sets $\mathbf{w}^r_{[t_0, t_{N^r+N-1}]}$ are created from the full time-series, making sure of selecting the relevant features of the signals. Their length must be at least twice as long as the reference MPC prediction horizon since performance is evaluated on the closed-loop system across the length of the MPC prediction horizon. Persistently exciting conditions can be validated on the computed reference trajectory. Note that tuning on short

datasets might depend on the system's initial state. For this reason, the proposed simplified algorithm considers the effect of the initial state.

Let $\mathcal{X}_p \subseteq \mathbb{X}$ be the state constraint set depending on the value of the design parameter p. Algorithm 1 details the MPC tuning procedure for a training set. The procedure can be repeated on additional training sets if the system behaviour is significantly variable. The final parameter choice verifies that the desired performances are guaranteed in all the investigated operating conditions.

Note that when the states are related to the stored energy in the system, the number m_x of needed initial conditions is often small. Indeed, the most demanding operating conditions for energy storage often sit on the extremes given by minimal or maximal low energy levels, which require higher control effort to balance internal energy requirements. In addition, the optimal p_c^* also gives information on the length of the importance subsample, which cannot be shorter than twice the length of the MPC prediction horizon.

3.3. Importance subsample

Classic data reduction approaches use individual years or cluster data into representative days and lead to significant errors in estimates of optimal system design due to data omissions affecting the output of the problem since they neglect how the problem depends on the data [15]. Conversely, the importance subsampling approach selects and groups subsamples according to their effect on the problem output, as discussed in Hilbers et al. [15] for models without interdependence between the sampled data.

3.3.1. Subsample definition

In the present contribution, the importance of a subsample is evaluated by optimising the system investment and operation cost computed, considering the closed-loop operation across a short subsample of weather and electricity prices. Note that the subsample length, as pointed out in the previous subsection, needs to be longer than twice the MPC prediction horizon to consider the correlation between time instants through the MPC prediction horizon on the closed-loop performance. The operation cost of the closed-loop system across the short subsample is weighted by R_h according to the subsample length to estimate the annual operational cost, assuming that the considered operating conditions repeat across the year.

Let $\mathcal{S} := \{\mathbf{S^{(1)}}, \mathbf{S^{(2)}}, \dots \mathbf{S^{(m)}}\}$ be a collection of m subsamples of $\mathbf{w}_{[t_0,T]}$ where T accounts for the full length of the time-series. Each $\mathbf{S^{(h)}}$ has $N_h + N - 1$ samples where N_h is the simulation length. The problem, denoted as $\mathbb{P}_h^S$, evaluating the importance of the subsample $\mathbf{S^{(h)}}$ is defined as

$$V_h^*(\hat{x}_h, p_c^*) = \min_p R_h \mathbf{1}'_{\mathbf{N_h}} \mathbf{V}_{\mathbf{N_h}}(\hat{x}_h, t_0^{(h)}, \mathbf{S^{(h)}}, p, p_c^*) + V_I(p)$$
$$\text{subject to } p \in \mathcal{P}, \tag{19}$$

where $\mathbf{1}_{\mathbf{N_h}}$ is a vector of ones of length N_h, $t_0^{(h)}$ is the initial time of $\mathbf{S^{(h)}}$ and $\hat{x}_h$ is the assigned initial state. Solving Eq. (19) corresponds to finding the parameters p that give the smallest closed-loop cost for a given controller described by p_c^* and for the initial condition $\hat{x}_h$. The choice of $\hat{x}_h$ should represent a condition providing a degree of robustness according to a risk measure. A natural simplifying choice for energy-related problems is considering the initial state returning the $\mathbb{P}_h^S$ with a higher cost.

If the specific problem instance is solvable for a prediction horizon covering the whole subsample, a meaningful choice is to consider the initial state a decision variable and impose an equality constraint with the state at the end of the subsample. The selection of subsamples requires the optimal solution of $\mathbb{P}_h^S$ for $h = 1, \dots, m$ and the definition of criteria to determine their importance. The criterion is a problem-dependent choice, and the most common choices consider the optimal cost $V_h^*(\hat{x}_h, p_c^*)$ as discussed in Hilbers et al. [14,15]. However, the volatility of electricity prices can render optimal cost alone an insufficient measure of subsample importance. Therefore, we propose choosing the importance subsample using the optimal cost and the design parameters p. This is because, independently of the achievable prices, they represent the design choice of interest for the problem under consideration.

In Algorithm 4, $|\mathcal{G}_i|$ denotes the cardinality of the set $\mathcal{G}_i$ and the function $d(\cdot)$ computes the distance between points. Various distance measures and methods appear in the literature to evaluate the quality of a cluster [32,33]. A commonly used distance is the within-cluster sum of squares, which estimates cluster tightness and accounts for the variance of the data, but different distance measure choices are possible.

The choice of the number of clusters depends on more than just the desired clustering resolution, but is a separate task from clustering. The shape of the clusters and the scale of the distribution of data points are important factors, as the case study shows in Section 4.

The most popular partitioning algorithms are known as k-means and k-medoids, and they minimise the sum over each cluster of the squared distance between the selected cluster centre and candidate points of the cluster. The k-medoid algorithm has the advantage that the centre is an element of the cluster, and it is more robust to outliers and noise [34], and was chosen in this work.

4. Case study: residential building co-design

The proposed framework from Section 3 is now applied to a residential building.

4.1. Residential building operating with time-varying energy prices

4.1.1. Building dynamics

The co-design framework is illustrated on a residential building contributing to a grid with time-varying energy prices. The thermal dynamics of a three-bedroom dwelling with a high insulation level are modelled by adopting a single-zone lumped-capacitance method [22]. The building is equipped with electrically driven heat pumps (HP), providing temperature regulation. We consider the option of installing photovoltaic panels (PV) and rechargeable lithium batteries. The surface area S^{PV} covered by the PV panels and the battery capacity S^B are the design parameters defining $p := [S^{PV}, S^B]$. The system state $x(t) := [T(t), SoC(t)]'$ includes the building internal temperature $T(t)$ and the battery state of charge $SoC(t)$.

The input $u(t) := [u^{eH}(t), u^{CeH}(t), u^{dch}(t), u^{ch}(t), u^b(t), u^s(t)]'$ consists of the electricity power $u^{eH}(t)$ and $u^{CeH}(t)$ consumed by the heating and cooling pumps, respectively, the battery charging $u^{ch}(t)$ and discharging $u^{dch}(t)$ rates and the bought $u^b(t)$ and sold $u^s(t)$ power. The uncertain exogenous vector $w(t) := [T^e(t), I(t), c^{el}(t), c^{em}(t)]$ considers the external temperature $T^e(t)$, solar irradiance $I(t)$, electricity prices $c^{el}(t)$ and carbon emissions $c^{em}(t)$. The dynamics of the building are:

$$\begin{bmatrix} \dot{T}(t) \\ \dot{SoC}(t) \end{bmatrix} = \begin{bmatrix} -(UA + \rho_{air} V C_{air}^p n_{ac})/C_{build} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} T(t) \\ SoC(t) \end{bmatrix} + \begin{bmatrix} COP(T^e(t))/C_{build} & -COP_{cool}/C_{build} & 0 & 0 \\ 0 & 0 & -1/\eta^{ds} & \eta^{ch} \end{bmatrix} \times \begin{bmatrix} u^{eH}(t) \\ u^{CeH}(t) \\ u^{dch}(t) \\ u^{ch}(t) \end{bmatrix} + \xi(t), \tag{20}$$

where

$$\xi(t) = \begin{bmatrix} T^e(t)(UA + \rho_{air} V C_{air}^p n_{ac})/C_{build} \\ 0 \end{bmatrix} \tag{21}$$

where $\text{COP}(T^e(t)) = m_{\text{COP}}(T^e(t) - 7) + 3$. The values of all parameters are given in Table 1.

Table 1
Dwelling parameters.

