

Document Version

Final published version

Licence

CC BY

Citation (APA)

Ruiz Manuel, I., Chen, M., Lombardi, F., & Pfenninger, S. (2026). Optimising for the long game: methodological challenges in energy system optimisation pathways. *Applied Energy*, 416(127980), Article 127980. <https://doi.org/10.1016/j.apenergy.2026.127980>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership. Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse




Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



Optimising for the long game: methodological challenges in energy system optimisation pathways

Ivan Ruiz Manuel* , Meijun Chen , Francesco Lombardi, Stefan Pfenninger-Lee 

TU Delft, Faculty of Technology, Policy and Management, Department of Engineering Systems and Services, Jaffalaan 5, Delft, 2628 BX, South Holland, Netherlands

HIGHLIGHTS

- Quantitative assessment of Energy System Optimisation Model literature on long-term pathways published over recent decades.
- Theoretical synthesis of critical aspects to consider when evaluating pathway exercises.
- Evidence of a shift towards near-sighted model setups, driven by the increased use of myopic approaches and shorter model horizons.
- Models have achieved finer operational and spatial detail, but long-term aspects remain coarse.
- Better practices could reduce model distortions and misinterpretations of pathway model results.

ARTICLE INFO

Keywords:

Energy system modelling
Optimisation
Capacity expansion
Pathways
Systematic literature review

ABSTRACT

Pathways that describe the optimal evolution of energy systems across multiple decades are important in energy system research and policy literature, with net-zero and similar climate policies being common drivers behind them. While there are many studies on aspects such as spatial and operational resolution, model features, and model transparency, there has been little attention on the methodological considerations of formulating pathway studies in mathematical optimisation terms, and how these methods have evolved over time. To address this, we conduct a systematic review of optimal pathway literature at or above the national level focusing on the following: i) the implications of model foresight choices, ii) end effects and related issues that may bias model outcomes, iii) trade-offs in model resolution, and iv) investment dynamics. We showcase how modellers have dealt with these aspects in a large sample of studies spanning multiple decades, and provide recommendations to both modellers and model users on identifying issues that can bias model results and how to improve upon them. In particular, we identify opportunities to better balance long-term anticipatory planning with high operational and spatial detail in models, and to improve the communication and systematic treatment of those mathematical design choices that potentially distort model decisions across time.

1. Introduction

Pathways, i.e., trajectories that describe the plausible evolution of a system over a period of years to decades, are important means of planning and assessing changes in techno-economic and socio-technical systems, particularly in the context of mitigating climate change [1,2]. The concept of “pathways” is often used interchangeably with that of “scenarios” [3,4], although the latter can also be used to describe the end state of a possible future without necessarily detailing the developments that lead to it [5]. Pathways can be produced qualitatively by using narratives to describe the interactions between actors and the systems

surrounding them [6], or quantitatively with the aid of computer models in combination with the aforementioned narratives [2].

The transition of energy systems to climate neutrality has been at the centre of many quantitative pathway exercises due to the large contribution of such systems to greenhouse gas (GHG) emissions [7]. Given the highly complex nature of the energy transition, pathways detailing energy investments several decades into the future have become ubiquitous, with governments and international bodies commissioning them as part of policy processes [8]. Examples of pathway exercises widely featured in research and policy documents include the Shared Socioeconomic Pathways (SSPs) [9] and Representative Concentration

* Corresponding author.

Email address: i.ruizmanuel@tudelft.nl (I. Ruiz Manuel).

Pathways (RCPs) [10] produced as part of Intergovernmental Panel on Climate Change (IPCC) reports, the International Energy Agency (IEA)[’s] annually published World Energy Outlook [11] and the U.S. Energy Information Administration (EIA)[’s] International Energy Outlook [12].

One of the most commonly used methods to produce energy system pathways is bottom-up Energy System Optimisation Models (ESOMs), with the aim of minimising monetary investment and operation costs from a central planner perspective [13,14]. This type of goal-oriented — i.e., “normative” — pathway has a storied history in governmental and electric utility planning [15], with models such as TIMES [16], OSeMOSYS [17] and MESSAGE [18] being widely recognised in scientific literature. The approach has also influenced international climate policy, as many of the Integrated Assessment Models (IAMs) used to investigate the long-term evolution of GHG emissions have an Energy System Optimisation Model (ESOM) at their core [19].

The application of ESOMs to highly uncertain long-term questions has been subjected to a range of criticisms, including how their insufficient spatio-temporal resolution biases results against renewable technologies [13], their frequent use of highly subjective parameters and formulations [20], their use of inaccurate or outdated cost projections of renewable technologies [21,22], and a general lack of transparency that obfuscates the interpretation of their results [23,24]. Energy modelling literature has responded to these criticisms by, for example, developing and assessing novel techniques for time series aggregation to better represent the patterns of demand and renewable resources [25,26], by comparing the effects of different model formulations [27] or by classifying models by their scope or features [28–30].

Despite all this attention, the methods and particularities of pathways produced by ESOMs are seldom discussed with the same degree of detail. Nevertheless, producing quantitative pathways is one of the most important applications of such models, and a model’s long-term setup can affect outcomes just as much (or more) as the issues discussed above. For example, assessments of energy system models have noted a general lack of discourse on the topic of model foresight and its implications [20], which is reflected in how reviews of energy models often omit the topic entirely [31–34] or only discuss it tangentially [35]. Yet, studies assessing this aspect regularly highlight its significant influence on model results [36,37]. Similarly, despite the mounting evidence on the importance of models having adequate spatial and temporal resolution [26,30,38], to date there is no systematic analysis of whether models have actually improved in that regard. Other nuances of long-term optimisation model formulations, like the presence and prevention of modelling distortions such as end effects [39,40] or the implications of different technological learning approaches [41] have garnered even less discussion. To our knowledge, there is no systematic assessment of ESOM literature regarding how pathway exercises are conducted and the considerations therein, which is worrisome given that this lack of awareness could result in disconnects between the narrative modellers intend to create within their modelling exercises, the ways an optimisation model will behave in the presence of distortions and other modelling artefacts, and how model results may be interpreted by users who lack enough context on the impacts of modeller choices.

Here we conduct a systematic review of pathway studies using ESOMs with a focus on the methodological aspects of these exercises. In particular, we focus on: i) the assumptions and justifications given when using different types of foresight and their implications; ii) the concept of the decision horizon and how model distortions such as end effects can constrain it; iii) tradeoffs in model resolution; and iv) the techniques used by modellers to influence investment dynamics. By describing these aspects in an accessible way, we hope to provide model developers with insights on how to further improve their frameworks, and to aid model users in contextualising pathway results in the face of weaknesses that might be present in the models that produce them. Our focus is on discussing key aspects of pathway exercises and how they have changed over the years, not on evaluating the sectoral coverage of

models, the quality of their data, or features that may improve the operational or technological formulations within models. Finally, we limit our focus to ESOM studies at national or supranational scale to ensure a reasonably homogeneous sample of studies in terms of policy scope, and to focus on model formulation issues that might impact international pledges such as Nationally Determined Contributions (NDCs). Although this implies spatial mismatches between the nations featured in these models, we believe this is justified as subnational entities are subject to the goals of federal institutions. A comparison of the match between spatial resolution and nation size is outside the scope of this study.

The paper is organised as follows. We start by detailing several methodological considerations of pathway exercises in Section 2, which will serve as a lens to detect potential issues in the current practices emerging from our subsequent literature analysis. Section 3 provides an overview of our systematic review method, including research queries, the protocols and tools utilised, and a description of our approach to data collection. Section 4 compiles the results of our analyses, including bibliometric approaches, and the resulting statistics of our sample of studies. Finally, Section 5 concludes with discussions and recommendations for pathway exercises going forward.

2. Background and theory

Constructing an ESOM involves many choices and assumptions by model developers, which in turn affect the questions the said model is best suited to answer; these choices and assumptions are also important aspects to consider when interpreting results [42]. In this section, we detail multiple aspects we have identified as essential when it comes to pathway formulations in ESOMs, placing particular focus on details often ignored in the literature. Our aim is to sufficiently explain these methodological aspects while providing a consistent nomenclature to describe them. These concepts will guide us in our subsequent analysis of the literature, helping us to detect any current practices that may be overlooking important methodological considerations. In particular, we focus on i) describing model horizons and foresight, ii) detailing end effects and how to mitigate them, iii) explaining resolution tradeoffs with a particular focus on how models aggregate investment decisions to improve tractability, and iv) examining how models regulate investment dynamics.

2.1. Time horizons and foresight

An important concept to keep in mind when evaluating pathways is the time horizon under study and how the model considers it when making decisions. A distinction should be made between *investment horizons*, which are the periods when the model can decide to invest in or remove certain technologies, and *operational horizons*, in which the model can only decide how to best utilise the technologies already at its disposal (Fig. 1).

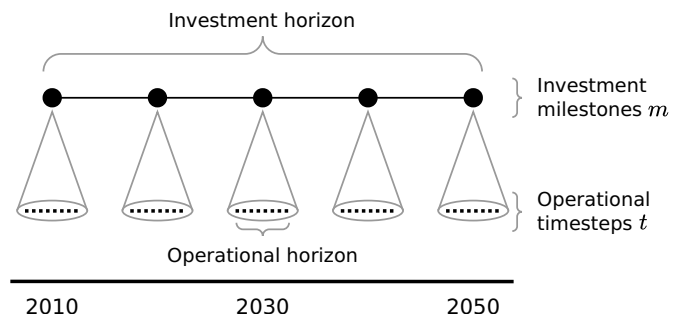


Fig. 1. The different horizons considered in a typical ESOM pathway exercise.

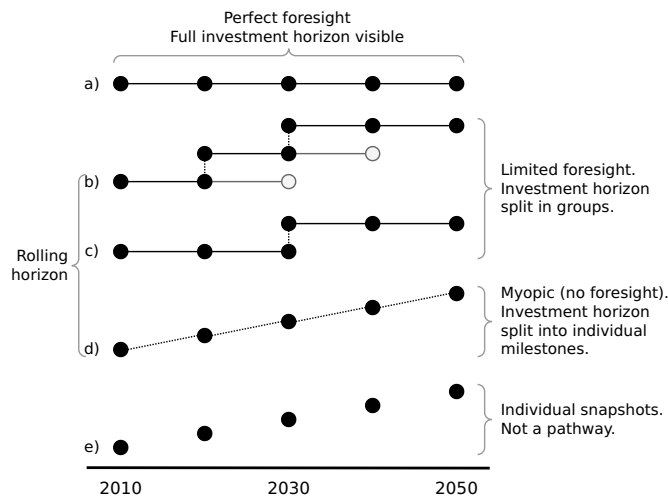


Fig. 2. Classification of typical foresight approaches seen in ESOM pathways. Dots represent model milestones, with white dots representing reconsidered investments. Adapted from [35,37].

Modellers might want to limit a model’s ability to anticipate events depending on their research goals and the computational capabilities at hand. For instance, when evaluating how delayed investment and systemic inertia can affect policy goals [36,43]. This means reducing the model’s *foresight*—often called the planning or decision horizon [37,40]—which is the degree of visibility that the model has into future system conditions. Reduced foresight leads the model to obtain suboptimal results within the bounds of the narrative constructed by the modellers. In general, three kinds of foresight approaches can be distinguished (Fig. 2).

Under *perfect foresight* the model has full visibility of the investment horizon, including price developments, future policies, and decommissioning requirements. It can be thought of as the most normative approach [5], since it is the only one that guarantees optimality within the bounds of the problem. However, perfect anticipation over multiple decades does not reflect reality, meaning this type of exercise is ill-suited to reproduce actual policymaker behaviour [44] and is often overly optimistic when exploring system disruptions [36]. A simplified version of most perfect foresight formulations is shown in Eq. (1), where the model minimises capacity and operational expenditures ($CAPEX_m$ and $OPEX_m$, respectively) multiplied by a monotonically decreasing discount factor (DF_m , discussed further in Section 2.2) for all the milestone years under consideration (m):

$$\text{Min} : \sum_m DF_m (CAPEX_m + OPEX_m) \tag{1}$$

Rolling horizon approaches mitigate an ESOM’s anticipatory behaviour by subdividing the investment horizon into groups and then passing—or “rolling”—results from one group to the next [35]. We can subdivide this category into two distinct methods: *limited foresight* and *myopic foresight*. Under limited foresight, models retain some anticipatory ability and may even be allowed some overlap between groups of decisions to enable reconsideration in the face of new information [37,43]. This means they are in essence a sequence of connected perfect foresight problems. In contrast, fully myopic models essentially lose all anticipatory ability, optimising only single investment periods. In essence, such models assume that agents do not know about or actively decide to ignore expected developments and instead face them as they occur [45]. Eq. (2) shows a simplified equivalent of a myopic model:

$$\text{Min} : CAPEX_m + OPEX_m \tag{2}$$

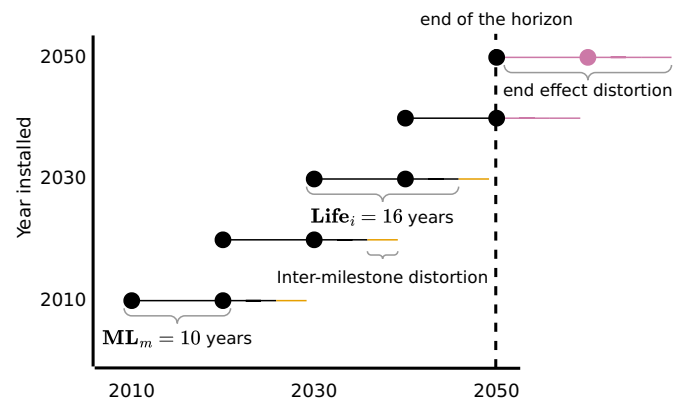


Fig. 3. Simplified example of common model distortions caused by interactions between the lifetime of investments ($Life_i$), foresight horizon (40 years), and milestone length (ML_m) in a model with perfect foresight.

Although myopic models are sometimes seen as more realistic [46], one should keep in mind that these models often retain perfect anticipation for their operational decisions (e.g., [47,48]). Whether or not forgoing anticipatory investment is a desired or realistic characteristic depends on the research question, as studies comparing the effects of foresight in ESOMs have highlighted both the dangers of using perfect foresight models with unrealistic “unicorn” expectations [36], and how myopic planning poses risks of locking-in undesirable system states [43].

2.2. End effects and decision horizons

Another aspect complicating pathway exercises is the finite horizon of the optimisation itself and the distortions it brings. Energy system pathways, like all capacity expansion formulations, are a type of infinite horizon problem because the period in which one needs to install and decommission investments is never-ending [49]. Since solving these models over an infinite horizon is intractable and would involve increasingly uncertain assumptions, modellers are forced to choose a finite period to focus on instead [50–52]. Energy modellers rarely discuss the implications of this choice even though operations research literature has long acknowledged its importance [53].

Making an infinite horizon problem finite results in what is generally called *end effects* [51,52,54], which are distortions that cause the model to make increasingly short-sighted choices the closer it is to the end of its foresight horizon [54]. In the case of ESOMs, these will typically impact investment decisions due to improper consideration of investment lifetimes ($Life_i$) versus their operational costs (Fig. 3), often artificially benefiting options with low initial investment costs but high operational costs such as natural gas turbines [40] or short-term decarbonisation solutions such as co-firing and carbon capture and storage retrofits [52]. In turn, they negatively affect investments with higher installation costs but low operational expenses (e.g., solar photovoltaics or wind turbines), and investments with long lifetimes such as hydropower [52].

The longer a model’s foresight horizon is, the less its end effects will impact earlier decisions [53], meaning that it is necessary to extend the horizon of a model beyond the target year of interest in order to avoid distorted assessments [54]. Operations research often refers to the *decision horizon* as the minimum length of the foresight horizon that guarantees that the first choice of the model (i.e., the first investment year in the case of an ESOM) is unaffected by end effects [55]. This can make ESOMs intractable since they often deal with long-term targets and assets with long lifetimes [40]. Modellers often turn to alternatives to mitigate end effects, and thus reduce the decision horizon, in two ways: by discounting and by adjusting investment cost calculations [52].

Discounting implies using the aforementioned discount factor of Eq. (1) to reduce the present value of future costs based on an assumed

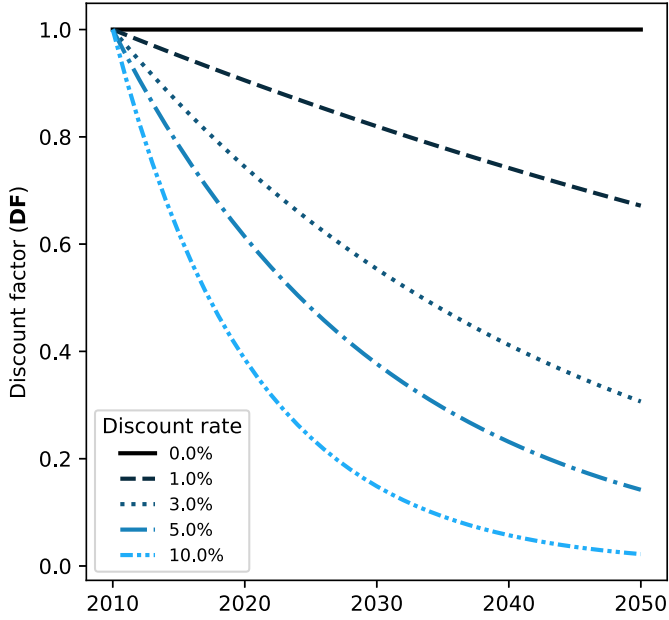


Fig. 4. Discount factor (DF_m) trends at different discount rates, relative to an assumed initial model year of 2010. For illustration, a €1000 investment cost incurred in 2050 corresponds to €671, €142, and €22 in 2010-equivalent terms under discount rates of 1%, 5%, and 10%, respectively.

social discount rate (d), with values between 2–10% being commonly observed in the literature [56], and a typical formulation for DF_m as illustrated in Eq. (3):

$$DF_m = \frac{1}{(1+d)^{m-m_0}} \quad \forall m \in M \quad (3)$$

The choice of the discount rate is complex and value-laden. From a mathematical perspective, higher discount rates will reduce the decision horizon required to make an initial choice by attenuating later costs relative to earlier ones (see Fig. A.1). However, discounting implies intergenerational inequities which are particularly important to consider in the context of climate change [57,58]. ESOMs are particularly sensitive to this parameter, with even small increments in the discount rate leading to significant changes in the discount factor between investment milestones, and thus in the options selected by the model (see Fig. 4). Higher discount rates often delay the implementation of decarbonisation technologies [56], leading to calls to use lower values [59], with some modellers even choosing to use no discounting at all [60].

To reduce artificial disadvantages affecting investments near the end of the horizon—both relative to earlier investments and between late investments themselves—discounting is often complemented by adjustments to investment cost calculations. The most common approaches are salvage values and annuity factors, which influence the timing at which the model accounts for technology investment costs [27,61]. An ESOM without these adjustments will typically estimate total capacity expenditures in a given year ($CAPEX_m$) by multiplying the total newly installed capacity (nc_{imr}) by its overnight installation cost (IC_{im}), as shown in Eq. (4):

$$CAPEX_m = \sum_{i,r} IC_{im} nc_{imr} \quad \forall m \in M \quad (4)$$

In this formulation, technologies closer to the end of the horizon will require installation expenses in full while seeing reduced use, leading to distorted estimates [39]. Studies evaluating end effects often

call this method “truncation” [40,52] and, while problematic, it is not uncommon to see it in the literature (e.g., [36,62,63]).

The first adjustment method, *salvage values*, adjusts Eq. (4) by subtracting a residual for investments with a remaining lifetime at the end of the horizon (SV_{imr}). These residual values attempt to represent the expected benefits that were not captured within the foresight horizon, and are used by popular modelling frameworks such as TIMES [16] and OSeMOSYS [17]. Eqs. (5) and (6) show a simplified example of this method based on [17]:

$$CAPEX_m^{SV} = \sum_{i,r} (IC_{imr} nc_{imr} - SV_{imr}) \quad \forall m \in M \quad (5)$$

$$SV_{imr} = \begin{cases} \text{if } m + \text{Life}_i - 1 \leq m_{max}: 0, \\ \text{else: } IC_{im} nc_{imr} \left(1 - \frac{(1+d)^{m_{max}-m+1} - 1}{(1+d)^{\text{Life}_i - 1}} \right) \end{cases} \quad \forall i \in I, m \in M, r \in R \quad (6)$$

The second adjustment method, *annuity factors* (AF_i), subdivides capacity costs into annuitized yearly payments, often adjusted by some investment-specific weighted average cost of capital ($WACC_i$) [27,64]. Technologies are compared only by accounting for the payments up to the end-of-horizon. This formulation is also popular, with modelling frameworks such as TEMOA [64] and PyPSA [60] making use of it. Eqs. (7)–(9) show an example of this method based on [64]:

$$CAPEX_m^{AF} = \sum_{i,r} (AF_i IC_{imr} nc_{imr}) \quad \forall m \in M \quad (7)$$

$$AF_i = \frac{WACC_i}{1 - (1+WACC_i)^{\text{LifeEoH}_{im}}} \quad \forall i \in I \quad (8)$$

$$\text{LifeEoH}_{im} = \begin{cases} \text{Life}_i & \text{if } m + \text{Life}_i - 1 \leq m_{max} \\ m_{max} - m + 1 & \text{otherwise} \end{cases} \quad \forall i \in I, m \in M \quad (9)$$

Even though the use of discounting, salvage values or annuity factors can help in mitigating end effects, they do not eliminate them. ESOM literature is scarce in evaluating the impacts of these formulations on end effects, but the few studies that exist confirm that end effects are still present even when both discounting and salvage values are used [40,51], and highlight that annuity factors are likely to suffer from similar issues [52]. It is also important to keep in mind that investment lifetimes are just one aspect of end effects. Generally, the more complex a model’s long-term formulation is (e.g., decreasing technology efficiencies over time, long-term storage, constraints endogenously affecting technology costs or carbon budget allocation, etc.), the more likely it is that modellers will need to extend the model’s foresight beyond the target year of interest in order to avoid distorted results. Although longer horizons can introduce additional uncertainty into modelling exercises, it should be noted that end effects are primarily driven by structural limitations in the model. Keeping parameters constant beyond the target year of interest can generally reduce distortions without increasing parametric uncertainty beyond the target year [40,54].

2.3. Spatial and temporal aggregation

Using an ESOM is often a balancing act between the computational power at hand and the model’s resolution, and modellers often turn to heuristics or clustering techniques to improve tractability. It is common to see models aggregating investment decisions to only a few milestone years (M) within the full horizon under study (Y), and a reduced subset

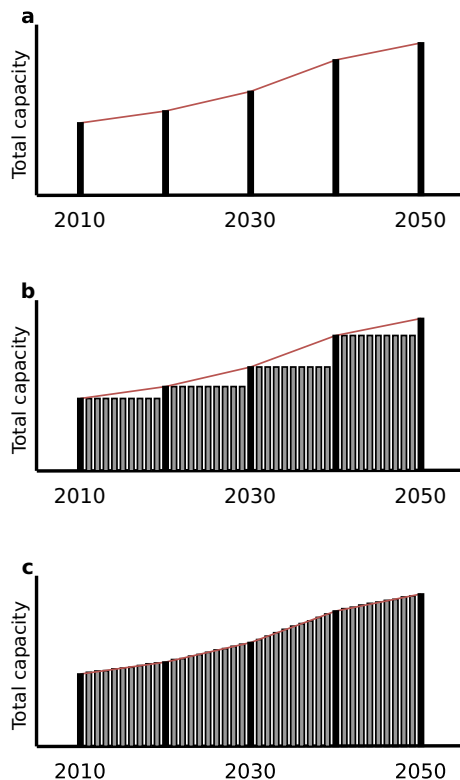


Fig. 5. Illustrative example of different milestone interpolation methods seen in the literature. Black bars represent model decisions, grey bars represent interpolated estimations, and the red line is a linear interpolation case (for comparison). (a) No interpolation [17]. (b) Constant interpolation [64]. (c) Linear interpolation [69].

of timesteps (T) within the hours of each year (H) to keep an exercise tractable [26,64], as shown in Eqs. (10) and (11):

$$m \in M \subseteq Y \quad (10)$$

$$t \in T \subseteq H \quad (11)$$

Doing so comes at a cost in accuracy. The literature has long argued that overly coarse models distort the contribution of renewable technologies [13], yet aggregation to three timesteps per season (day/night/peak, roughly 16 timesteps) or to one day per month (288 timesteps) remains a popular approach [30]. Accordingly, operational time series aggregation has received substantial attention, with studies highlighting how distortions induced by overly coarse time-series data can be significant [26,65]. Studies exploring spatial aggregation have reached similar conclusions when it comes to investments in renewable technologies and transmission networks [66,67], and have highlighted that higher spatial resolution is always preferred if one wishes to diminish distorted results in relation to costs, siting accuracy, and emission trends [68].

The aggregation of milestone years in pathway models has received much less attention, however. The need for it is simple: every year with new investments within the horizon of a model necessitates recalculation of operational decisions, which can quickly make a model intractable. Similar to other dimensions, aggregating investments implies sacrifices to a model's precision. Accounting for artefacts induced by year aggregation is not simple, and has led to models like TIMES reformulating their approach in order to reduce distortions [69].

The methods used by models to interpolate between milestones are diverse (Fig. 5). Some models, such as OSeMOSYS [17], forgo any kind of interpolation and only calculate costs for the requested milestones

(Fig. 5(a)). Others such as TEMOA and PyPSA [60,64], assume capacities and generation remain constant between milestones (Fig. 5(b)). Finally, models like TIMES and EFOM possess involved formulations that attempt to distribute new investments equally between milestones [70], with TIMES even offering multiple distinct approaches [69] (Fig. 5(c)).

Milestone aggregation has important implications for the cost estimations of a model and may introduce additional distortions to its results, as it limits the capacity of a model to adequately assess decommissioning and similar investment dynamics [61]. This issue, which we call *inter-milestone effects*, is rarely described in ESOM literature. In particular, it limits the capacity of a model to adequately assess decommissioning [61]. If unaccounted for, a model may arbitrarily benefit certain investments with an erroneous extended lifetime if the length between milestones (ML_m) and the remaining lifetime of an asset ($LifeRem_{im}$) are misaligned (Fig. 3, Eq. (12)).

$$LifeExtErr_{im} = (-LifeRem_{im}) \bmod ML_m \forall i \in I, m \in M \quad (12)$$

Corrections to these distortions vary per modelling framework and interpolation approach. For example, both TEMOA and MESSAGE scale down the distorted capacity by a factor of its remaining lifetime and the length of the distorted milestone: if the milestone length is 10 and the remaining lifetime of an investment is 7, its remaining capacity will be scaled down by 0.7 for that milestone [18,64]. Although this does not remove the distortion—the asset's lifetime is still artificially extended—it changes its nature in an attempt to reduce its influence.

2.4. Investment dynamics

When evaluating cost-optimised pathways, one should keep in mind their general tendency to delay investments when exogenous performance improvements can be anticipated [71]. In essence, if a model with enough foresight identifies that cost decrements or efficiency improvements occur due to external factors, it may delay deployment until said improvements materialise without representing the mechanisms that drive technological improvement in reality [36,43]. We can call this *opportunistic investment behaviour*.

As a way to counteract opportunistic behaviour, modelling exercises often incorporate constraints affecting investment trends with the aim to guide solutions towards what the modellers see as preferable or more realistic. Examples include, first, constraints on deployment growth [37]; in other words, the amount of new capacity of a given technology that is considered realistic to deploy between two investment milestones is arbitrarily fixed. Second, knowledge stock effects [43], which consist of non-monetary expertise a given nation acquires or loses in relation to a technology, enabling a model to endogenously alter deployment growth based on recent trends. Third, endogenous learning [72], namely a mathematical relation between the amount of capacity deployment of a given technology and its cost, which reduces the more the technology becomes deployed, mimicking real-world technological learning and cost reductions [73]. Fourth and finally, endogenous early retirement [74]; where investments may be retired before the end of their technical lifetimes when doing so is financially attractive, allowing newer technologies to diffuse earlier. Their use varies across modelling frameworks and studies. For example, TIMES offers both endogenous learning and early retirement as optional features that depend on user configuration [16]; similarly, ESO-X, MESSAGE, and PyPSA have been employed both with and without endogenous learning as a feature [36,60,72].

3. Methods

Building and expanding on the key methodological concepts outlined in Section 2, we conduct a systematic review of energy systems modelling literature following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol [75], with the aim of quantitatively assessing how pathway modelling studies published across multiple decades approach such methodological dilemmas. Our protocol-based analysis also offers a structured and replicable approach

Table 1
Terms used to query article titles, abstracts and keywords in the Scopus database [76].

Aspect	Operator	Query
System	–	(energy OR power OR electric*) PRE/2 system
Model	AND	(energy OR power OR electri* OR planning) PRE/2 model*
Approach	AND	optim* OR “bottom up” OR “cost competitive*”
Exercise	AND	pathway OR “multi* period” OR “multi* stage” OR “long term” OR scenario OR “transition path” OR “generation expansion” OR “capacity expansion”

to reviewing this field of research, which may be easily and consistently updated by others in the future. We build upon other reviews of energy systems modelling literature that take quantitative approaches [29, 34,41] by narrowing the subject to a particular aspect of ESOMs (i.e., pathway exercises) while ensuring our inclusion criteria capture a broad range of studies. The following sections detail our procedure: first, how our sample of studies was identified and filtered, and second, the data collection processes.

3.1. Gathering the sample of studies

Data gathering started by developing an appropriate search string to capture terminology often used to describe pathway studies using ESOMs. As seen in Table 1, we combined queries related to four particular aspects of interest: (i) the systems under study, which were energy or power systems; (ii) the models used, namely power and energy planning models; (iii) the modelling approach, which was optimisation via mathematical programming; (iv) the type of exercise, in this case pathways or multi-stage investments. When relevant, a proximity operator—*PRE/n*—was used to capture cases where terminology might vary slightly due to sectoral scope or other particularities (e.g., *energy PRE/2 model** will identify studies using *energy systems model*, *energy modelling* and *energy-economy-environment models* as valid). We purposefully avoided restrictions on the publication date of papers, subject area, or journal ranking in order to diminish biases.

Fig. 6 depicts each stage in our study identification procedure. The Scopus database [76] was used as the primary source of data, resulting in a total of 3632 records after being consulted on the 6th of January 2025. This sample was reduced to 2591 records by filtering out publications that were not written in English or that were not scientific articles or reviews. Although pathways are often featured in grey literature and government reports, both potentially valuable sources of information, the focus of this review is on assessing practices and methods in peer-reviewed scientific literature, meaning that such documents were outside the scope of this study.

To aid in the screening of this large sample of studies, we used ASReview LAB version 1.6.6 [77], a machine-learning tool that employs active learning to assist systematic reviews and meta-analyses in an efficient and transparent manner [78]. The tool has a human-in-the-loop approach, meaning reviewers are ultimately the ones guiding record selection by reading article titles and abstracts; the tool re-arranges the order in which record titles and abstracts are presented to the reviewer to expedite the review process. During this stage, we selected studies mentioning the use of ESOMs in the context of pathways or capacity expansion. If the title and abstract of a study omitted such information, the main body of the text was briefly consulted to assess the study’s relevance. To reduce the risk of the ASReview presets biasing our selection, we split the screening into three phases, with the first two using different classifiers and feature extractors, and the third being an additional validation phase where random records were assessed to confirm the effectiveness of ASReview (see Fig. B.2). Our screening yielded a total of 855 records, which we further reduced to 825 by removing 30 records that were inaccessible. To balance the study’s goals, timeliness, and the

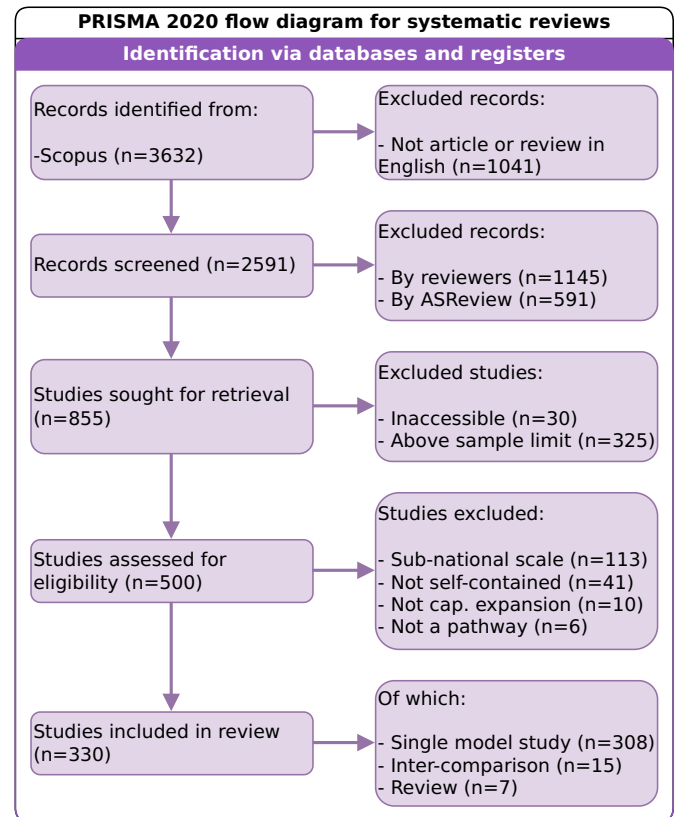


Fig. 6. PRISMA diagram depicting each stage in the systematic review, including the number of studies and reasons for elimination. Adapted from Page et al. [75].

resources at our disposal (i.e., two coders), we established a cutoff of 500 studies for further assessment, for which we selected the studies with the highest ASReview priority scores.