Description	Parameter	Value	Unit
Average U-value	U	0.93195	W/(m^2 K)
Wall surface area	A	82.06959707	m^2
Air density	ρ_{air}	1.225	kg/m^3
Building volume	V	224.05	m^3
Air heat capacity	C^p_{air}	1.005	kJ/(kg K)
Air changes per hour	n_{ac}	1	h^{-1}
Building thermal mass	C_{build}	15286.6114	kJ/K
Floor surface area	S_F	89.62	m^2
HP electricity bound	$\overline{u}^{eH}$	4	kW
CP electricity bound	$\overline{u}^{ceH}$	6	kW
HP capacity	$\overline{Q}^{HP}$	6	kW

The input and state constraints are

$$\underline{T}(t) \le T(t) \le \overline{T}(t) \tag{22}$$

$$0 \le SoC(t) \le S^B \le \overline{SoC} \tag{23}$$

$$0 \le u^{dch}(t), u^{ch}(t) \le S^B/T_{ds} \tag{24}$$

$$P^{PV}(t) = \theta_1(1 + \theta_2 I(t) + \theta_2 T^e(t))I(t)S^{PV} \tag{25}$$

$$u^b(t) - u^s(t) + u^{dch}(t) - u^{ch}(t) + P^{PV}(t) = u^{eH}(t) + u^{CeH}(t) \tag{26}$$

$$0 \le u^b(t) \le \overline{u}^b,\ 0 \le u^s(t) \le \overline{u}^s \tag{27}$$

$$0 \le u^{eH}(t) \le \overline{u}^{eH},\ 0 \le u^{CeH}(t) \le \overline{u}^{CeH} \tag{28}$$

$$\mathrm{COP}(T^e_t)u^{eH}_t \le \overline{Q}^{HP} \tag{29}$$

$$0 \le S^{PV} \le S_F, \tag{30}$$

where $\underline{T}(t)$ and $\overline{T}(t)$ define thermal comfort limits according to standards [35], T_{ds} is the number of hours required to fully discharge the battery at the maximum rate and $P^{PV}(t)$ is the power produced by the PV panels. All the parameters used in the study are reported in Table 2.

The nonlinear function Eq. (25) is a good model of the maximum power generated by PV panels [36,37], while the operating limits refer to the design specs for the multi-crystalline JAP6 4BB module range produced by JA [38].

The stage cost used by the EMPC depends on the electricity prices

$$\ell(x(t), u(t), w(t), t, p) := c^{el}(t)u^b(t) - 0.9c^{el}(t)u^s(t) + c_{CO_2}c^{em}(t)u^b(t) \tag{31}$$

where $c^{el}(t)$ are time-varying electricity prices, c_{CO_2} is the carbon price.

The expenditures $V_I(p) = c_B S^B + c_{PV} S^{PV}$ faced to buy the technologies use annualised capital costs c_B and c_{PV} computed by dividing the capital cost (CAPEX) by the "present value of annuity factor":

$$a_{y,r} = \frac{1 - \frac{1}{(1+r)^y}}{r}$$

Table 2
Problem parameters.

Description	Parameter	Value	Unit
HP COP slope	m_{COP}	0.067	°C
CP COP	COP_{cool}	0.7	-
Battery charging	η^{ch}	0.88	-
Battery discharging	η^{ds}	0.88	-
Discharging hours	T_{ds}	2	h
Bought power bound	$\overline{u}^b$	30	kW
Sold power bound	$\overline{u}^s$	30	kW
Max battery size	$\overline{SoC}$	60	kWh
Power/(I_r) gain	θ_1	0.12	kW/m^2
Power/(I_r) correction	θ_2	$-1.345e^{-4}$	-
Power/($T\ I_r$) correction	θ_2	$-3.25e^{-3}$	-
Carbon price	c_{CO_2}	100	/(ton CO2e)
Battery lifespan	y	15	years
PV lifespan	y	30	years
Battery CAPEX	C_B	460	/kWh
PV CAPEX	C_{PV}	325	/m^2
Interest rate	i_r	2 %	-

Table 3
Deterministic sizing for $T^{(d)}_k = T^{(s)} = 15$ min for all $k = 0, 1, \dots, N$ and $t_f = 24$ h.

Data year	Battery (kWh)	PV area (m^2)	Optimal cost (£/year)	Mean cost(£/year)
2008	60	89.0	-563.6	415.3
2009	60	0.0	248.1	491.0
2010	0	3.4	451.1	386.1
2011	0	89.0	261.3	379.5
2012	0	8.4	411.8	381.6
2013	51	89.0	350.4	393.9
2014	0	1.7	316.4	388.1
2015	0	1.7	309.6	388.1
2016	60	5.6	237.6	491.0
2017	0	0.0	382.4	390.5
2018	9	89.0	208.8	361.2

considering the technology lifespan and the interest rate reported in Table 2. The technologies come in units of 1 kWh for the battery capacity and 1.68 m^2 for the PV panel dimension.

4.1.2. Pricing and weather data

The time-varying electricity prices $c^{el}(t)$ in Eq. (31) are piece-wise constant with 15 minutes resolution. The price data assumes the Octopus Agile tariff pricing mechanism [39] using the Market Index Price and data from BSC [40]. The significant volatility of the prices is illustrated by the boxplots and histograms in Fig. 5, which shows data over 11 years grouped by month. The histograms also highlight that rare events are not uncommon, confirming the importance of developing a design approach that considers their impact on overall performance. Weather data have been obtained from the Centre for Environmental Analysis (CEDA) archive [41]. Fig. 7 reports the monthly boxplots and frequency distributions of the temperature. Fig. 6 reports irradiance frequency and relative boxplots without zero irradiance values to enhance the visibility of non-zero values. The grid CO2 intensity is based on data from the Carbon Intensity API developed by the ESO National Grid [42]. Statistics on Carbon prices are illustrated in Fig. 8 The adopted risk measure in the cost Eq. (12) is the expectation over 11 scenarios corresponding to the available data. Energy demand consists of the energy needed to meet the imposed thermal comfort requirements. It is determined by the MPC controller, which manages energy storage, consumption, and comfort while optimising economic performance.

4.2. Numerical results

The framework has been implemented in MATLAB. In particular, the EMPC formulation realising the closed-loop simulation uses ICLOCS2.5 [43] with the solver IPOPT [44] while NOMAD [45] has been used to solve the black-box optimisation problems. The adopted transcription method is explicit Euler since it allows input discontinuities. The clusters were computed using the MATLAB function `kmedoids`. The studies were performed on a server with an AMD EPYC 7443 24-core processor running Windows Server 2019.

4.2.1. Deterministic and robust solution

In the first study, we compare the technology sizing using multiple scenarios against the deterministic case to demonstrate the importance of considering different operating conditions. For this study we have set $T^{(d)}_k = T^{(s)} = 15$ min for all $k = 0, 1, \dots, N$ and $t_f = 24$ h.

Table 3 reports 11 optimal technology designs according to a deterministic formulation using the information of a single year. For each year, the mean cost in Table 3 corresponds to the annual mean cost achieved by applying the obtained optimal deterministic design across the 11 years with the different realisations of the exogenous signals. The results show a large variability in the size choice depending on the year under consideration. The mean cost over the considered 11 years is much higher than the optimal cost of the planning phase.

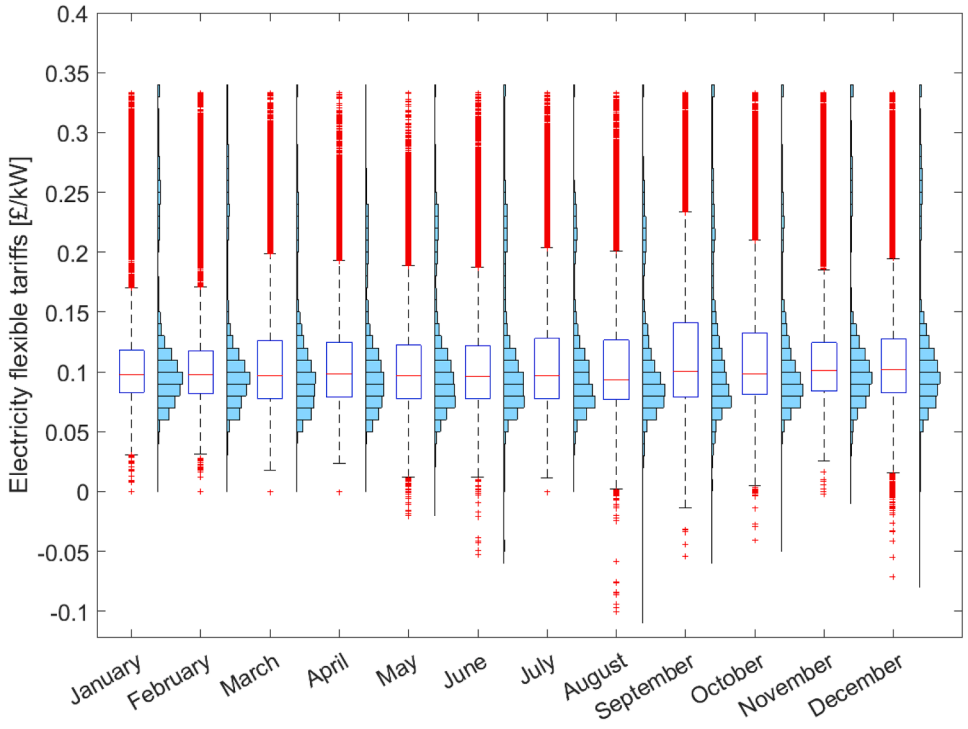


Fig. 5. Statistics of electricity prices.

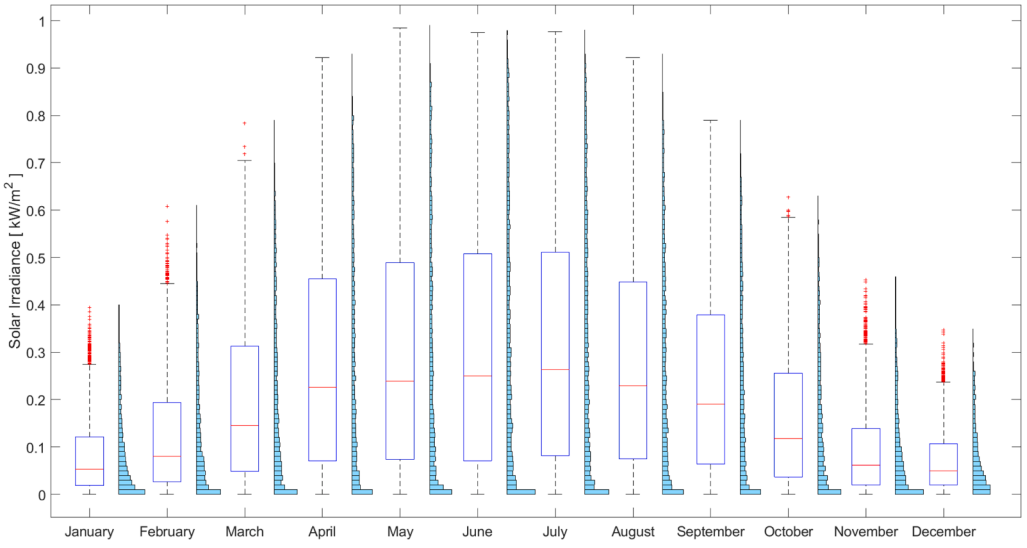


Fig. 6. Statistics of irradiance.

Table 4
Technology sizing using robust formulations with pre-defined EMPC parameters $T_k^{(d)} = T^{(s)} = 15$ min for all $k = 0, 1, \ldots, N$ and $t_f = 24$ h.

Co-design Problem (Pb)	Scaling in sub-sampling	Battery (kWh)	PV area (m^2)	Effective mean cost(£/year)	Estimated mean cost(£/year)	Comp. time (days)
Pb Eq. (16)	-	16	89.0	347.1	-	45
Pb Eq. (17) $n_c = 574$	no	13	89.0	347.6	356.7	30
Pb Eq. (17) $n_c = 50$	no	23	89.0	349.3	378.2	4
Pb Eq. (17) $n_c = 50$	yes	14	89.0	347.3	303.4	2.8

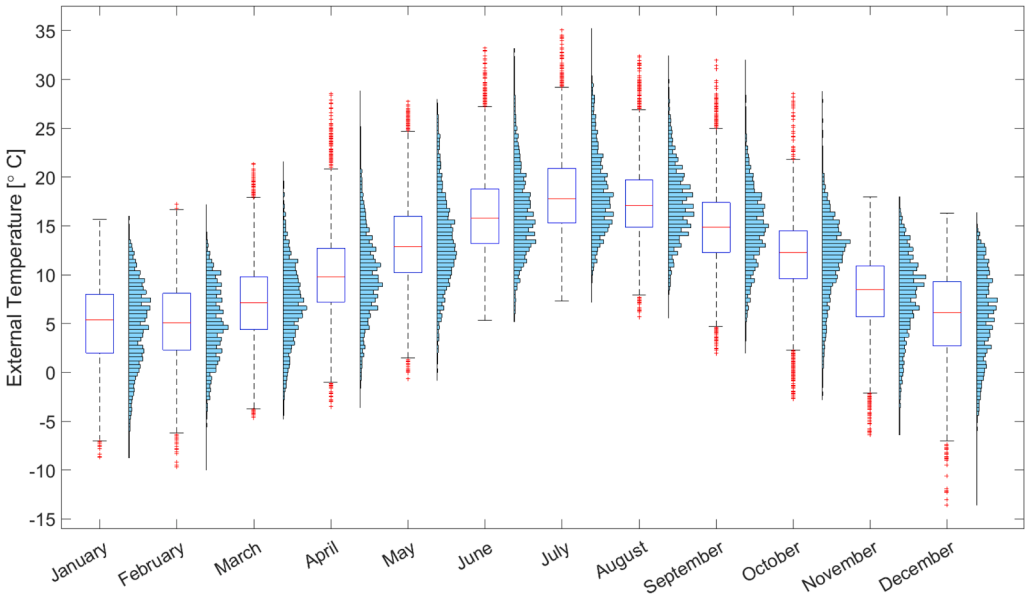


Fig. 7. Statistics of the external temperature.

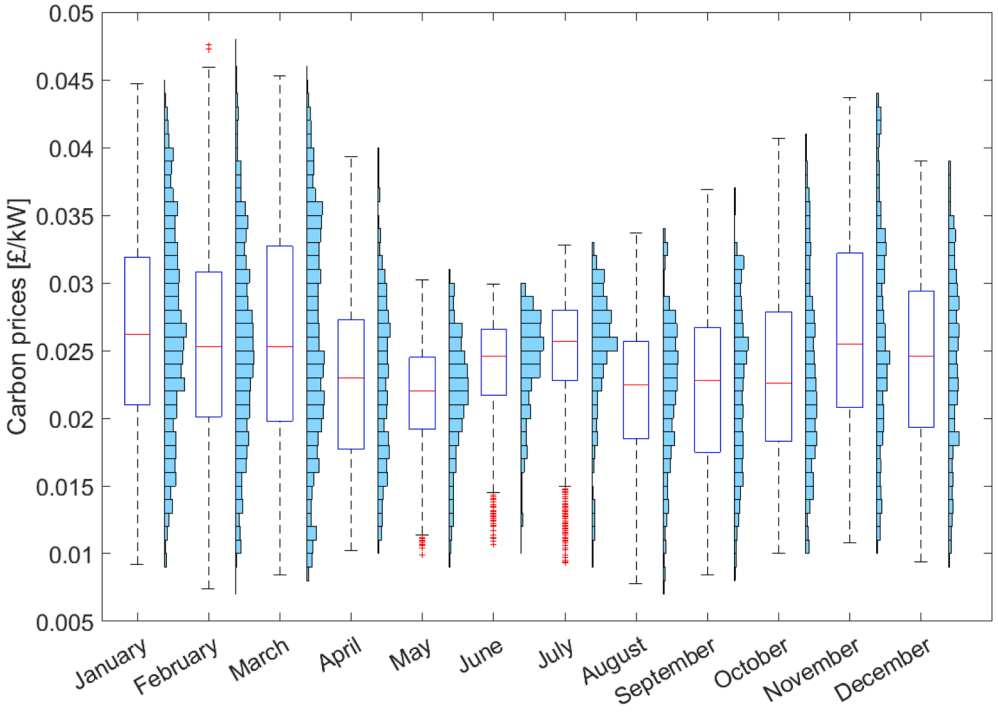


Fig. 8. Statistics of carbon prices.