3.2. Assessment and data collection

We developed a coded questionnaire, depicted in Table C.1, which was given to two coders to extract condensed data from each study. Coders were required to mark the sentence or paragraph that led them to provide a specific answer. Inter-coder reliability was assessed and is reported in the results and in Appendix F. We subdivided the questionnaire into three stages, with studies that did not meet certain criteria being excluded from further analysis: (i) an eligibility stage aimed at improving the homogeneity of our sample; (ii) a classification stage aimed at separating single-model studies from reviews and model inter-comparisons; and (iii) an analysis stage that extracted technical information from single-model studies.

The eligibility stage and the classification stage were utilised to improve the comparability of the studies in our sample. To do so, we required studies to have a national or super-national scope, removing models depicting subnational entities like states [79] or cities [80]. We similarly removed studies that either did not use an ESOM or cases where its influence was difficult to establish in a self-contained way. This meant removing bi-directionally coupled models solved in iterations [81] and multi-model setups [82], since establishing cause-and-effect in these setups is difficult and complicates comparisons. Models whose aim was tangential to energy planning, such as car stock models [83], were also removed. Finally, we ensured that studies that did not feature pathways were withdrawn (e.g., studies optimising one or several “snapshot” years [84]). The remaining sample of studies was classified into three categories: studies utilising a single ESOM, literature reviews, and model inter-comparisons.

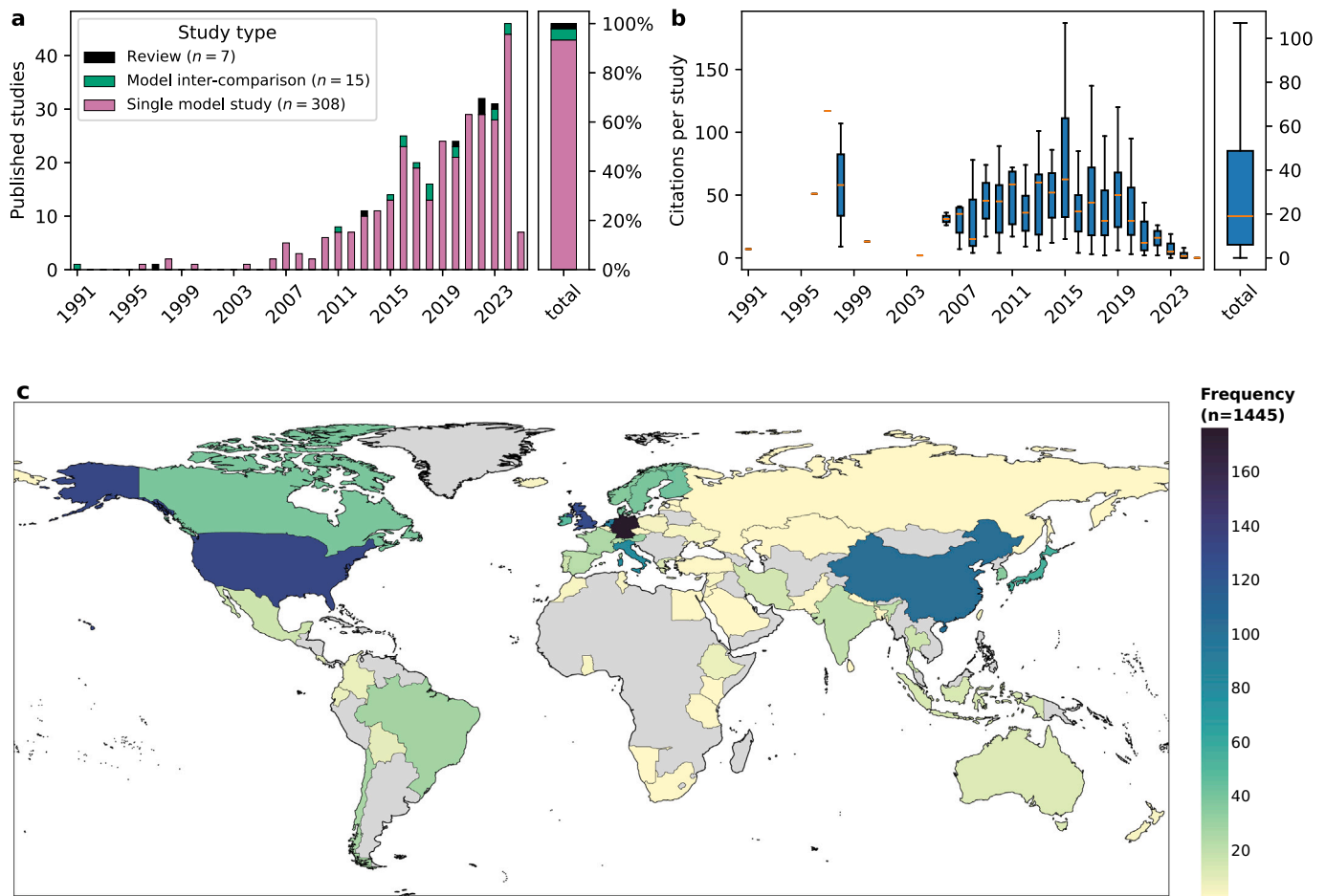


Fig. 7. Bibliographic trends in our full sample of studies ($n = 330$). (a) Study type by year of publication. (b) Citations per study by year of publication. Boxes show the median and interquartile range (25–75%), whiskers extend to 1.5 times the interquartile range, and outliers are omitted. (c) Author affiliation by country, counting each co-author per study separately. Countries in grey had no authorship in our sample.

Our analysis stage focused on extracting data from our sample of single ESOM studies by consulting their methodological descriptions, coding each item in our questionnaire across four topic areas (Table C.1) based on how it was described in the paper. First, we classified studies qualitatively on multiple details, including the modelling framework used, and the type of formulation employed (e.g., linear programming (LP), mixed-integer linear programming (MILP), etc.). Second, we extracted spatial details, including the region under study (see Table D.2) and the model's spatial resolution ($r \in R$). Third, the temporal details of each study were identified, including: the foresight utilised, the horizon of the exercise from baseline to target year—noting modelling extensions beyond said target year if stated—the total number of milestones modelled ($|M|$), and the total number of timesteps ($|T|$). For milestones, we assumed that the baseline year was used only as a reference and was not included in the optimisations (e.g., if a study stated “modelled from 2015 to 2050 in 5-year steps”, then $|M| = (2050-2015)/5 = 7$). For timesteps, we noted the cardinality implied by the aggregation method used in the paper (e.g., statements like “12 representative days at hourly resolution” were coded as $|T| = 12 \cdot 24 = 288$, “full hourly annual resolution” was coded as $|T| = 365 \cdot 24 = 8760$, and “three time bins per season” was coded as $|T| = 3 \cdot 4 = 12$). Fourth, and last, we made note of the use of endogenous learning.

4. Results

4.1. Descriptive analysis

Our systematic review identified 330 publications: 308 single-ESOM studies, 7 review articles, and 15 model inter-comparisons. The publication years range from 1991 to 2025, with a persistent upward trend in publications over time (Fig. 7(a)). Studies published before 2000 are scarce, possibly due to terminology not being captured by our search strategy. After 2000, the number of studies increases from 2 in 2006 to 46 in 2024, reflecting the growing scientific and policy interest in energy system decarbonisation. Our analysis showed a relatively steady mean citation count of 50.3 for studies between 2006–2020, with a wide spread of ± 50.1 SD (Fig. 7(b)). Citations decline from 2021 onwards, which is to be expected due to recency. It should be noted that the search was conducted in January 2025, explaining the low number of studies in that year ($n = 7$).

Studies in our sample are distributed across 93 journal sources, mostly focused on disciplines such as energy research, policy, power systems, and sustainability. Among these, thirteen journals account for 67% of the sample with *Applied Energy* ($n = 40$) and *Energy* ($n = 35$) holding the largest shares and exhibiting a relatively steady publication history (Fig. E.3). Conversely, *Energy Policy* ($n = 30$) was the

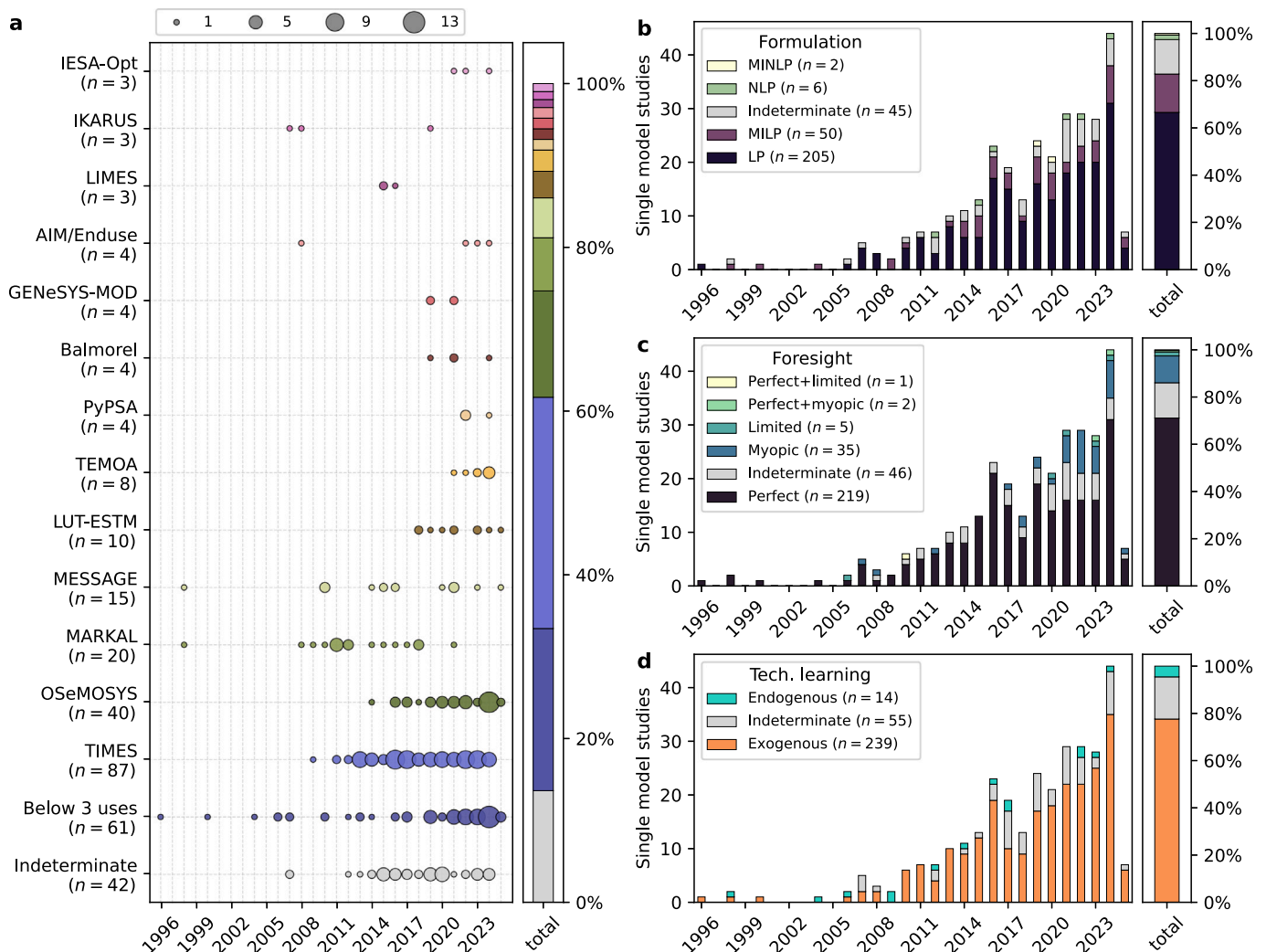


Fig. 8. Summary of model characteristics in single-model studies by year of publication ($n = 308$). (a) Model framework used. Frameworks with lower usage ($n < 3$) have been aggregated into the “Below 3 uses” category. (b) Model formulation as identified by coders. (c) Foresight approach as identified by coders. Cases where authors contrasted multiple foresight formulations have been given their own categories. (d) Technological learning approach as identified by coders.

third largest and showed a decreasing trend in publications within our sample in the latest half-decade. A total of 59 journals held a single article within our sample, and were composed mostly focused journals of on specialised technical topics (e.g., geothermal engineering, solar photovoltaics, buildings, etc.) or a dedicated regional scope (e.g., Türkiye, Canada, South Africa). Neither of these findings is surprising given the wide breadth of technologies included in ESOMs and our eligibility criteria.

We similarly assessed the distribution of author affiliation within our sample, which showed that a reduced number of countries hold significant influence over this field of literature (Fig. 7(c)). Of the total of 1445 identified co-author counts, ten nations accounted for 62% in terms of author affiliation, with Germany (12.2%), the U.S. (9.1%), the U.K. (8.9%), China (7.1%) and the Netherlands (6.3%) being the largest among them. Co-authors with European affiliations are over-represented in studies, with a total share of 58.1% of the total ($n = 839$), followed by Asia with 19.6% ($n = 283$), and North America with 13.6% ($n = 197$). In contrast, Oceania, Africa and South America are under-represented with 1.1%, 2.1%, and 5.5% respectively (see Tables E.3 and E.4 for details).

4.2. General qualitative features of ESOMs

We evaluate single-model studies ($n = 308$) on multiple qualitative aspects including model framework and formulation, foresight, technology learning approach and modeller region. This section reports substantive characteristics only; reporting completeness (i.e., indeterminate items) is analysed in Section 4.3.

Modelling frameworks. We identified 63 named modelling frameworks, 14 of which were used more than three times (Fig. 8(a)). TIMES was the most used framework at 28% ($n = 87$), which expands to 34% when combined with its predecessor MARKAL ($n = 20$) [85]. OSeMOSYS was the second most used framework at 13% ($n = 40$), which increases to 14% when accounting for GENE SYS-MOD ($n = 4$), which was based on the same codebase [86]. Other frameworks included MESSAGE ($n = 15$), LUT-ESTM ($n = 10$) and TEMOA ($n = 8$). Notably, the number of frameworks below 3 uses increased sharply in the last half-decade (20%, $n = 20$), indicating growing diversification in the field.

Mathematical formulation. This category displayed a broadly consistent preference towards tractable formulations. Linear programming (LP) dominated (67%, $n = 205$), followed by MILP (16%, $n = 50$). Non-linear approaches were rare (2.5%, $n = 8$), which is to be expected as it implies

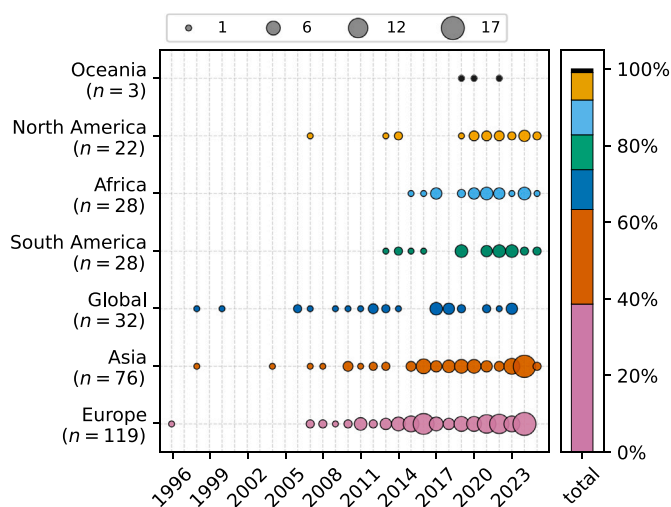


Fig. 9. World regions modelled in single-model studies by year of publication ($n = 308$). See Table D.2 for further details on country assignment.

higher computational costs [35]. There is no strong evidence of complex formulations becoming more popular, with LP remaining the preferred approach across the years.

Foresight. Perfect foresight was the most used approach (71%, $n = 219$), with myopic rolling horizon being a distant second (11%, $n = 35$), and limited-foresight exercises a scarce third (2%, $n = 5$). Myopic exercises have become more common in the past half decade, with 69% ($n = 22$) of such studies appearing between 2020 and 2025. Only a few studies compared foresight approaches directly ($n = 3$) [37,87,88]. Overall, foresight choice is quite binary between perfect and myopic foresight, with limited approaches remaining rare occurrences.