Conversely, the solution returned by the robust formulation, reported in Table 4, is substantially different from the solutions to the deterministic problem. In Table 4, the cost denoted as "Effective mean cost" is the optimal cost achieved by solving the co-design problem Eq. (16) using the mean as the risk measure and optimising the parameters p or fixing the value of p to the solution returned by the optimisation problem Eq. (19). The optimal cost obtained by solving Eq. (19) for different cluster choices appears in Table 4 as "Estimated mean cost". The robust formulations achieve a better cost in all cases, up to a reduction of about 30 % in some cases.

Table 4 also compares the optimal sizing and the computational times obtained by solving the co-design problem Eq. (16) warm started with the value $p = [9,\ 44]$ against the robust co-design problem Eq. (19) by exploiting parallel computation. Despite the warm starting, solving Eq. (16) required over 10 times more computational time (last column) than using the proposed approach from Eq. (19). Moreover, the decomposed robust formulation Eq. (19) returns optimal solutions with performance comparable to the optimal solution to Eq. (16).

The problem considering a parallel implementation over 574 subsamples returns an optimal solution close to the original problem in terms of technology sizes and optimal cost. In particular, the solution to problem Eq. (17), considering 574 week-long subsamples (second row in Table 4), returns a mean cost of 356.7 £/year which is a close estimate of the effective cost of 347.6 £/year achieved by the optimisation problem Eq. (16) using a battery of 13 kWh and 89.0 m^2 of PV panels (first row in Table 4). The difference in the optimal solution is due to the building's

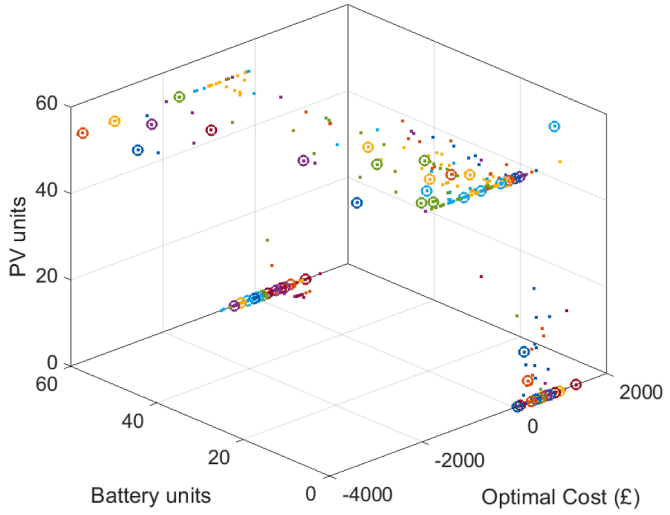


Fig. 9. Data points (coloured dots) consisting in the optimal cost and technologies' size solution to Eq. (19) clustered in 50 groups differentiated by colours. Circles indicate the clusters' centroids. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

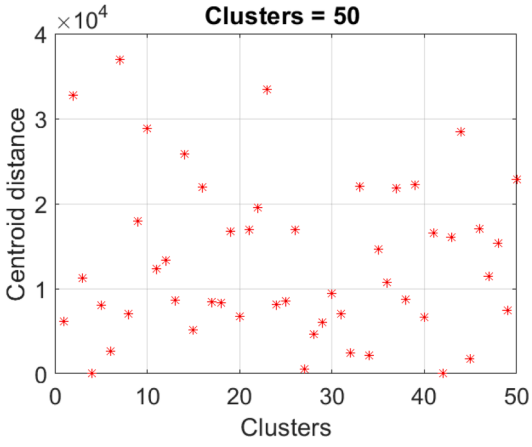


Fig. 10. Within-cluster sums of point-to-medoid distances of the clusters in Fig. 9.

initial state condition. The potential implications of the sensitivity will be discussed in Section 5.1.

Table 4 also reports the optimal design applying the importance subsample technique described in Section 3.3. The computational time is about 10 times lower compared to problem Eq. (16), with a deterioration in the performance of only 0.6 %.

4.2.2. Clustering performance

The clusters are determined on data points considering the optimal costs and sizes computed by solving the problem Eq. (19). The output of Algorithm 4 considering 50 groups is reported in Fig. 9.

The large variability in the values of the optimal costs gives a large sum of distances between a data point and the center of its cluster, as shown in Fig. 10. In particular, Fig. 10 reports the sum of the Euclidean distances from the centers, called centroids, considering 50 clusters.

The cluster spread can be only reduced by substantially increasing the number of clusters, as shown in Fig. 11.

Moreover, the clustering algorithm returns different outputs every time the algorithm is executed, as shown in Fig. 11, depending on the choice of the first centroid. Even if the clustering routine uses the K-means++ algorithm [46] to avoid the problem of sensitivity to the initialisation, different runs of the clustering algorithm return different centroids with substantially different sizing solutions. Also, note that in Fig. 9, the selected centroids do not adequately represent all data points due to the dominance of the high-cost values in the clustering procedure. Consequently, the clustering has been performed on data points with a scaled cost assuming values in $[-60,\ 60]$ to assign the same importance to all the quantities and reduce sensitivity to the initialisation. The centroids, obtained by rescaling the data points, cover the space more uniformly, reducing the sizing's sensitivity to the specific clustering output. Scaling essentially defines the importance of a quantity in the clustering process and regularises data considering their semantic meaning.

Fig. 12(a) shows the sum of the squared distances. The optimal number of clusters is commonly identified as the value associated with a sharp slope change (the Elbow method [47]) in the sum of squared distance as a function of the number of clusters. However, Fig. 12(a) does not show any sharp slope change. Thus, the elbow method is inconclusive because the error keeps decreasing with the number of clusters. Instead, the maximum of the summed point-to-centroid distances in Fig. 12(b) highlights the variability of the clustering outcome. Figs. 12(b) and 13 suggest that a clusters' number larger than 50 reduces uncertainty, because the maximal centroid distance is less spread along the vertical axis for all experiments. Each boxplot in Fig. 13 considers 100 runs of the clustering algorithm using the maximum of the summed point-to-centroid distances.

4.2.3. Technology-specific MPC tuning

An interesting outcome of the case studies is that the choice of MPC parameters depends on the technology size. The MPC tuning of the sampling time, discretisation step and prediction horizon, considering the constraints induced by the discontinuities at the changes in the electricity prices, uses the discrete variables n_s, n_x and n_f characterising the variables of interest

$$T^{(s)} := n_s T^{(d)} \tag{32}$$

$$\delta_T := n_x T^{(s)} \tag{33}$$

$$t_f = n_f \tag{34}$$

where δ_T denotes the time interval at which the change in the price occurs. A resolution of an hour in the prediction horizon is reasonable for the co-design problem as a first approximation to limit the computational burden. The discretisation step is implicitly defined as

$$T^{(d)} = \frac{\delta_T}{n_x n_s}.$$

The physical limitations on the tuning parameters induce the following bounds on the decision variables

$$1 \le n_s \le \frac{\delta_T}{\underline{T}^{(d)}} \tag{35}$$

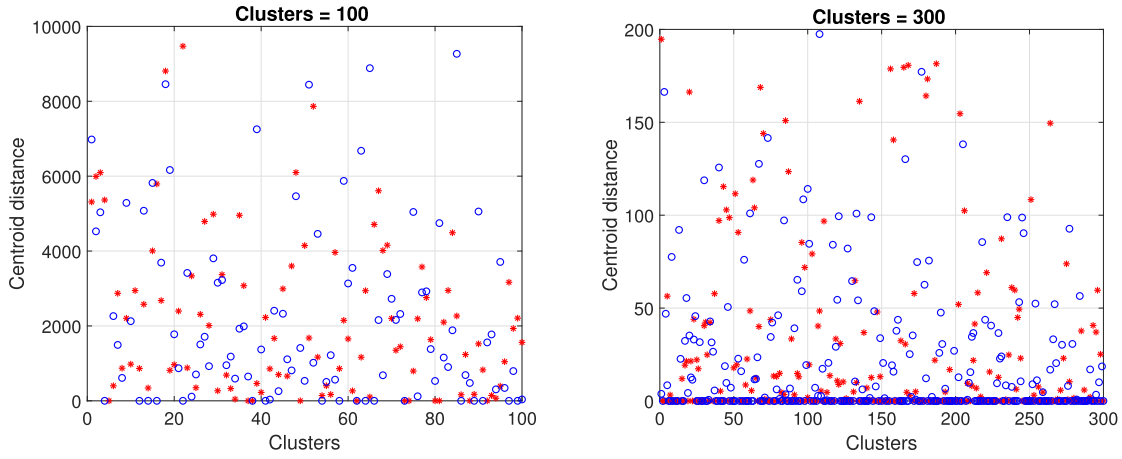
$$1 \le n_x \le \frac{\delta_T}{\underline{T}^{(d)}} \tag{36}$$

$$n_s n_x \le \frac{\delta_T}{\underline{T}^{(d)}} \tag{37}$$

where $\underline{T}^{(d)}$ denotes the lower bound of the discretization step. The performed studies consider as objective Eq. (15) $Q(p_c) := n_f - 1/(4n_x) - 1/(4n_x n_s)$ with $p_c = [n_s, n_x, n_f]$, corresponding to minimising the prediction horizon and maximising the sampling time and the discretisation step. The value of $\underline{T}^{(d)}$ is 5 minutes, and the reference trajectory used in the objective $J^{(2)}(p, p_c)$ considers $T^{(d)} = T^{(s)} = 5$ min and $t_f = 3$ days. The tuning algorithm runs the closed-loop MPC considering a whole week.

Fig. 14 shows the Pareto fronts for two different technology sizes.

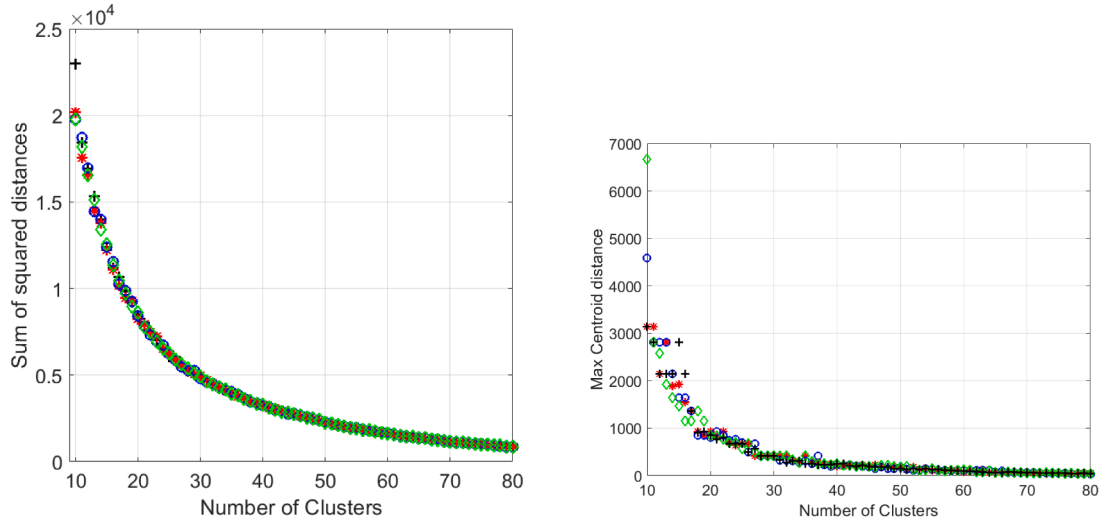
It is interesting to observe that the choice of the MPC parameters depends on the technology size. The co-design requires a controller that



(a) Distances for 100 clusters. Different markers indicate the outcome of different runs of the clustering algorithm.

(b) Distances for 300 clusters. Different markers indicate the outcome of different runs of the clustering algorithm.

Fig. 11. Within-cluster sums of point-to-medoid distances for different numbers of clusters' choices.



(a) Sum of errors as a function of the number of clusters with scaling for multiple experiments

(b) Maximum of errors as functions of the number of clusters with scaling for multiple experiments

Fig. 12. Error analysis.

provides good performance for all possible sizes. Conversely, once the system design is completed, the controller parameter needs to perform well only for the specific size. Consequently, the final decision can compromise accuracy with real-time computational efficiency. The results indicate that a closed-loop cost error of £0.15 per week can be achieved by $T^{(d)} = 5$ min $T^{(s)} = 15$ min and $t_f = 17$, which corresponds to a solution considering the maximum size of the technologies used. Instead, the choice of $T^{(d)} = T^{(s)} = 15$ min and $t_f = 24$ hours gives a closed-loop cost of £1.31 per week Fig. 15. Such a choice provides a closed-loop error of 0.046 per week if the technology's size is halved, and a closed-loop error of £0.15 per week can be achieved for a smaller prediction horizon of 15 and 16 hours. The Pareto front of the cost error performed only considering the prediction horizon as a tuning parameter is illustrated in Fig. 15. The studies highlight the importance of tuning the controller in an integrated fashion with the building design.

4.3. Implementation challenges

The implementation of the designed buildings under the proposed framework presents several challenges, the main one being the practical implementation of the MPC. Despite the extensive body of research demonstrating the potential benefits of MPC, its widespread adoption in the building industry is limited. The limited commercial implementation of MPC is due to the complexities of the predictive model design, the time constraints associated with developing and deploying these models, and the computational complexities that arise during online optimisation. However, the practical MPC implementation does not need to use the same prediction model used for the design but can rely on data-driven approaches [48–50].

Data-driven approaches offer a pathway to reduce the labour-intensive effort associated with modelling and improve scalability. The feasibility of implementing Model Predictive Control locally within

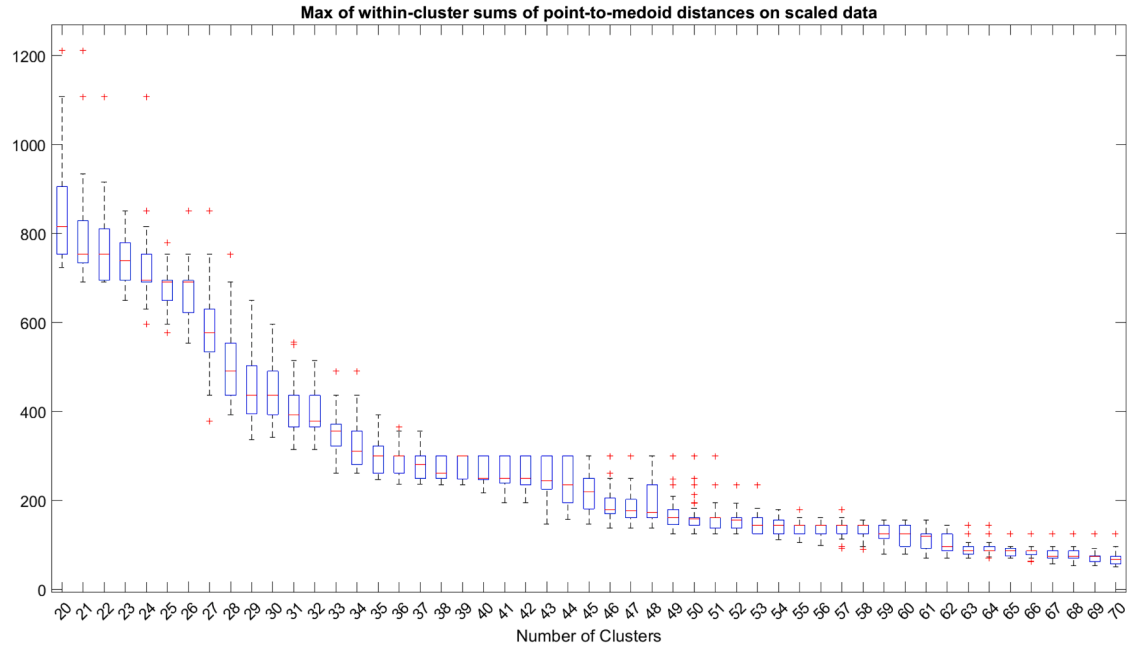


Fig. 13. Maximum of errors' boxplots as functions of the number of clusters with scaling for multiple experiments.