Endogenous learning. Endogenous learning was uncommon (5%, $n = 14$), implying that most models have a weakened causal link between technology deployment and cost reductions from economies of scale. This type of setup is known to lead to delayed system change and a potential for the model to exhibit opportunistic behaviour [71]. The trend is unsurprising, as endogenous learning increases computational complexity, either by introducing non-linear relations or necessitating iterative model runs. Despite these challenges, our sample included cases where modellers employed more involved approaches—such as Benders decomposition—to enable endogenous learning in models with high operational resolution [89]. Such combinations of decomposition techniques and endogenous learning have been highlighted by other studies as promising avenues for improvement [41].

Modelled region. Fig. 9 shows how the regional focus of these studies generally reflects the authorship trends identified in Section 4.1, with Europe (39%, $n = 119$) and Asia (25%, $n = 76$) being the main areas of focus, followed by global models as the third largest category (10%, $n = 32$). North America (7%, $n = 22$) had a lower share than its co-authorship counterpart, which may be explained as an artefact of our decision to exclude sub-national studies, which therefore excludes studies of sub-national U.S. grids such as ERCOT [90]. Of the remaining regions, South America (9%, $n = 28$), and Africa (9%, $n = 28$) had larger shares than their co-authorship, and Oceania had a roughly equal amount (1%, $n = 3$).

4.3. Degree of transparency

Our questionnaire assessed reporting clarity for 12 items, allowing coders to mark an item as *indeterminate* when a study did not facilitate

Table 2

Summary of indeterminate cases per question for single model exercises ($n = 308$).

Subject	Question	Total (%)
Model specification	Modelling framework	42 (13.6%)
	Model name	163 (52.9%)
	Formulation	45 (14.6%)
Spatial detail	World region	0 (0.0%)
	Number of regions	46 (14.9%)
Temporal detail	Foresight	46 (14.9%)
	Baseline year	8 (2.6%)
	Target year	6 (1.9%)
	Horizon year	8 (2.6%)
	Number of milestones	12 (3.9%)
Investment dynamics	Number of timesteps	85 (27.6%)
	Learning approach	55 (17.9%)

an unambiguous entry. Table 2 summarises the indeterminate shares by item.

Studies did not assign a name to the framework utilised in 14% of cases ($n = 42$). This is not necessarily problematic as models can be tailor-made for a specific study (making naming irrelevant), authors may prefer not to assign a name, or journal policies may discourage naming altogether. However, it implies that at least 86% of studies used or developed a framework intended for reuse, assuming that naming signals that intention.

Geographic scope was always identifiable—by design, as ambiguous regions were excluded during screening. Items related to a model’s horizon (baseline year, target year, horizon year, number of milestones) were coded in most cases, each with <4% indeterminate.

Items central to determining computational tractability showed more opacity. Mathematical formulation (e.g., LP, MILP) and foresight were indeterminate in 15% of studies, which is compounded by additional opacity regarding the number of regions modelled (15%) and the number of operational timesteps (27.2%). These omissions hinder the ability of readers to interpret both operational detail and the computational burden of models.

Endogenous learning was indeterminate in 18% of studies, reflecting difficulties in confirming its absence. Despite previous reviews highlighting its importance [42], it is not a widely used method, likely due to computational burden [41]. Most studies implicitly assumed exogenous learning, increasing the burden of proof for systematic exercises like this one.

The numbers in Table 2 also mask other difficulties that were hard to quantify. Studies frequently left key model aspects open to interpretation, and it was not uncommon for authors to cite a framework without specifying configuration or resolution, even though most frameworks allow multiple mathematical formulations, temporal/spatial resolutions, and optional constraints. In other cases, readers were referred to multiple model versions with differing configurations, or to grey literature that was effectively inaccessible due to link rot or access restrictions. Inter-coder analyses showed a similar pattern: model formulation, foresight and learning approach produced disagreement, which completely disappeared when indeterminate cases were excluded (Table F.5). That is, coders disagreed about clarity rather than the feature itself when it was disclosed. Quantitative questions showed comparable issues, with the number of model milestones proving particularly challenging and often leading coders to report different counts when the long-term resolution was vaguely stated.

4.4. Quantitative analysis of model resolution

Our large sample of studies provides a unique opportunity to assess the evolution of multiple aspects of ESOMs across the years. To do so, we isolate a subset of studies that provide sufficient data for the following items: model foresight, number of milestones ($|M|$), number of regions

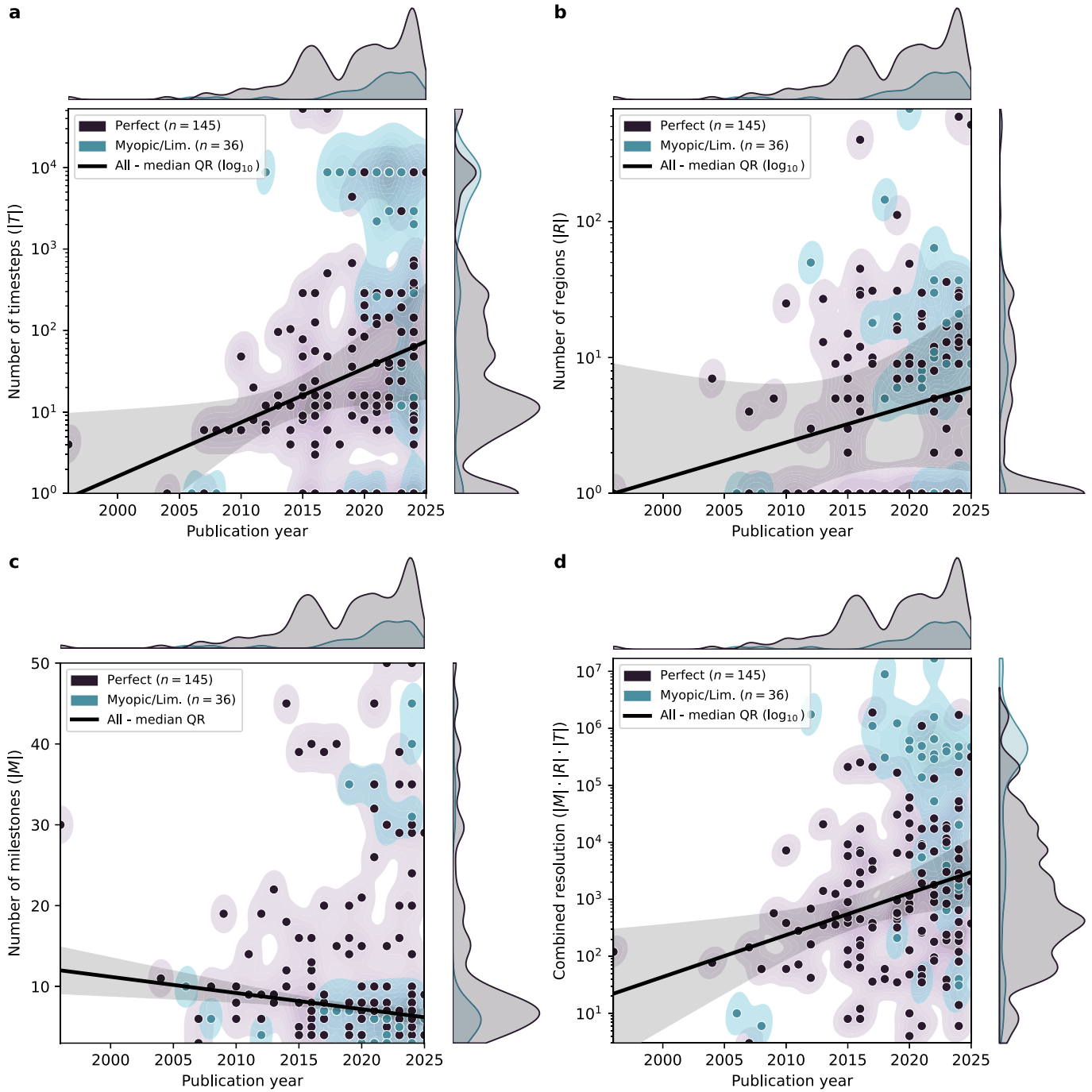


Fig. 10. Dimensional characteristics of single-model studies with sufficient transparency by publication year and foresight approach ($n = 181$). The solid black line represents a median quantile regression ($\tau = 0.5$) with 95% confidence bands from a paired bootstrap in grey. Side panels show kernel density estimates by publication year (top) and model dimension (right), subdivided by foresight category (colour). (a) Operational resolution (total timesteps $|T|$). (b) Spatial resolution (total regions $|R|$). (c) Long-term resolution (total milestones $|M|$). (d) Evolution of combined resolution ($|M| \cdot |R| \cdot |T|$)—an indicator of study setup, not computational burden. Y-axes use \log_{10} scaling in a, b, and d; linear scale in c.

($|R|$), number of operational timesteps ($|T|$), baseline year, target year, and horizon year. This resulted in a reduced subset of 181 studies (59% of 308). We group studies into those with perfect foresight ($n = 145$) and those using myopic or limited foresight ($n = 36$).

We evaluate trends in the dimensionality of ESOMs in our subset in recent decades (Fig. 10), revealing distinct trends despite the wide heterogeneity of these models. In particular, there is an increasing preference towards higher short-term resolution—in both temporal and spatial aspects—and a decreasing trend in long-term resolution.

Short-term temporal resolution—measured as the number of operational timesteps $|T|$ —shows clear increases over time (Fig. 10a), with an annual multiplicative change of 16.4% (95% CI [7.2%, 31.8%]; $n = 181$), equivalent to doubling roughly every 4.6 years (95% CI [2.5, 10]). Despite this upward trend, dispersion remains large, indicating substantial heterogeneity across modelling studies. Foresight strongly influences operational resolution: studies with myopic or limited foresight tend to favour near-hourly annual operation at a median of 2920 (IQR 31–8760; $n = 36$), with studies using myopic versions of PyPSA and LUT-ESTM

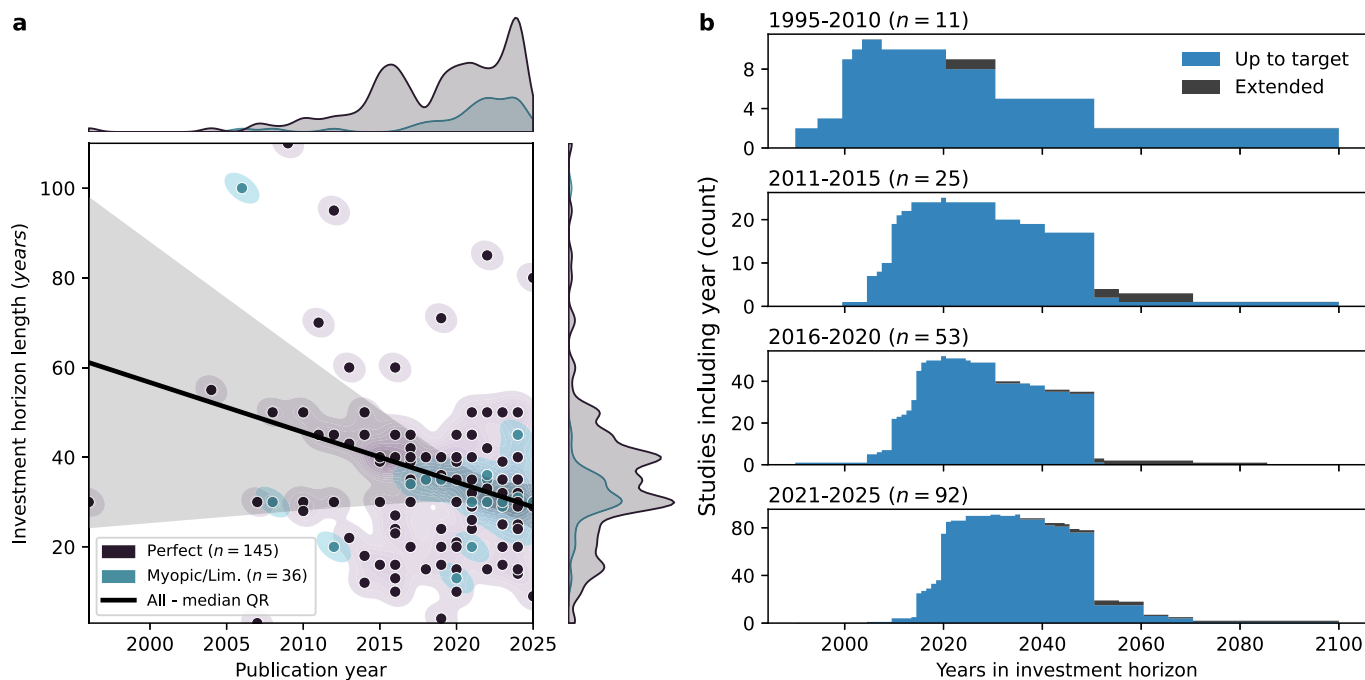


Fig. 11. Evaluation of long-term characteristics of single-model studies with sufficient transparency that relate to end effects ($n = 181$). (a) Length of the modelled investment horizon by year of publication and foresight approach. The solid black line represents a median quantile regression ($\tau = 0.5$) with 95% confidence bands from a paired bootstrap in grey. Side panels show kernel density estimates by publication year (top) and investment horizon length (right), subdivided by foresight category (colour). (b) Counts of each modelled year, subdivided into groups of close publication dates. Cases where authors extended the horizon beyond the target year of interest have been highlighted in black.

often favouring full hourly resolution (Fig. G.5). Conversely, perfect foresight models are more dispersed in their approach and use lower resolutions at a median of 16 timesteps (IQR 6–120; $n = 145$). This variability also holds within frameworks: models built with OSeMOSYS, TEMOA and TIMES vary by over two orders of magnitude in this metric (Fig. G.5).

Spatial resolution—total regions $|R|$ —shows a modest upward tendency over time. The median trend indicates an annual increase of 6.4% (95% CI -3.3 – 14.7% ; $n = 181$). Over a decade, this compounds to about 86% higher on the median. Because the confidence interval includes 0%, the evidence is suggestive rather than definitive (Fig. 10b). There is a clear trade-off between foresight and spatial detail: models either pursue finer spatial resolution at the cost of less or no anticipation in investments, or retain anticipation by neglecting energy transmission costs and weather differences across regions (Fig. G.4). By foresight type, perfect foresight studies tended to use single-region “copperplate” setups (median 1, IQR 1–10; $n = 145$), a pattern that persists even in recent work, whereas myopic and limited foresight models were more spatially disaggregated (median 9, IQR 5.75–23.24; $n = 36$).

Fig. 10c shows that long-term temporal resolution—milestones $|M|$ across the investment horizon—declines over time. The median trend is -0.2 milestones per year (95% CI $[-0.08, -0.27]$; $n = 181$). This corresponds to roughly two fewer milestones per decade assuming milestones are spaced five years apart ($ML = 5$). Aggregate metrics suggest that this is the only dimension in which models with perfect foresight utilise higher resolutions (median 8 milestones, IQR 6–16) than models with myopic or limited foresight (median 6 milestones, IQR 5.75–7), although in general this aspect remains coarse in both approaches (Fig. G.4).

To answer the question of whether or not there is a focus on increasing resolution in these studies as a whole, we look at their combined resolution (the product $|M| \cdot |R| \cdot |T|$) in Fig. 10(d). Importantly, this measure reflects study setup rather than the computational burden of a model, since myopic and limited-foresight models do not resolve all milestones simultaneously. The increases in short-term resolution

($|T|$) dominate, leading to a similar annual increase of 18.4% (95% CI $[10.53\%, 34.59\%]$, which would compound to about a 443% increase on the median over a decade. Dispersion is large (log10-MAE 1.13) due to the substantial heterogeneity across studies and years. This large increase must be put into context, as this category has the most clear outlines between foresight categories (Fig. G.4). Perfect foresight studies had an aggregate median of $10^{2.81}$ (IQR $[10^{2.16}, 10^{3.81}]$) which was about three orders of magnitude below that of myopic or limited foresight models ($10^{5.49}$, IQR $[10^{3.52}, 10^{5.72}]$). Looking at specific frameworks (Fig. G.5) reveals a few outliers, with the IESA-Opt model [91] and the model of Komiyama and Fujii [92] as standout perfect foresight cases with a higher combined resolution than most myopic models solved at an hourly operational resolution.

4.5. The heuristics of optimised pathways

Section 2 described two types of issues that can surface in long-term exercises: (i) end effects, which occur when models with foresight do not extend beyond the year of interest, and (ii) inter-milestone effects, which occur when years in the investment horizon are aggregated into groups (i.e., $ML_m > 1$). Although the presence of said issues broadly depends on the experimental setup of each study and the formulation of each model, we can assess modeller practices that surround them.

As noted in Section 2.2, end effects are mitigated—but not eliminated—by heuristics such as salvage values or heavy discounting; robust handling typically also requires extending the model beyond the target year [51]. To assess practice, we examine two indicators: (i) the length of the investment horizon (total years spanned; Fig. 1), and (ii) whether authors explicitly state an extension beyond the target year.

For the first indicator, Fig. 11(a) suggests a decrease in horizon length over publication years: a linear median regression estimates -1.11 years per calendar year (95% CI $[-1.25, 0.00]$). For the second, we grouped studies with close publication proximity and counted explicit horizon extensions (Fig. 11(b)). Only 15 studies reported an extension (8.2% of

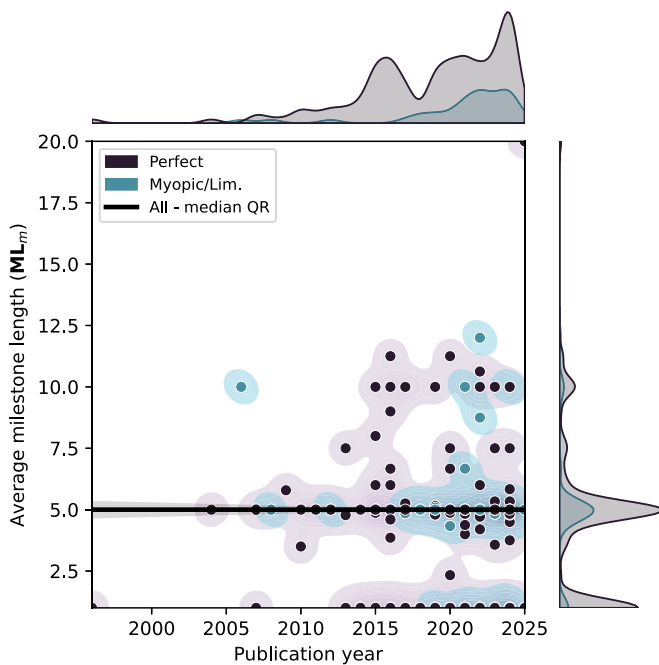


Fig. 12. Long-term aggregation trends in single-model studies with sufficient transparency ($n = 181$) by year of publication and foresight approach. The solid black line represents a median quantile regression ($\tau = 0.5$) with 95% confidence bands from a pairs bootstrap in grey. Side panels show kernel density estimates by publication year (top) and average milestone length (right), subdivided by foresight category (colour).

$n = 181$), including frameworks with distinct approaches for end effect mitigation. These include models with both salvage values (OSeMOSYS $n = 6$, TIMES $n = 3$), and annuity factors (SWITCH $n = 1$, GridPath $n = 1$); see Table H.7 for details.

Taken together, these patterns suggest that explicit extensions are either uncommon or unreported, while planning horizons are likely shortening in the literature. There is a strong tendency to terminate analyses at years prominently featured in policy targets (e.g., 2050, 2100), with 2050 being a standout across categories. Thus, ESOM studies appear increasingly short-term in scope, primarily handling end effects via heuristics (e.g., salvage values) rather than via a conscious re-structuring of the problem such as horizon extensions. These practices can bias results towards near-term solutions near important policy years, favouring technologies with short payback periods and undervaluing long-lived assets or variable renewable technologies [52].

Inter-milestone effects depend on interactions between model structure and exogenous parameters. For example, a model with no or very coarse milestone interpolation (i.e., Fig. 5(a) and (b)) will be ill-suited to evaluate technologies with short lifetimes if used in combination with long milestone lengths. Thorough comparison between model formulations is beyond the scope of this review, as it requires a curated harmonisation between models to enable proper evaluation. Nevertheless, we can assess the need for such comparisons by looking at trends in model setup.

Fig. 12 suggests that time aggregation is typically chosen heuristically. Most studies either optimise every year—often intractable when other dimensions are highly resolved—or group years into multiples of five as a rule of thumb. The latter does not align with the expected decommissioning dynamics of the energy transition [93], and can lead to distorted results depending on the model's structure and application. As an example, the second most popular framework—OSeMOSYS—does not interpolate or represent the period between investment milestones. Its core design goal was to have a straightforward formulation [17],

so the choice to use discrete, non-interpolated periods is unsurprising. Nevertheless, this can lead to biased estimates of cumulative annual trends unless additional assumptions or a more detailed formulation are used to describe the evolution between milestones. If these are omitted or applied inconsistently, the resulting cumulative quantities may be systematically misestimated. This matters for common policy questions such as national carbon budgets [94].

5. Discussion and conclusion

We conducted a systematic review of energy system pathway exercises using ESOMs in the academic literature, focusing on technical aspects of model setup that influence results, but are often left implicit. We highlight four critical aspects: i) the implications of model foresight assumptions, ii) decision horizons associated with end effects that can distort model outcomes, iii) tradeoffs in model resolution, and iv) investment dynamics and the need to counteract opportunistic investment behaviour by incorporating endogenous learning.

From this review, we reveal four broad patterns corresponding to the critical aspects identified above. For i), long-term ESOMs generally fall into two categories: either perfect foresight exercises with strong anticipatory capacity but relatively low spatial and operational resolution, or, increasingly, myopic exercises that offer high spatial and operational detail but struggle to account for long-term policy goals. For ii), modellers typically follow heuristic rules of thumb when choosing the investment horizon in pathway exercises, leading to a strong trend towards shorter model horizons as key carbon-neutrality deadlines draw nearer; in many formulations, this can bias models towards short-term solutions because end effects distort investment choices. For iii), there is a clear trend towards higher operational accuracy, with finer short-term temporal resolution receiving particular attention, reflecting the need to model variable renewable technologies such as solar and wind with greater precision. For iv), the use of endogenous learning remains uncommon, likely because of computational and algorithmic complexity, leaving models with a tendency to delay investments in expectation of future price decreases, which goes against well-understood economies of scale.

This highlights several areas for improvement and further research. The choices of model foresight, combined spatial–operational resolution, and endogenous learning—points i, iii, and iv—are closely interrelated, in large part because of computational limitations, and different configurations imbue models with different biases. The current binary between perfect foresight and full myopia suggests room for improvement, as limited foresight approaches might offer a better trade-off between a model's anticipatory ability and its operational and spatial resolution, reducing the risk of lock in and stranded assets. A few examples within our sample show how methodological improvements can preserve perfect foresight even at high resolutions, either by selectively coarsening non-electrical sectors [95], or by using decomposition techniques to enable endogenous learning at these resolutions [89]. Recent work has already shown that substantial computational gains are possible by making the temporal representation across the model more flexible [96], making this a promising area of study. Additionally, greater awareness is needed of potential model distortions caused by short horizons and coarse long-term aggregation. Although scarce, literature showcasing more rigorous treatment of end effects does exist [52], and established models like TIMES have documented the impacts of improper milestone aggregation [69]. Model developers and users stand to benefit from adopting more systematic approaches to evaluate and contrast these aspects. Finally, there is a pressing need for greater transparency in how the long-term setup of these models is communicated: pathway exercises often leave out contextual details that are essential for interpreting their results, as has also been noted by other work on energy system modelling best practices [42,97].

This work has several limitations. Our search terms and article selection criteria were designed to capture and harmonise studies in order to

identify modeller practices related to pathway setup, not to provide evidence about the quality of those studies per se. Other important aspects affecting the quality of ESOM studies, such as operational formulations, approaches to uncertainty mitigation, and up-to-date model parameterization, are outside the scope of this work. Furthermore, models are mutable pieces of software, meaning that the articles and model documentation we use as examples may not reflect the latest versions of those models. The critical aspects and theoretical concepts highlighted here are, however, expected to remain relevant. Similarly, our review does not capture modelling setups utilised in grey literature or government reports, meaning that the assessment of pathway practices at the academic-policy interface remains an important area of future study. Finally, we do not consider interactions between ESOMs and other policy-focused models, particularly IAMs. These heterogeneous models often incorporate ESOMs and face similar tractability trade-offs in foresight, endogenous learning, and resolution [19,98]. We argue that both fields would benefit from quantitative analyses aimed at identifying best practices for model development and communication of limitations. Promising steps in this direction have been made in the context of recent projects which focused explicitly on ESOM-Integrated Assessment Model (IAM) intercomparisons [99,100]. However, these focused primarily on model outputs, and less so on mathematical formulations, for which we encourage further mutual learning.

Our study systematically assessed practices related to optimised energy system pathways in the academic literature, highlighting important considerations that modellers and model users should take into account when interpreting results produced by ESOMs. In particular, we underscore that clear communication of the type of foresight used is essential. Myopic formulations can be useful when assessing the effects of unexpected disruptions on the expected benefits of investments [101], such as policy shifts or technological developments that were foreseen but not achieved. However, they implicitly assume that investments are not made in anticipation or consideration of potential future policies, which limits their ability to capture strategies that avoid lock-in and stranded assets under tightening environmental targets. Advice about planning pathways coming from such models should be interpreted with this nuance in mind.

Similarly, careful consideration of biases in model formulation is necessary to avoid misinterpreting the effectiveness of the proposed pathways. This is particularly important when discussing technologies known to artificially benefit from end of the horizon effects (e.g., gas turbines, carbon-capture and storage) [40,52], or when representing technologies that might be distorted by milestone interpolation due to uneven year-by-year stock turnover (e.g., road transport, distributed power generation) [102,103]. Despite advancements in the operational resolution of models and a few promising highly resolved cases, the computational challenge facing ESOMs remains substantial due to the need for horizon extensions, the widespread coarseness in long-term resolution, and the additional complexity induced by endogenous learning formulations.

CRedit authorship contribution statement

Ivan Ruiz Manuel: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Meijun Chen:** Writing – review & editing, Validation, Investigation, Data curation. **Francesco Lombardi:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Stefan Pfenninger-Lee:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

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.

Acknowledgements

This work was sponsored by the Swiss Federal Office of Energy's "SWEET" call 1–2020 programme and performed in the PATHFNDR (PATHways to an Efficient Future Energy System through Flexibility AND Sector Coupling) consortium.

Appendix A. Discount rate dynamics

Here we present a streamlined example of discount rates reducing decision horizons based on [49]. There are two mutually exclusive choices: C_1 , which steadily rises over time, and C_2 , which follows the same steady rise but alternates around C_1 by a fixed amplitude (Fig. A.1a). Because the model adds up costs only up to a chosen end year (the horizon), and because higher discount rates put less weight on later years, the preferred option can flip depending on where one stops and how strongly one discounts (Fig. A.1b). A discount rate of 0 makes the model alternate between choices indefinitely (Fig. A.1c), meaning that the decision horizon is effectively infinite. A positive discount (3%) stabilises the choice if the horizon is sufficiently long at around ~ 2027 , suggesting a decision horizon of 17 years. Conversely, a negative discount (-1%) exacerbates the alternation between choices the longer the horizon is.

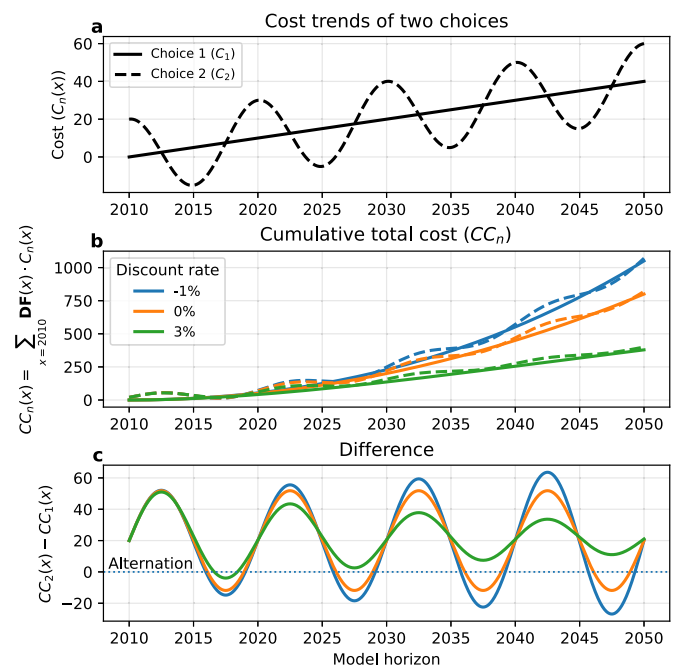
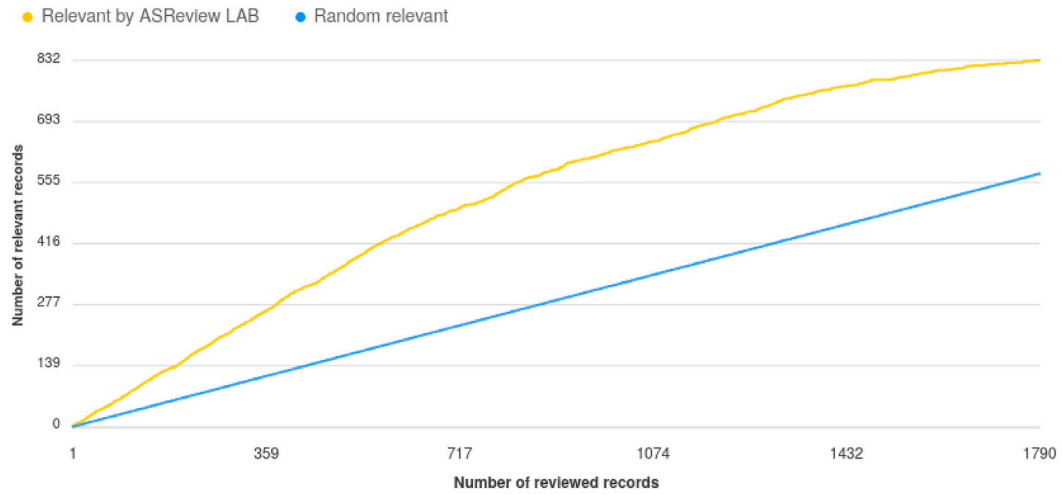


Fig. A.1. Example of discount rates influencing the length of a model's decision horizon, Adapted from [49]. (a) Cost trends of two mutually exclusive choices. (b) Cumulative discounted cost of both choices over time. (c) Difference between cumulative costs over time.

Appendix B. Review screening phase

Screening was subdivided into three steps: (i) an initial pass using the default Naive Bayes classifier with a Term Frequency-Inverse Document Frequency (TD-IDF) feature extractor, with a stopping condition of 66% of total papers screened; (ii) a second pass using a more computationally expensive two layer neural network with Doc2Vec as the feature extractor [104], with a stopping condition of 75% of papers screened; (iii) a final pass where a sample of 50 random leftover records was screened to assess the effectiveness of the active learning tool. The progress at each stage of the screening phase is depicted in Fig. B.2. It can be seen that, at the end of stage (i), the rate of relevant studies significantly decreases in relation to those analysed meaning that further screening would be mostly immaterial, which was confirmed during stage (ii). None of the



(a)

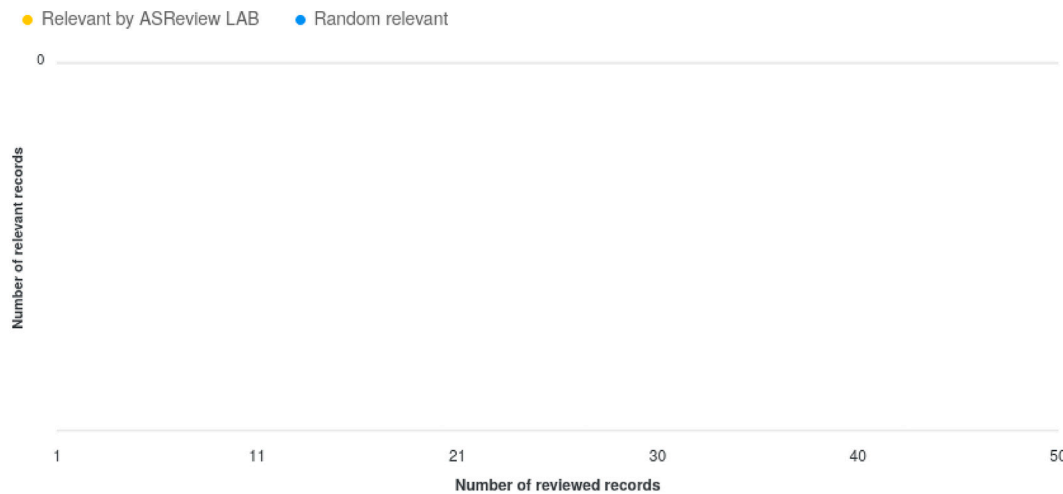
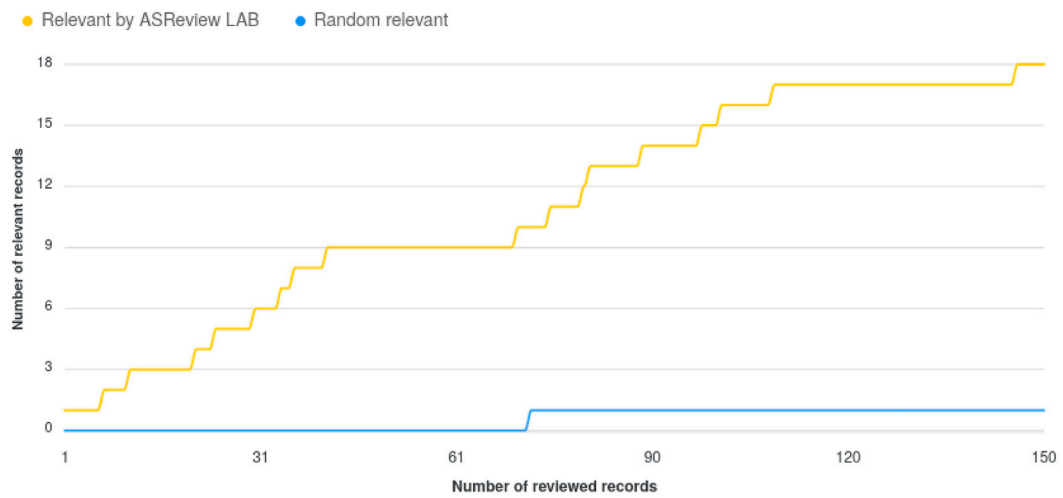


Fig. B.2. Screening phases: relevant records (yellow) versus random relevant (blue). Figures produced using ASReview LAB version 1.6.6 [77]. **Screening phase i** using Naive Bayes as a classifier and a TD-IDF feature extractor. This phase ended after screening 66% of the sample. **Screening phase ii** using a two-layer neural network as a classifier and a Doc2Vec feature extractor. This phase ended after screening 75% of the sample. **Screening phase iii** using a random sample of 50 leftover studies. No relevant records were found.

random records evaluated at stage (iii) was deemed relevant, which led us to conclude that the screening tool was an effective aid.