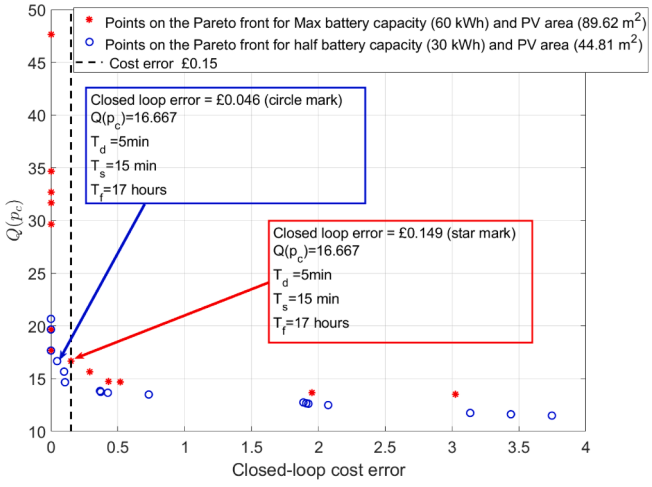


Fig. 14. MPC tuning - Pareto front for the maximum technology size (red *) and half technology size (blue o). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

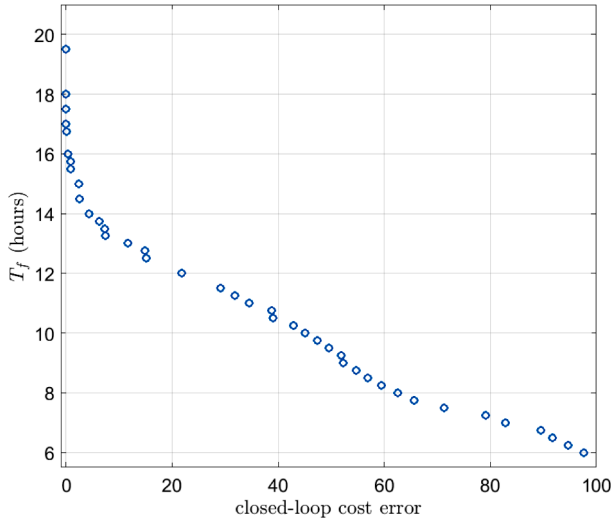


Fig. 15. The Pareto front for tuning of the prediction horizon T_f with $T_k^{(d)} = T^{(s)} = 15$ min.

buildings has been significantly enhanced by the continuous advancements in the computational power of Building Automation Systems (BAS) and the growing accessibility of monitored building data [51].

Furthermore, the proliferation of smart meters and sensors within buildings has led to an exponential increase in the availability of real-time operational data, providing the necessary information for MPC algorithms to make informed control decisions [51]. In parallel with these hardware advancements, the field of Artificial Intelligence (AI), particularly machine learning (ML) and deep learning (DL) techniques, has witnessed remarkable progress; these AI-based control algorithms are often used to enhance accuracy and robustness in building energy management systems by effectively modeling complex energy patterns, predicting demand fluctuations, and optimising control actions.

The implementation of MPC relies heavily on the availability of accurate and timely data, particularly real-time electricity pricing and short-term weather forecasts. Various sources, from public to commercial, offer APIs that provide this information, with different pricing models and levels of accessibility. Time-of-use electricity tariffs with public APIs are becoming more widely available, offering users access to dynamic pricing data (Table 5). While the availability of electricity pricing and weather forecast data is a crucial prerequisite for MPC, the successful implementation of these advanced control strategies also hinges on the ability to integrate these data streams with other relevant information from within the building, such as occupancy levels, sensor readings from HVAC and lighting systems, and data from the Building Management System (BMS).

This integration process often presents a significant challenge due to the inherent diversity in data formats, communication protocols, and update frequencies across these various sources. Furthermore, the raw data obtained from sensors and external APIs often requires extensive

Table 5
Electricity (E) and weather (W) data providers.

Provider	Data type	Geographic coverage	Cost	Key features
Agile Octopus	E	London	Based on wholesale cost + premium (4-7pm)	Dynamic pricing based on half-hourly wholesale energy prices
EPEX SPOT	E	UK (London)	Subscription (contact sales for details)	Day-ahead and intraday auction data, real-time market data services
Trading Economics	E	UK (London)	API Gateway (Subscription plans available)	Real-time spot benchmark electricity price
FlatPeak	E	Global (UK)	Not specified	Real-time access to customer's tariff and grid carbon intensity
Commodities-API	E	EU (UK)	Subscription	Real-time and historical electricity rates data
LSEG	E	Europe (UK)	Subscription (contact sales for details)	Daily and weekly OTC price assessments, wholesale prices, renewable energy info
UK Elec Costs API	E	Great Britain (London)	Free	Estimated half-hourly electricity costs
Weatherstack	W	Global (London)	Free (limited), Paid (from $8.99/month)	Real-time, historical, up to 14-day forecast
WeatherAPI	W	Global (London)	Free (limited), Paid (from $7/month)	Real-time, 3 to 14-day forecast, historical data
OpenWeatherMap	W	Global (London)	Free (limited), Paid (from $40/month)	Real-time, hourly/daily forecast, historical data, weather maps
Visual Crossing	W	Global (London)	Free (limited), Paid (from $35/month)	Real-time, historical, 15-day forecast, climate normals
Meteomatics	W	Global (London)	Quote-based	Real-time, forecasts (up to 15 days), historical, climate data, various models

preprocessing, cleaning, and validation to ensure its quality and accuracy before MPC algorithms can effectively utilize it [52]. Issues such as missing data points, sensor noise, and data inconsistencies can compromise the reliability of the MPC system if not adequately addressed. Another layer of complexity is introduced when integrating newer MPC technologies with existing building automation systems, particularly legacy systems that may employ outdated or proprietary communication protocols incompatible with modern open standards. Achieving data exchange between these disparate systems might necessitate using specialized gateways or middleware solutions capable of translating data between different protocols, adding to the complexity and cost of implementation. The increasing reliance on data collection and analysis for optimising building energy management through MPC raises ethical and privacy concerns that must be addressed [53].

Robust data privacy and security measures, such as anonymization techniques to de-identify personal information, strong encryption protocols to secure data both at rest and during transmission, and strict access controls to prevent unauthorized access or misuse, are essential for mitigating privacy risks and ensuring compliance with data protection regulations like the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA).

Addressing implementation challenges requires a multidisciplinary effort involving control engineers, building scientists, data scientists, policymakers, and building owners working collaboratively to advance the field and facilitate the transition towards a more energy-efficient and sustainable built environment.

Another important implementation aspect is ensuring reliable operation under unexpected conditions within the proposed MPC framework. This relies on a combination of robust co-design that accounts for uncertainties, thorough validation against unseen scenarios, the potential integration of safety fail-safe mechanisms, and the system's ability to adapt to changing conditions over time. The framework aims to create a system that is inherently more resilient to variations by optimising the physical design, the controller, and the building's operation in an integrated manner. The framework includes a validation step where the performance of the optimised system is tested against different datasets not used in the design process. This step is crucial for assessing the robustness and adaptability of the designed system to unseen conditions. If the performance is unsatisfactory under specific unexpected scenarios encountered during validation, the framework allows for incorporating these problematic scenarios and re-optimising the system enhancing, system reliability. Another aspect to consider for reliable operation is the intrinsic adaptability of the data-driven MPC system over time, adjusting control parameters in response to changes in the building's behavior or external conditions.