Appendix C. Assessment questionnaire

Table C.1

Questionnaire filled for each study in our sample. Questions had a multiple-choice setup with two exceptions: (-) indicates an answer was open text, and (*) indicates an answer could be any natural number.

Stage	Subject	Question	Options
Eligibility		Is the study relevant at or above national scale?	Yes / No / Unclear
		Is the study relevant to at least one self-contained ESOM?	Yes / No / Unclear
		Does the study relate to energy generation capacity expansion planning?	Yes / No / Unclear
		Does the study relate to pathways?	Yes / No / Unclear
Classification		What type of study is this?	Single model study / Review / Model inter-comparison / Unclear
Analysis	Model detail	What is the name of the modelling framework used?	- / Unclear
		What name did the authors use for this specific model?	- / Unclear
		What type of formulation was used?	LP / MILP / NLP / MINLP / Other / Unclear
Spatial detail		What region of the world was modelled?	Global / Africa / Asia / Europe / North America / Oceania / South America / Synthetic / Unclear
		How many regions does the model have?	* / Unclear
Temporal detail		Which types of foresight were used?	Perfect / Limited / Myopic / Unclear
		What is the baseline year of the model?	* / Unclear
		What is the target year of the exercise?	* / Unclear
		What was the final year modelled?	* / Unclear
		How many investment decision years were used?	* / Unclear
Investment dynamics		How many operational timesteps were used?	* / Unclear
		What type of technological learning was used?	Exogenous / Endogenous / Unclear

Appendix D. Regions used for spatial classification

Table D.2

World region classification used in this study as defined in our world in data [105]. Please note that a country being in this list does not imply it was individually featured within in our sample of single-model studies.

Region	Countries
Africa	Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo, Cote d'Ivoire, Democratic Republic of Congo, Djibouti, Egypt, Equatorial Guinea, Eritrea, Eswatini, Ethiopia, Ethiopia (former), Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mayotte, Morocco, Mozambique, Namibia, Niger, Nigeria, Orange Free State, Reunion, Rwanda, Saint Helena, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, South Sudan, Sudan, Sudan (former), Tanzania, Togo, Tunisia, Uganda, Western Sahara, Zambia, Zimbabwe
Asia	Afghanistan, Armenia, Azerbaijan, Bahrain, Bangladesh, Bhutan, British Indian Ocean Territory, Brunei, Cambodia, China, Christmas Island, Cocos Islands, Democratic Republic of Vietnam, East Timor, Georgia, Hong Kong, India, Indonesia, Iran, Iraq, Israel, Japan, Jordan, Kazakhstan, Korea (former), Kuwait, Kyrgyzstan, Laos, Lebanon, Macao, Malaysia, Maldives, Mongolia, Myanmar, Nepal, North Korea, Oman, Pakistan, Pakistan (former), Palestine, Philippines, Qatar, Republic of Vietnam, Saudi Arabia, Singapore, South Korea, Sri Lanka, Syria, Taiwan, Tajikistan, Thailand, Turkey, Turkmenistan, United Arab Emirates, Uzbekistan, Vietnam, Yemen, Yemen Arab Republic, Yemen People's Republic
Europe	Aland Islands, Albania, Andorra, Austria, Austria-Hungary, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czechia, Czechoslovakia, Denmark, Duchy of Modena and Reggio, Duchy of Parma and Piacenza, East Germany, Estonia, Faroe Islands, Finland, France, Germany, Gibraltar, Grand Duchy of Baden, Grand Duchy of Tuscany, Greece, Guernsey, Hungary, Iceland, Ireland, Isle of Man, Italy, Jersey, Kingdom of Bavaria, Kingdom of Sardinia, Kingdom of Saxony, Kingdom of Wurttemberg, Kingdom of the Two Sicilies, Kosovo, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Moldova, Monaco, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Russia, San Marino, Serbia, Serbia and Montenegro, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkmenistan, United Arab Emirates, Vatican, West Germany, Yugoslavia
North America	Anguilla, Antigua and Barbuda, Aruba, Bahamas, Barbados, Belize, Bermuda, Bonaire Sint Eustatius and Saba, British Virgin Islands, Canada, Cayman Islands, Costa Rica, Cuba, Curacao, Dominica, Dominican Republic, El Salvador, Federal Republic of Central America, Greenland, Grenada, Guadeloupe, Guatemala, Haiti, Honduras, Jamaica, Martinique, Mexico, Montserrat, Netherlands Antilles, Nicaragua, Panama, Puerto Rico, Saint Barthelemy, Saint Kitts and Nevis, Saint Lucia, Saint Martin (French part), Saint Pierre and Miquelon, Saint Vincent and the Grenadines, Sint Maarten (Dutch part), Trinidad and Tobago, Turks and Caicos Islands, United States, United States Virgin Islands
Oceania	American Samoa, Australia, Cook Islands, Fiji, French Polynesia, Guam, Kiribati, Marshall Islands, Micronesia (country), Nauru, New Caledonia, New Zealand, Niue, Norfolk Island, Northern Mariana Islands, Palau, Papua New Guinea, Pitcairn, Samoa, Solomon Islands, Tokelau, Tonga, Tuvalu, Vanuatu, Wallis and Futuna
South America	Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Falkland Islands, French Guiana, Great Colombia, Guyana, Paraguay, Peru, South Georgia and the South Sandwich Islands, Suriname, Uruguay, Venezuela

Appendix E. Bibliometric analyses

Additional bibliometric analyses were conducted using the pybibx library to analyse our sample [106]. To do so, bibliometric data for each study were downloaded from the SCOPUS database [76] and subsequently fed to the pybibx library [106] for further processing and cleaning. In the case of institutional affiliation analyses (Tables E.3 and Table E.4) a total of five cases—0.3% of the sample—could not be assigned to a country, meaning the total co-authorship sample including unknowns was 1450, meaning missing affiliation cases are immaterial to our results.

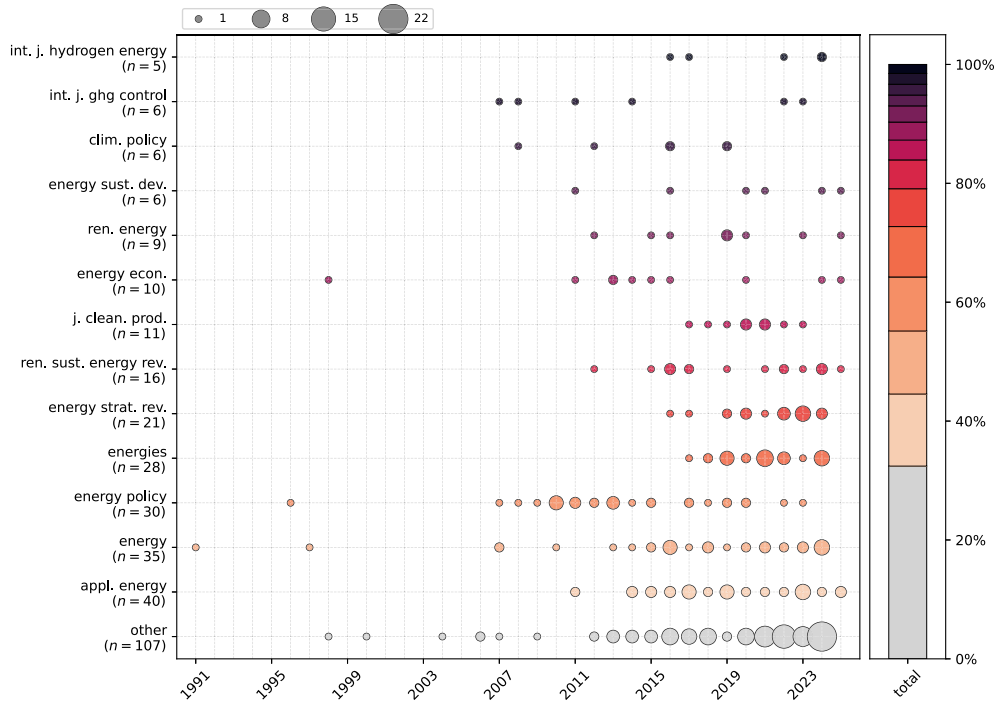


Fig. E.3. Publishing trends in our full sample of studies ($n = 330$). Size depicts study count per year. Journals with less than 5 total publications were aggregated into “other”.

Table E.3

Co-authorship share by institution for 1445 co-authors identified within our full sample of studies ($n = 330$) for each world region as defined by our world in data [105].

World region	Countries identified	N. of countries	Frequency	Share
Africa	Ethiopia, Ghana, Morocco, Tunisia, Egypt, South Africa, Tanzania, Kenya, Namibia	9	31	2.1%
Asia	China, Japan, South Korea, India, Thailand, Iran, Indonesia, Taiwan, Singapore, Syria, Pakistan, Turkey, Bangladesh, Jordan, Nepal, Kazakhstan, Saudi Arabia, Sri Lanka	18	283	19.6%
Europe	Germany, United Kingdom, Netherlands, Italy, Ireland, Finland, Denmark, Sweden, Norway, Austria, France, Spain, Switzerland, Portugal, Greece, Cyprus, Poland, Belgium, Iceland, Ukraine, Czechia, Faroe Islands, Russia, Estonia, Latvia, Luxembourg	26	839	58.1%
North America	United States, Canada, Mexico, Costa Rica	4	197	13.6%
Oceania	Australia, New Zealand	2	16	1.1%
South America	Brazil, Chile, Bolivia, Colombia, Ecuador	5	79	5.5%

Table E.4
Countries with the largest co-authorship share by institutional affiliation. Computed using pybibx [106].

Country	Frequency	Share ($n = 1445$)
Germany	176	12.2%
United States of America	132	9.1%
United Kingdom	129	8.9%
China	102	7.1%
Netherlands	91	6.3%
Italy	80	5.5%
Japan	54	3.7%
Ireland	48	3.3%
Finland	42	2.9%
Denmark	39	2.7%
Sweden	39	2.7%
Canada	39	2.7%
Norway	36	2.5%
South Korea	35	2.4%
Austria	31	2.1%
Brazil	28	1.9%
Chile	26	1.8%
France	25	1.7%
Spain	22	1.5%
Switzerland	21	1.5%
India	19	1.3%
Portugal	16	1.1%
Mexico	16	1.1%
Thailand	16	1.1%
Iran	15	1.0%
Indonesia	13	0.9%
Greece	12	0.8%
Australia	12	0.8%
Costa Rica	10	0.7%
Bolivia	9	0.6%

Appendix F. Intercoder analysis

We conducted inter-coder analyses for ten randomised articles within our sample of single model studies ($n = 305$) to evaluate causes of disagreement between our two coders, with each coder evaluating five from the other. We subdivided evaluations into qualitative questions and quantitative questions.

For qualitative questions we made use of Cohen's kappa (κ), as shown in Table F.5. Four out of seven questions showed perfect agreement, and the remaining three showed comparatively poor agreement. An additional control test, κ_c , was conducted by removing cases where either reviewer marked the question as indeterminate, which effectively removed all disagreement.

Table F.6 shows the computation of Pearson's r for our quantitative questions, with broad agreement in all categories except for Q17. Disagreement was mostly caused by a study that used ambiguous language when stating the horizon of the exercise [107], which led to coders noting different milestones based either on textual descriptions or the figures shown in the article (35 and 7, respectively).

Table F.5
Inter-coder analysis of qualitative questions conducted for ten single model studies. κ is Cohen's kappa, p_o is the observed agreement among raters, p_e is the hypothetical probability of chance agreement, κ_c is a re-computation of Cohen's kappa with indeterminate cases removed. Cases with a single labelled category ($p_e = 1$) are marked as "inf".

Question	κ	p_o	p_e	κ_c
Study type	inf	1	1	inf
Modelling framework	1	1	0.28	1
Model name	1	1	0.52	1
Formulation	0.44	0.8	0.64	1
World region	1	1	0.26	1
Foresight	0.23	0.6	0.48	1
Endogenous learning	0.41	0.8	0.66	1

Table F.6
 Inter-coder analysis of quantitative questions conducted for ten single model studies. r is Pearson's r , p is the p -value. Indeterminate cases were removed during calculations.

Question	r	p	Indeterminate count
Number of regions	1	0	4
Baseline year	0.96	8.19e-06	0
Target year	1	0	0
Horizon year	0.94	0.0002	1
Number of milestones	0.57	0.087	0
Number of timesteps	0.999	1.07e-06	4

Appendix G. Model resolution analysis

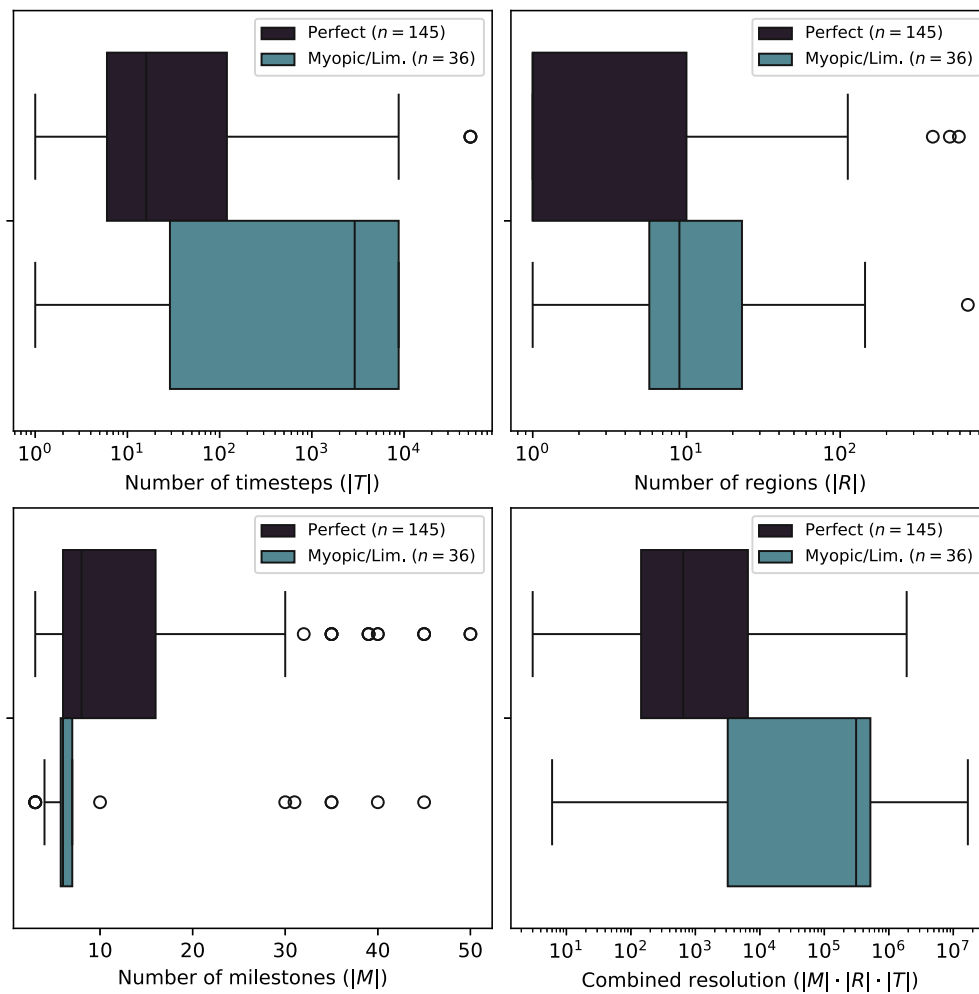


Fig. G.4. Aggregated dimensional characteristics of single model studies with sufficient transparency ($n = 181$). Combined resolution ($|M| \cdot |R| \cdot |T|$) does not necessarily represent computational burden in myopic or limited foresight models given that they are rolling horizon approaches.

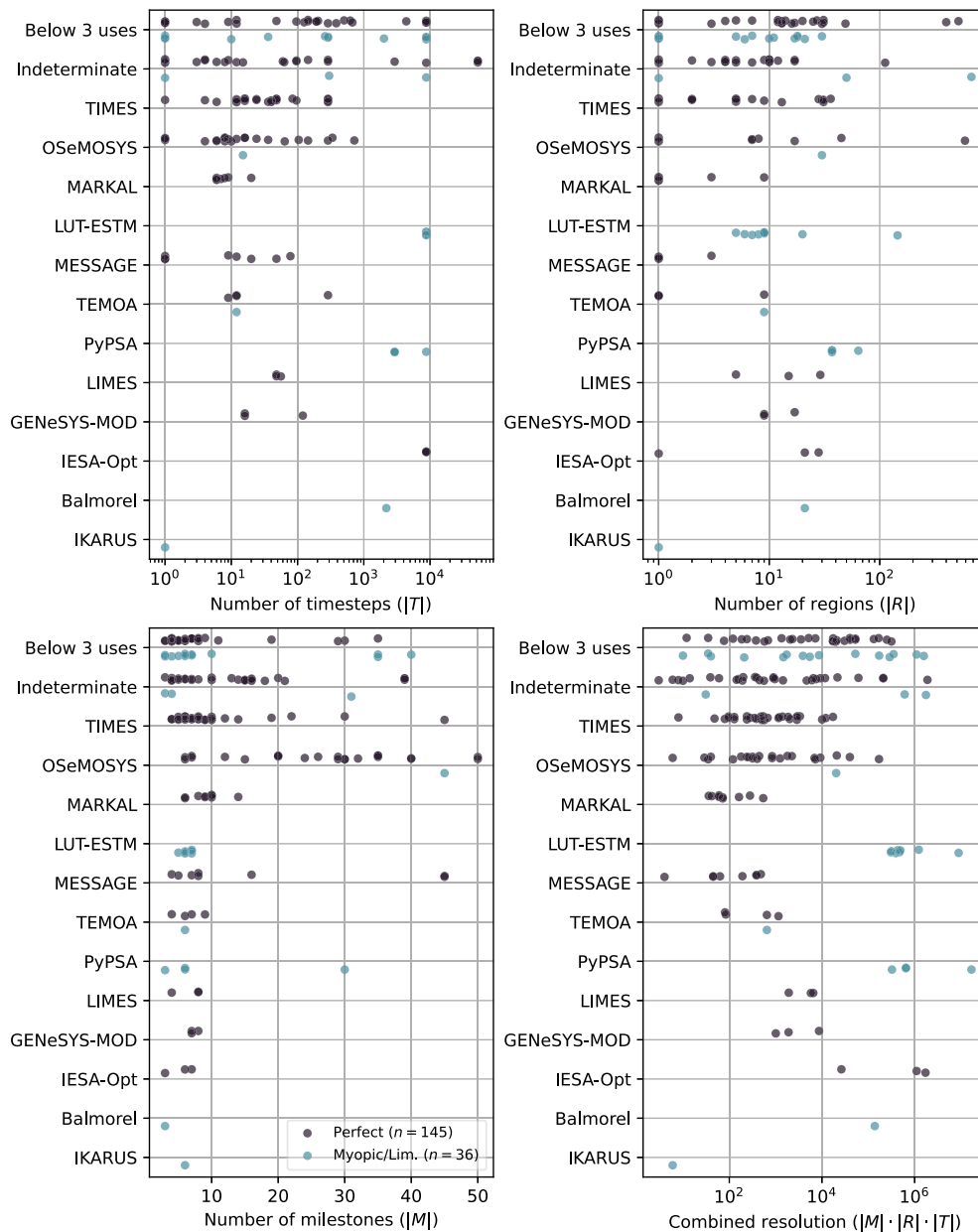


Fig. G.5. Summary of dimensional characteristics of single model studies with sufficient transparency by model framework ($n = 181$). Combined resolution ($|M| \cdot |R| \cdot |T|$) does not represent computational burden in myopic or limited foresight models given that they are rolling horizon approaches.

Appendix H. Long-term horizon extensions

Table H.7

Modelling frameworks with detected end-of-horizon extensions in our sample if single-model studies with sufficient transparency ($n = 181$).

Framework	Counts	References
DIMENSION	1	[108]
GridPath	1	[109]
IKARUS	1	[110]
MARKAL	1	[111]
MESSAGE	1	[112]
OSeMOSYS	6	[113–118]
SWITCH	1	[119]
TIMES	3	[120–122]

Data availability

Curated datasets and code are publicly available on [GitHub](#), [Zenodo](#), and in appendixes.

References

- Rosenbloom D. Pathways: an emerging concept for the theory and governance of low-carbon transitions. *Glob Environ Change* 2017;43:37–50. <https://doi.org/10.1016/j.gloenvcha.2016.12.011>. <https://www.sciencedirect.com/science/article/pii/S0959378016306926>.
- Turnheim B, Berkhout F, Geels F, Hof A, McMeekin A, Nykvist B, van Vuuren D. Evaluating sustainability transitions pathways: bridging analytical approaches to address governance challenges. *Glob Environ Change* 2015;35:239–53. <https://doi.org/10.1016/j.gloenvcha.2015.08.010>. <https://www.sciencedirect.com/science/article/pii/S0959378015300315>.
- Maier HR, Guillaume JHA, van Delden H, Riddell GA, Haasnoot M, Kwakkel JH. An uncertain future, deep uncertainty, scenarios, robustness and adaptation: how do they fit together? *Environ Model Softw* 2016;81:154–64. <https://doi.org/10.1016/j.envsoft.2016.03.014>. <https://www.sciencedirect.com/science/article/pii/S1364815216300780>.
- Swart RJ, Raskin P, Robinson J. The problem of the future: sustainability science and scenario analysis. *Glob Environ Change* 2004;14:137–46. <https://doi.org/10.1016/j.gloenvcha.2003.10.002>. <https://www.sciencedirect.com/science/article/pii/S0959378003000670>.
- Börjesson L, Höjer M, Dreborg K-H, Ekvall T, Finnveden G. Scenario types and techniques: towards a user's guide. *Futures* 2006;38:723–39. <https://doi.org/10.1016/j.futures.2005.12.002>. <https://www.sciencedirect.com/science/article/pii/S0016328705002132>. number: 7.
- Geels FW, Schot J. Typology of sociotechnical transition pathways. *Res Policy* 2007;36:399–417. <https://doi.org/10.1016/j.respol.2007.01.003>. <https://www.sciencedirect.com/science/article/pii/S0048733307000248>.
- IPCC. Climate change 2022 - mitigation of climate change: working group III contribution to the sixth assessment report of the intergovernmental panel on climate change. 1 ed. Cambridge University Press; 2023. <https://doi.org/10.1017/9781009157926> ed. <https://www.cambridge.org/core/product/identifier/9781009157926/type/book>.
- Süsser D, Ceglaz A, Gaschnig H, Stavrakas V, Flamos A, Giannakidis G, Lilliestam J. Model-based policymaking or policy-based modelling? How energy models and energy policy interact. *Energy Res Soc Sci* 2021;75:101984. <https://doi.org/10.1016/j.erss.2021.101984>. <https://www.sciencedirect.com/science/article/pii/S2214629621000773>.
- O'Neill BC, Kriegler E, Riahi K, Ebi KL, Hallegatte S, Carter TR, Mathur R, van Vuuren DP. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Clim Change* 2014;122:387–400. <https://doi.org/10.1007/s10584-013-0905-2>.
- van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose SK. The representative concentration pathways: an overview. *Clim Change* 2011;109:5. <https://doi.org/10.1007/s10584-011-0148-z>.
- International Energy Agency. World energy outlook 2024. Technical Report, International Energy Agency; 2024. <https://www.iea.org/reports/world-energy-outlook-2024>.
- U.S. Energy Information Administration. International energy outlook 2023. Technical Report, U.S. Energy Information Administration; 2023. https://www.eia.gov/outlooks/ieo/pdf/IEO2023_Narrative.pdf.
- Pfenninger S, Hawkes A, Keirstead J. Energy systems modeling for twenty-first century energy challenges. *Renew Sustain Energy Rev* 2014;33:74–86. <https://doi.org/10.1016/j.rser.2014.02.003>. <https://www.sciencedirect.com/science/article/pii/S1364032114000872>. text.ids = Pfenninger.S.2014.EnergySystemsModeling.
- Shu DY, Reinert C, Mannhardt J, Leenders L, Lüthje J, Mitsos A, Bardow A. Overcoming the central planner approach – bilevel optimization of the European energy transition. *iScience* 2024;27:110168. <https://doi.org/10.1016/j.isci.2024.110168>. <https://linkinghub.elsevier.com/retrieve/pii/S2589004224013932>. text.ids = Shu.DY.2024_OvercomingCentralPlanner.
- Hoffman KC, Wood DO. Energy system modeling and forecasting. *Annu Rev Environ Resour* 1976;1:423–53. <https://doi.org/10.1146/annurev.ev.01.110176.002231>. <https://www.annualreviews.org/content/journals/10.1146/annurev.ev.01.110176.002231>. publisher: Annual Reviews.
- Loulou R, Lehtilä A, Kanudia A, Remme U, Goldstein G. Documentation of the times model - part ii: reference manual, 2016. https://iea-etsap.org/docs/Documentation_for_the_TIMES_Model-PartII.pdf.
- Howells M, Rogner H, Strachan N, Heaps C, Huntington H, Kypreos S, Hughes A, Silveira S, DeCarolis J, Bazillian M, Roehrl A. OSeMOSYS: the open source energy modeling system: an introduction to its ethos, structure and development. *Energy Policy* 2011;39:5850–70. <https://doi.org/10.1016/j.enpol.2011.06.033>. <https://www.sciencedirect.com/science/article/pii/S0301421511004897>. number: 10.
- Huppmann D, Gidden M, Fricko A, Kolp P, Orthofer C, Pimmer M, Kushin N, Vinca A, Mastrucci A, Riahi K, Krey V. The MESSAGEix integrated assessment model and the *IX modeling platform* (ixmp): an open framework for integrated and cross-cutting analysis of energy, climate, the environment, and sustainable development. *Environ Model Softw* 2019;112:143–56. <https://doi.org/10.1016/j.envsoft.2018.11.012>. <https://www.sciencedirect.com/science/article/pii/S1364815218302330>.
- Pietzcker RC, Ueckerdt F, Carrara S, de Boer HS, Després J, Fujimori S, Johnson N, Kitous A, Scholz Y, Sullivan P, Luderer G. System integration of wind and solar power in integrated assessment models: a cross-model evaluation of new approaches. *Energy Econ* 2017;64:583–99. <https://doi.org/10.1016/j.eneco.2016.11.018>. <https://linkinghub.elsevier.com/retrieve/pii/S0140988316303395>.
- Ellenbeck S, Lilliestam J. How modelers construct energy costs: discursive elements in energy system and integrated assessment models. *Energy Res Soc Sci* 2019;47:69–77. <https://doi.org/10.1016/j.erss.2018.08.021>. <https://www.sciencedirect.com/science/article/pii/S2214629618306546>.
- Jaxa-Rozen M, Trutnevyte E. Sources of uncertainty in long-term global scenarios of solar photovoltaic technology. *Nat Clim Change* 2021;11:266–73. <https://doi.org/10.1038/s41558-021-00998-8>. <https://www.nature.com/articles/s41558-021-00998-8>. number: 3 Publisher: Nature Publishing Group.
- Creutzig F, Agoston P, Goldschmidt JC, Luderer G, Nemet G, Pietzcker RC. The underestimated potential of solar energy to mitigate climate change. *Nat Energy* 2017;2:17140. <https://doi.org/10.1038/nenergy.2017.140>. <https://www.nature.com/articles/nenergy2017140>.
- Pfenninger S, DeCarolis J, Hirth L, Quoilin S, Staffell I. The importance of open data and software: is energy research lagging behind? *Energy Policy* 2017;101:211–5. <https://doi.org/10.1016/j.enpol.2016.11.046>. <https://www.sciencedirect.com/science/article/pii/S0301421516306516>.
- DeCarolis JF, Hunter K, Sreepathi S. The case for repeatable analysis with energy economy optimization models. *Energy Econ* 2012;34:1845–53. <https://doi.org/10.1016/j.eneco.2012.07.004>. <https://www.sciencedirect.com/science/article/pii/S0140988312001405>. number: 6.
- Kotzur L, Markewitz P, Robinius M, Stolten D. Time series aggregation for energy system design: modeling seasonal storage. *Appl Energy* 2018;213:123–35. <https://doi.org/10.1016/j.apenergy.2018.01.023>. <https://www.sciencedirect.com/science/article/pii/S0306261918300242>.
- Pfenninger S. Dealing with multiple decades of hourly wind and PV time series in energy models: a comparison of methods to reduce time resolution and the planning implications of inter-annual variability. *Appl Energy* 2017;197:1–13. <https://doi.org/10.1016/j.apenergy.2017.03.051>. <https://www.sciencedirect.com/science/article/pii/S0306261917302775>.
- Candas S, Muschner C, Buchholz S, Bramstorf R, van Ouwkerk J, Hainisch K, Löffler K, Günther S, Berendes S, Nguyen S, Justin A. Code exposed: review of five open-source frameworks for modeling renewable energy systems. *Renew Sustain Energy Rev* 2022;161:112272. <https://doi.org/10.1016/j.rser.2022.112272>. <https://www.sciencedirect.com/science/article/pii/S1364032122001927>.
- Plazas-Niño FA, Ortiz-Pimiento NR, Montes-Páez EG. National energy system optimization modelling for decarbonization pathways analysis: a systematic literature review. *Renew Sustain Energy Rev* 2022;162:112406. <https://doi.org/10.1016/j.rser.2022.112406>. <https://www.sciencedirect.com/science/article/pii/S1364032122003148>. text.ids = Plazas-Niño.FA.2022.NationalEnergySystema.
- Lopion P, Markewitz P, Robinius M, Stolten D. A review of current challenges and trends in energy systems modeling. *Renew Sustain Energy Rev* 2018;96:156–66. <https://doi.org/10.1016/j.rser.2018.07.045>. <https://www.sciencedirect.com/science/article/pii/S1364032118305537>.
- Prina MG, Manzolini G, Moser D, Nastasi B, Sparber W. Classification and challenges of bottom-up energy system models - a review. *Renew Sustain Energy Rev* 2020;129:109917. <https://doi.org/10.1016/j.rser.2020.109917>. <https://www.sciencedirect.com/science/article/pii/S1364032120302082>. text.ids = Prina.MG.2020.ClassificationChallengesBottomupa.
- Fattahi A, Sijm J, Faaij A. A systemic approach to analyze integrated energy system modeling tools: a review of national models. *Renew Sustain Energy Rev* 2020;133:110195. <https://doi.org/10.1016/j.rser.2020.110195>. <https://www.sciencedirect.com/science/article/pii/S1364032120304858>.
- Fodstad M, Crespo del Granado P, Hellemo L, Knudsen BR, Pisciella P, Silvast A, Bordin C, Schmidt S, Straus J. Next frontiers in energy system modelling: a review on challenges and the state of the art. *Renew Sustain Energy Rev* 2022;160:112246. <https://doi.org/10.1016/j.rser.2022.112246>. <https://www.sciencedirect.com/science/article/pii/S136403212200168X>.
- Gacitua L, Gallegos P, Henriquez-Auba R, Lorca A, Negrete-Pincetic M, Olivares D, Valenzuela A, Wenzel G. A comprehensive review on expansion planning: models and tools for energy policy analysis. *Renew Sustain Energy Rev* 2018;98:346–60. <https://doi.org/10.1016/j.rser.2018.08.043>. <https://www.sciencedirect.com/science/article/pii/S1364032118306269>.
- Chang M, Thellufsen JZ, Zakeri B, Pickering B, Pfenninger S, Lund H, Østergaard PA. Trends in tools and approaches for modelling the energy transition. *Appl Energy* 2021;290:116731. <https://doi.org/10.1016/j.apenergy.2021.116731>. <https://www.sciencedirect.com/science/article/pii/S0306261921002476>.
- Kotzur L, Nolting L, Hoffmann M, Groß T, Smolenko A, Priesmann J, Büsing H, Beer R, Kullmann F, Singh B, Praktikno A, Stolten D, Robinius M. A modeler's guide to handle complexity in energy systems optimization. *Adv Appl Energy* 2021;4:100063. <https://doi.org/10.1016/j.adapen.2021.100063>. <https://www.sciencedirect.com/science/article/pii/S266679242100055X>.
- Heuberger CF, Rubin ES, Staffell I, Shah N, Mac Dowell N. Power capacity expansion planning considering endogenous technology cost learning. *Appl Energy* 2017;204:831–45. <https://doi.org/10.1016/j.apenergy.2017.07.075>. <https://linkinghub.elsevier.com/retrieve/pii/S0306261917309479>.
- Keppo I, Strubegger M. Short term decisions for long term problems – the effect of foresight on model based energy systems analysis. *Energy* 2010;35:2033–42. <https://doi.org/10.1016/j.energy.2010.01.019>. <https://www.sciencedirect.com/science/article/pii/S0306544210000216>. text.ids = Keppo.I.2010.ShortTermDecisionsa number: 5.
- Poncelet K, Delarue E, Six D, Duerinck J, D'haeseleer W. Impact of the level of temporal and operational detail in energy-system planning models.