5. Discussion and conclusions

The achievement of net-zero carbon emissions requires decarbonising the entire housing stock. Improved environmental performance is achieved by including renewable energy sources in the way buildings are operated. However, the intermittent nature of the renewables often requires additional equipment, such as batteries. Choosing appropriate equipment is a trade-off between the needs of the building and the costs. Furthermore, the environmental performance of a building with the additional equipment will be affected by the way the building is operated. Model Predictive Controllers have been widely used in building control because they leverage predictions of future operating conditions to ensure optimal operation while satisfying constraints [2]. In particular, long prediction horizons allow accounting for weather seasonality and uncertainty in energy prices, inherent in the modern energy landscape. However, MPC performance is often limited by available computational capacities, because a long prediction horizon leads to large optimal control problems that need to be solved numerically. We have presented a robust framework to simultaneously optimise the design, controller, and operation of residential buildings, considering external weather conditions and electricity prices that vary over time.

Scalability of the proposed co-design framework is a known challenge due to the inherent exponential complexity introduced by high-dimensional systems, multiple long temporal scales and uncertainty modelling. However, the framework's architecture and modular decomposition are designed to mitigate complexities due to multiple temporal scales and uncertainties and enable potential applications to larger systems, such as commercial buildings or even components of smart city infrastructure. The framework decouples the tuning of the controller

(MPC) from the sizing problem by leveraging importance-based subsample selection, allowing larger systems to be handled by solving smaller subproblems concurrently, rather than relying on a monolithic formulation. With careful aggregation of thermal zones and technology clusters, the framework can be adapted to commercial settings without a prohibitive increase in computational burden.

From an economic perspective, the proposed co-design framework aligns with national and international energy objectives. It can enable long-term operational savings and higher grid responsiveness by leveraging dynamic electricity tariffs, PV-battery storage, and demand flexibility. However, financial barriers such as upfront capital requirements and technical complexity may limit near-term adoption. Policy tools, such as subsidies, performance-based incentives, and integrated design software, could lower these barriers, encouraging developers and homeowners to adopt such optimisation techniques in practice.

5.1. Discussion

Case studies demonstrate the ability of the presented co-design framework to seek trade-offs in an integrated fashion with a temporal resolution spanning years to minutes. In particular, the case study reported lower costs for the robust co-design framework than for the deterministic approach in all cases, up to a reduction of about 30 % in some cases. The developed approximations and solution approaches report a computational time reduction at least 10 times lower than the original problem, with a performance deterioration of only 0.6 %. The obtained computational time improvements make the presented framework suitable for extension to non-residential/ commercial cases.

The results show the optimal solution's sensitivity to data and initial conditions. The optimal solution to the problem that simulates the closed-loop for short periods is very sensitive to the choice of the initial condition. The initial condition describes the energy available in the building at the beginning of the period, and its value affects the overall control outcome. The performed studies consider an empty battery and the temperature at its minimum value at the beginning of each subsample. The used initial state describes the worst condition in the winter season and adds a degree of robustness to the design process. The computation time achieved by the parallel implementation, including all the available data, is lower but still substantial.

The high sensitivity also demonstrates that the range of price variations is such that the value of the initial energy stored is comparable to the savings achieved. It also suggests that the combination of the considered technologies is only convenient in highly dynamic and uncertain electricity markets unless other sources of revenue, such as ancillary services, are accounted for as possible additional income.

5.2. Conclusions

The proposed framework allows solving the co-design problem over a longer time horizon than traditional approaches. By selecting relevant samples from the entire dataset, the approach limits the size of the optimisation problem that needs to be solved, thus improving computational efficiency.

The simultaneous optimisation of the design, control and operation of a building considering uncertainty is a computationally challenging optimisation problem. The challenges are primarily related to the multi-objective nature of the operation of a building, as well as to long operating timescales and corresponding exogenous data. The framework proposed in this paper mitigates the computational complexity by decoupling multiple objectives and iteratively solving reduced-size optimisation problems. Furthermore, to reduce complexity further, we sample the data sets of the exogenous data to include data with the most critical information for the decision process.

The performance of the proposed framework was validated for a residential building. However, the framework was developed for co-design for Model Predictive Control regardless of the application domain. Thus, the framework can be used for co-design in other systems with complicated dynamics and long-time operating horizon, such as power systems, modern transports, robotics, medical devices and manufacturing processes. Overall, the flexibility of MPC and the optimal design of residential buildings indicate that the presented framework is a good candidate for future work. The importance subsample approach and the MPC tuning algorithm require further analysis and improvements in terms of their accuracy and computational performance.

Future work could explore the integration of hybrid optimisation methods, such as surrogate-assisted metaheuristics and multi-objective evolutionary algorithms, which could enhance scalability in high-dimensional settings. Additionally, real-world implementation will require model calibration using sensor data and a robust interface with building automation systems.

6. Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Gemini to find relevant information in the literature and Co-Pilot to obtain some figures in TikZ. After using this tool/service, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Paola Falugi: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization; **Edward O'Dwyer:** Writing – review & editing, Validation, Methodology, Investigation, Data curation, Conceptualization; **Marta Zagorowska:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization; **Eric Kerrigan:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization; **Yuanbo Nie:** Writing – review & editing, Software; **Goran Strbac:** Project administration, Writing – review & editing, Supervision, Resources; **Nilay Shah:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Paola Falugi reports financial support was provided by Engineering and Physical Sciences Research Council. Marta A. Zagorowska reports financial support was provided by Engineering and Physical Sciences Research Council. Edward O'Dwyer reports financial support was provided by Engineering and Physical Sciences Research Council. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Edward O'Dwyer is now working for Octopus Energy.

References

[1] W.G.B. Council, The 2020 Global Status Report for Buildings and Construction, 2020. https://worldgbc.org/article/launch-of-the-2020-global-status-report-for-buildings-and-construction/, accessed 13 Nov 2024.

[2] J. Drgoňa, J. Arroyo, I. Cupeiro Figueroa, D. Blum, K. Arendt, D. Kim, E.P. Ollé, J. Oravec, M. Wetter, D.L. Vrabie, L. Helsen, All you need to know about model predictive control for buildings, Ann. Rev. Control 50 (2020) 190–232.