- sector units: a case study on uncertainty in china's carbon emission targets. *Comput Chem Eng* 2024;191:108840. <https://doi.org/10.1016/j.compchemeng.2024.108840>. <https://linkinghub.elsevier.com/retrieve/pii/S0098135424002588>.
- [83] Mulholland E, Rogan F, Gallachóir BÓ. From technology pathways to policy roadmaps to enabling measures – a multi-model approach. *Energy* 2017;138:1030–41. <https://doi.org/10.1016/j.energy.2017.07.116>. <https://linkinghub.elsevier.com/retrieve/pii/S0360544217312276>.
- [84] Berntsen PB, Trutnevte E. Ensuring diversity of national energy scenarios: bottom-up energy system model with modeling to generate alternatives. *Energy* 2017;126:886–98. <https://doi.org/10.1016/j.energy.2017.03.043>. <https://linkinghub.elsevier.com/retrieve/pii/S0360544217304097>.
- [85] Loulou R, Labriet M. ETSAP-TIAM: the times integrated assessment model part i: model structure. *Comput Manag Sci* 2008;5:7–40. <https://doi.org/10.1007/s10287-007-0046-z>. <http://link.springer.com/10.1007/s10287-007-0046-z>.
- [86] Bartholdsen H-K, Eidens A, Löffler K, Seehaus F, Wejda F, Burandt T, Oei P-Y, Kemfert C, Hirschhausen C. Pathways for germany's low-carbon energy transformation towards 2050. *Energies* 2019;12:2988. <https://doi.org/10.3390/en12152988>. <https://www.mdpi.com/1996-1073/12/15/2988>.
- [87] Limpens G, Rixhon X, Contino F, Jeanmart H. EnergyScope pathway: an open-source model to optimise the energy transition pathways of a regional whole-energy system. *Appl Energy* 2024;358:122501. <https://doi.org/10.1016/j.apenergy.2023.122501>. <https://linkinghub.elsevier.com/retrieve/pii/S0306261923018652>.
- [88] Marchenko OV, Solomin SV. The development options of nuclear power under carbon dioxide emissions constraints. *Nucl Energy Technol* 2023;9:27–31. <https://doi.org/10.3897/nucet.9.100754>. <https://nucet.pensoft.net/article/100754/>.
- [89] Felling T, Levers O, Fortenbacher P. Multi-horizon planning of multi-energy systems. *Electr Power Syst Res* 2022;212:108509. <https://doi.org/10.1016/j.epr.2022.108509>. <https://www.sciencedirect.com/science/article/pii/S0378779622006241>. `text.ids = FellingT.2022.MultihorizonPlanningMultienergysa`.
- [90] O. US EPA. U.S. grid regions, 2022. <https://www.epa.gov/green-power-markets/us-grid-regions>.
- [91] Béres R, Van Der Wel A, Fattahi A, Van Den Broek M. The impact of national policies on europe-wide power system transition towards net-zero 2050. *Energy* 2024;310:133216. <https://doi.org/10.1016/j.energy.2024.133216>. <https://linkinghub.elsevier.com/retrieve/pii/S0360544224029918>.
- [92] Komiyama R, Fujii Y. Optimal integration of variable renewables in electric power systems of Japan. *J Energy Eng* 2017;143:F4016004. [https://doi.org/10.1061/\(ASCE\)EY.1943-7897.0000361](https://doi.org/10.1061/(ASCE)EY.1943-7897.0000361). <https://ascelibrary.org/doi/10.1061/>.
- [93] Farfan J, Breyer C. Structural changes of global power generation capacity towards sustainability and the risk of stranded investments supported by a sustainability indicator. *J Clean Prod* 2017;141:370–84. <https://doi.org/10.1016/j.jclepro.2016.09.068>. <https://www.sciencedirect.com/science/article/pii/S0959652616314093>.
- [94] van den Berg NJ, van Soest HL, Hof AF, den Elzen MGJ, van Vuuren DP, Chen W, Drouet L, Emmerling J, Fujimori S, Höhne N, Köberle AC, McCollum D, Schaeffer R, Shekhar S, Vishwanathan SS, Vrontisi Z, Blok K. Implications of various effort-sharing approaches for national carbon budgets and emission pathways. *Clim Change* 2020;162:1805–22. <https://doi.org/10.1007/s10584-019-02368-y>. <http://link.springer.com/10.1007/s10584-019-02368-y>.
- [95] Sánchez Diéguez M, Fattahi A, Sijm J, Morales España G, Faaij A. Modelling of decarbonisation transition in national integrated energy system with hourly operational resolution. *Adv Appl Energy* 2021;3:100043. <https://doi.org/10.1016/j.adapen.2021.100043>. <https://linkinghub.elsevier.com/retrieve/pii/S2666792421000354>.
- [96] Gao Z, Gazzani M, Tejada-Arango DA, Soares Siqueira A, Wang N, Gibescu M, Morales-España G. Fully flexible temporal resolution for energy system optimization. *Appl Energy* 2025;396:126267. <https://doi.org/10.1016/j.apenergy.2025.126267>. <https://www.sciencedirect.com/science/article/pii/S0306261925009973>.
- [97] Pfenninger S. Open code and data are not enough: understandability as design goal for energy system models. *Prog Energy* 2024;6:033002. <https://doi.org/10.1088/2516-1083/ad371e>. <https://dx.doi.org/10.1088/2516-1083/ad371e>. publisher: IOP Publishing.
- [98] Harmsen M, Krieger E, van Vuuren DP, van der Wijst K-I, Luderer G, Cui R, Dessens O, Drouet L, Emmerling J, Morris JF, Fosse F, Fragkiadakis D, Fragkiadakis K, Fragkos P, Fricko O, Fujimori S, Gernaat D, Guivarch C, Iyer G, Karkatsoulis P, Keppo I, Keramidis K, Köberle A, Kolp P, Krey V, Krüger C, Leblanc F, Mittal S, Paltssev S, Rochedo P, van Ruijven BJ, Sands RD, Sano F, Strefler J, Arroyo EV, Wada K, Zakeri B. Integrated assessment model diagnostics: key indicators and model evolution. *Environ Res Lett* 2021;16:054046. <https://doi.org/10.1088/1748-9326/abf964>.
- [99] Dekker, Mark M. and Daioglou, Vassilis and Pietzcker, Robert and Rodrigues, Renato and de Boer, Harmen-Sytze and Dalla Longa, Francesco and Drouet, Laurent and Emmerling, Johannes and Fattahi, Amir and Potiou, Theofano and Fragkos, Panagiotis and Fricko, Oliver and Gusheva, Ema and Harmsen, Mathijs and Huppmann, Daniel and Kannavu, Maria and Krey, Volker and Lombardi, Francesco and Luderer, Gunnar and Pfenninger, Stefan and Tsiropoulos, Ioannis and Zakeri, Behnam and van der Zwaan, Bob and Usher, Will and van Vuuren, Detlef. Identifying energy model fingerprints in mitigation scenarios. *Nat Energy* 2023;8:1395–404. <https://doi.org/10.1038/s41560-023-01399-1>. <https://www.nature.com/articles/s41560-023-01399-1>.
- [100] Mikropoulos E, Roelfsema M, Chen H-H, Staffell I, Oreggioni G, Hdidouan D, Thellufsen JZ, Chang MA, Fragkos P, Giannousakis A, Chatterjee S, Ülge-Vorsatz D, Pfenninger S, Pickering B, Victoria M, Brown T, van Vuuren DP. Examining pathways for a climate neutral Europe by 2050; a model comparison analysis including integrated assessment models and energy system models. *Energy* 2025;319:134809. <https://doi.org/10.1016/j.energy.2025.134809>. <https://www.sciencedirect.com/science/article/pii/S0360544225004517>.
- [101] Heuberger CF, Staffell I, Shah N, Mac Dowell N. Impact of myopic decision-making and disruptive events in power systems planning. *Nat Energy* 2018;3:634–40. <https://doi.org/10.1038/s41560-018-0159-3>. <https://www.nature.com/articles/s41560-018-0159-3>. number: 8.
- [102] Held M, Rosat N, Georges G, Pengg H, Bouloukos K. Lifespans of passenger cars in europe: empirical modelling of fleet turnover dynamics. *Eur Transp Res Rev* 2021;13:9. <https://doi.org/10.1186/s12544-020-00464-0>.
- [103] Vázquez M, Rey-Stolle I. Photovoltaic module reliability model based on field degradation studies. *Prog Photovolt Res Appl* 2008;16:419–33. <https://doi.org/10.1002/pip.825>, eprint: <https://onlinelibrary.wiley.com/doi/abs/10.1002/pip.825>. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/pip.825>.
- [104] Lau JH, Baldwin T. An empirical evaluation of doc2vec with practical insights into document embedding generation, 2016. <https://doi.org/10.48550/arXiv.1607.05368>. <http://arxiv.org/abs/1607.05368>. arXiv:1607.05368 [cs].
- [105] Our World in Data Team. Definitions of world regions, Our World in Data, 2018. <https://ourworldindata.org/world-region-map-definitions>.
- [106] Pereira V, Basilio MP, Santos CHT. pyBiBx – a Python library for bibliometric and scientometric analysis powered with artificial intelligence tools. *Data Technol Appl* 2025;59:302–37. <https://doi.org/10.1108/DTA-08-2023-0461>. <http://arxiv.org/abs/2304.14516>. arXiv:2304.14516 [cs].
- [107] Limmeechokchai B, Bahadur Pradhan B, Churnak P, Chaichaloemprecha A, Rajbhandari S, Pita P. Energy system transformation for attainability of net zero emissions in Thailand. *Int J Sustain Energy Plan Manag* 2022;35:27–44. <https://doi.org/10.54337/ijsepm.7116>. <https://journals.aau.dk/index.php/sepm/article/view/7116>.
- [108] Jägemann C, Fürsch M, Hagspiel S, Nagl S. Decarbonizing europe's power sector by 2050 — analyzing the economic implications of alternative decarbonization pathways. *Energy Econ* 2013;40:622–36. <https://doi.org/10.1016/j.eneco.2013.08.019>. <https://linkinghub.elsevier.com/retrieve/pii/S0140988313001928>.
- [109] Chowdhury AFMK, Deshmukh R, Wu GC, Uppal A, Mileva A, Curry T, Armstrong L, Galelli S, Ndhulukula K. Enabling a low-carbon electricity system for southern Africa. *Joule* 2022;6:1826–44. <https://doi.org/10.1016/j.joule.2022.06.030>. <https://linkinghub.elsevier.com/retrieve/pii/S254243512200304X>.
- [110] Martinsen D, Krey V. Compromises in energy policy—using fuzzy optimization in an energy systems model. *Energy Policy* 2008;36:2983–94. <https://doi.org/10.1016/j.enpol.2008.04.005>. <https://linkinghub.elsevier.com/retrieve/pii/S030142150800181X>.
- [111] Kannan R. The development and application of a temporal MARKAL energy system model using flexible time slicing. *Appl Energy* 2011;88:2261–72. <https://doi.org/10.1016/j.apenergy.2010.12.066>. <https://linkinghub.elsevier.com/retrieve/pii/S0306261910005866>.
- [112] Ghadaksaz H, Saboohi Y. Energy supply transformation pathways in Iran to reduce GHG emissions in line with the Paris agreement. *Energy Strategy Rev* 2020;32:100541. <https://doi.org/10.1016/j.esr.2020.100541>. <https://linkinghub.elsevier.com/retrieve/pii/S2211467X20300948>.
- [113] Slimani J, Kadriani A, El Harraki I, Ezzahid E. Towards a sustainable energy future: modeling morocco's transition to renewable power with enhanced OSeMOSYS model. *Energy Convers Manag* 2024;317:118857. <https://doi.org/10.1016/j.enconman.2024.118857>. <https://linkinghub.elsevier.com/retrieve/pii/S0196890424007982>.
- [114] Fernandez Vazquez CAA, Brecha RJ, Fernandez Fuentes MH. Analyzing carbon emissions policies for the bolivian electric sector. *Renew Sustain Energy Transit* 2022;2:100017. <https://doi.org/10.1016/j.rset.2022.100017>. <https://linkinghub.elsevier.com/retrieve/pii/S2667095X22000010>.
- [115] Dallmann C, Schmidt M, Möst D. Between path dependencies and renewable energy potentials: a case study of the Egyptian power system. *Energy Strategy Rev* 2022;41:100848. <https://doi.org/10.1016/j.esr.2022.100848>. <https://linkinghub.elsevier.com/retrieve/pii/S2211467X22000451>.
- [116] Henke HTJ, Gardumi F, Ellefsen Ó, Lítla M, Lærke B, Karlsson K. Exploring european decarbonisation pathways in the power decisions game. *Energy Sustain Soc* 2024;14:41. <https://doi.org/10.1186/s13705-024-00469-w>. <https://energysustainoc.biomedcentral.com/articles/10.1186/s13705-024-00469-w>.
- [117] Moksnes N, Howells M, Usher W. Increasing spatial and temporal resolution in energy system optimisation model – the case of Kenya. *Energy Strategy Rev* 2024;51:101263. <https://doi.org/10.1016/j.esr.2023.101263>. <https://linkinghub.elsevier.com/retrieve/pii/S2211467X23002134>.
- [118] Taliotis S, Shivakumar A, Ramos E, Howells M, Mentis D, Sridharan V, Broad O, Mofor L. An indicative analysis of investment opportunities in the african electricity supply sector — using TEMBA (the electricity model base for africa). *Energy Sustain Dev* 2016;31:50–66. <https://doi.org/10.1016/j.esd.2015.12.001>. <https://linkinghub.elsevier.com/retrieve/pii/S0973082615300065>.
- [119] Verástegui F, Lorca Á., Olivares D, Negrete-Pincetic M. Optimization-based analysis of decarbonization pathways and flexibility requirements in highly renewable power systems. *Energy* 2021;234:121242. <https://doi.org/10.1016/j.energy.2021.121242>. <https://linkinghub.elsevier.com/retrieve/pii/S0360544221014900>.

- [120] Pattupara R, Kannan R. Alternative low-carbon electricity pathways in Switzerland and its neighbouring countries under a nuclear phase-out scenario. *Appl Energy* 2016;172:152–68. <https://doi.org/10.1016/j.apenergy.2016.03.084>. <https://linkinghub.elsevier.com/retrieve/pii/S030626191630410X>.
- [121] Carvajal PE, Li FGN, Soria R, Cronin J, Anandarajah G, Mulugetta Y. Large hydropower, decarbonisation and climate change uncertainty: modelling power sector pathways for Ecuador. *Energy Strategy Rev* 2019;23:86–99. <https://doi.org/10.1016/j.esr.2018.12.008>. <https://linkinghub.elsevier.com/retrieve/pii/S2211467X18301202>.
- [122] Sandberg E, Krook-Riekkola A. The impact of technology availability on the transition to net-zero industry in Sweden. *J Clean Prod* 2022;363:132594. <https://doi.org/10.1016/j.jclepro.2022.132594>. <https://linkinghub.elsevier.com/retrieve/pii/S095965262202193X>.