[3] D. Angeli, A. Casavola, F. Tedesco, Theoretical advances on economic model predictive control with time-varying costs, Ann. Rev. Control 41 (2016) 218–224.
[4] M.J. Risbeck, J.B. Rawlings, Economic model predictive control for time-Varying cost and peak demand charge optimization, IEEE Trans. Automat. Contr. 65 (2020) 2957–2968.
[5] B. Khusainov, E.C. Kerrigan, G.A. Constantinides, Automatic software and computing hardware codesign for predictive control, IEEE Trans. Contr. Syst. Technol. 27 (2019) 2295–2304.
[6] M. Garcia-Sanz, Control co-Design: an engineering game changer, Adv. Contr. Applic. 1 (2019) e18.
[7] S.S. Rao, Combined structural and control optimization of flexible structures, Eng. Optim. 13 (1988) 1–16.
[8] A. Suardi, S. Longo, E.C. Kerrigan, G.A. Constantinides, Energy-aware MPC co-design for DC-DC converters, in: European Control Conference, 2013, pp. 3608–3613.
[9] K.J. Kircher, K.M. Zhang, Convexity and monotonicity in nonlinear optimal control under uncertainty, 2019. 1904.00987
[10] N.A. Diangelakis, B. Burnak, J. Katz, E.N. Pistikopoulos, Process design and control optimization: a simultaneous approach by multi-parametric programming, AIChE J. 63 (11) (2017) 4827–4846.
[11] G.P. Henze, B. Biffar, D. Kohn, M.P. Becker, Optimal design and operation of a thermal storage system for a chilled water plant serving pharmaceutical buildings, Eng. Optim. 40 (2008) 1004–1019.
[12] K.M. Powell, W.J. Cole, U.F. Ekarika, T.F. Edgar, Optimal chiller loading in a district cooling system with thermal energy storage, Energy 40 (2013) 445–453.
[13] M. Hoffmann, L. Kotzur, D. Stolten, M. Robinius, A review on time series aggregation methods for energy system models, Energies 13 (2020).
[14] A. Hilbers, D. Brayshaw, A. Gandy, Importance subsampling for power system planning under multi-year demand and weather uncertainty, in: Conference on Probabilistic Methods Applied to Power Systems, 2020, pp. 1–6.
[15] A.P. Hilbers, D.J. Brayshaw, A. Gandy, Importance subsampling: improving power system planning under climate-based uncertainty, Appl. Energy 251 (2019) 113–114.
[16] P. Falugi, E. O'Dwyer, E.C. Kerrigan, E. Atam, M.A. Zagorowska, G. Strbac, N. Shah, Predictive control co-design for enhancing flexibility in residential housing with battery degradation, IFAC-PapersOnLine 54 (6) (2021) 8–13.
[17] G. Scarciotti, A. Astolfi, Interconnection-based model order reduction - a survey, Eur. J. Contr. 75 (2024) 100929.
[18] M.J.N. Oliveira Panão, N.M. Mateus, G. Carrilho da Graça, Measured and modeled performance of internal mass as a thermal energy battery for energy flexible residential buildings, Appl. Energy 239 (2019) 252–267.
[19] P. Klanatsky, F. Veynandt, C. Heschl, R. Stelzer, P. Zogas, G. Siokas, A. Balomenos, Real long-term performance evaluation of an improved office building operation involving a data-driven model predictive control, Energy Build. 338 (2025) 115590.
[20] J. Yang, H. Wang, L. Cheng, Z. Gao, F. Xu, A review of resistance-capacitance thermal network model in urban building energy simulations, Energy Build. 323 (2024) 114765.
[21] V.S.K.V. Harish, A. Kumar, A review on modeling and simulation of building energy systems, Renew. Sustain. Energy Rev. 56 (2016) 1272–1292.
[22] I. Hazyuk, C. Ghiaus, D. Penhouet, Optimal temperature control of intermittently heated buildings using model predictive control: part I - Building modeling, Build. Environ. 51 (2012) 379–387.
[23] R.T. Rockafellar, Coherent approaches to risk in optimization under uncertainty, INFORMS Tutor. Oper. Res. (2014) 38–61.
[24] J.T. Betts, Practical methods for optimal control and estimation using nonlinear programming: Second edition, Advances in Design and Control, Society for Industrial and Applied Mathematics, 2010.
[25] A.J. Conejo, L.B. Morales, S.J. Kazempour, A.S. Siddiqui, Investment in Electricity Generation and Transmission - Decision Making under Uncertainty, Springer, 2016.
[26] L. Grüne, S. Pirkelmann, Closed-loop performance analysis for economic model predictive control of time-varying systems, in: IEEE Conference on Decision and Control (CDC), 2017, pp. 5563–5569.
[27] Y. Nie, E.C. Kerrigan, Solving dynamic optimization problems to a specified accuracy: an alternating approach using integrated residuals, IEEE Trans. Automat. Contr. (2022) 1–1.
[28] R.T. Marler, J.S. Arora, Survey of multi-objective optimization methods for engineering, Struct. Multidiscip. Optim. 26 (2004) 369–395.
[29] C. Audet, W. Hare, Derivative-free and blackbox optimization, Springer Series in Operations Research and Financial Engineering, Springer International Publishing, 1st ed. 2017. edition, 2017.
[30] M. Zhu, D. Piga, A. Bemporad, C-GLISP: preference-based global optimization under unknown constraints with applications to controller calibration, IEEE Trans. Contr. Syst. Technol. 30 (5) (2022) 2176–2187.
[31] J.C. Willems, P. Rapisarda, I. Markovsky, B.L.M. De Moor, A note on persistency of excitation, Syst. Contr. Lett. 54 (2005) 325–329.
[32] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley & Sons, New York, 2009.
[33] T. Madhulatha, An overview on clustering methods, IOSR J. Eng. 2 (2012).
[34] X. Jin, J. Han, K-Medoids Clustering, Springer US, 2010, pp. 564–565.
[35] CIBSE, Temperature in Indoor Workplaces (Thermal Comfort), 2015. https://www.cibse.org/knowledge/knowledge-items/detail?id=a0q20000006obXh, Accessed: 27 MAY 2021.
[36] R. Dows, E. Gough, PVUSA procurement, acceptance, and rating practices for photovoltaic power plants, Tech. Report, Pacific Gas and Electric Company, 1995.
[37] D. Pepe, G. Bianchini, A. Vicino, Estimating PV forecasting models from power data, in: IEEE International Energy Conference, ENERGYCON, 2018, pp. 1–6.
[38] J.A. S. T.C. Ltd., Product datasheets, 2020. https://www.jasolar.com. Accessed: 26 Oct 2020.
[39] P. Steele, Agile pricing explained, Octopus Energy Ltd, 2019. https://octopus.energy/blog/agile-pricing-explained/, accessed 13 Nov 2024.
[40] E. BSC, Electricity industry Data. https://www.elexon.co.uk/data/, accessed 13 Nov 2024.
[41] C.f.E. Analysis, Weather Data in CEDA Archives, 2021. https://help.ceda.ac.uk/article/96-weather-data, accessed 13 Nov 2024.
[42] N.G. ESO, Carbon Intensity API, 2020,. http://carbonintensity.org.uk/, accessed 13 Nov 2024.
[43] Y. Nie, O. Faqir, E.C. Kerrigan, ICLOCS2: Solve your optimal control problems with less pain, in: Proc. 6th IFAC Conference on Nonlinear Model Predictive Control, 2018, pp. 345–362. 9th–22nd AUG 2018, Madison, Wisconsin (USA).
[44] A. Wächter, L. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, Math. Program. 106 (2006) 25–57.
[45] S. Le Digabel, Algorithm 909: NOMAD: nonlinear optimization with the MADS algorithm, ACM Trans. Math. Softw. 37 (4) (2011) 1–15.
[46] D. Arthur, S. Vassilvitskii, K-means + +: the advantages of careful seeding, in: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2007, pp. 1027–1035.
[47] T.S. Madhulatha, An overview on clustering methods, IOSR J. Eng. 2 (2012) 719–725.
[48] E. O'Dwyer, E.C. Kerrigan, P. Falugi, M. Zagorowska, N. Shah, Data-driven predictive control with improved performance using segmented trajectories, IEEE Trans. Contr. Syst. Technol. 31 (2023) 1355–1365.
[49] H. Zhang, S. Seal, D. Wu, F. Bouffard, B. Boulet, Building energy management with reinforcement learning and model predictive control: a survey, IEEE Access 10 (2022) 27853–27862.
[50] J. Engel, T. Schmitt, T. Rodemann, J. Adamy, Evaluating the impact of data availability on machine learning-augmented MPC for a building energy management system, in: 2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE), IEEE, 2024, p. 1-5.
[51] G. Serale, M. Fiorentini, A. Capozzoli, D. Bernardini, A. Bemporad, Model predictive control (MPC) for enhancing building and HVAC system energy efficiency: problem formulation, applications and opportunities, Energies 11 (2018).
[52] E. Mariam, M. Nader, A.-N. Mohammed, Sensor data validation and fault diagnosis using auto-Associative neural network for HVAC systems, J. Build. Eng. 27 (2020) 100935.
[53] M. Savadkoohi, M. Macarulla, B. Tejedor, M. Casals, Analyzing the implementation of predictive control systems and application of stored data in non-residential buildings, Energy Effic. 17 (2024) 27853–27862.