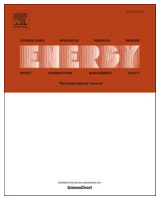
Representing net load variability in electricity system capacity expansion models accounting for challenging weather-years

Energy 316 (2025) 134346

Contents lists available at ScienceDirect

Energy

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



Representing net load variability in electricity system capacity expansion models accounting for challenging weather-years

Jonathan Ullmark *, Lisa Göransson, Filip Johnsson

"Chalmers University of Technology", Gothenburg, Sweden

ARTICLE INFO

Handling editor: Neven Duic

Keywords:
Capacity expansion modeling
Interannual net-load variability
Carbon-neutral electricity systems
Extreme weather-events
Representative time-series
Weather-year

ABSTRACT

Cost-minimizing electricity system models are important tools for understanding conditions for the development of the electricity system. Since the variability of wind and solar power outputs differs between years, a satisfactory representation of variability requires a high time resolution, as well as data that cover multiple decades. This work proposes a weather-year selection method that represents power generation variability by selecting a set of weather years to represent the net-load variability of a broader span of historical weather years (in this work, 39 years). The representativeness is captured in terms of net-load amplitude and duration, such that the electricity demand, as well as the wind- and solar-generation profiles, are considered in their chronologic order, rather than simply as discrete data-points. The weather-year selection method is applied to modeling the North European electricity system with the aims of evaluating the method and investigating the impacts of extreme net-load events on the electricity system composition. The results show that the proposed method can represent the net-load variability of multiple decades using a few selected weather-years. In addition, when the probability of extreme net-load events is accounted for, these extreme events mainly increase the peak thermal capacity and long-term biogas fuel storage capacity.

1. Introduction

Historically, variations in the electricity demand have been handled primarily by adjusting the fuel supply to fuel-based electricity generators. For efforts to make electricity production more sustainable, the limited access to sustainable and affordable fuel-based electricity generation poses multiple challenges to the stable and reliable delivery of electricity. In the temporal context, these challenges range from how to handle real-time grid frequency control to interannual challenges, such as how to implement and dispatch fuel storage units, how to design a system that works in both favorable and unfavorable weather-years, and how to plan and manage variable demands for biomass from forestry and agriculture. While some of the challenges related to achieving a sustainable, affordable, stable and reliable electricity system can be solved through the development and implementation of technological solutions (e.g., advanced inverter control to assist with frequency stability [1]), other challenges comprise a combination of economic aspects, societal risk-aversion, and the weighing of competing aspects of sustainability (such as those related to the different sources and uses of biofuels). Given that the challenges span several research disciplines, new methods and models, as well as knowledge sharing between disciplines, are needed. Understanding the sources of demand and supply variability is a crucial component of the design of a reliable and cost- and resource-efficient electricity system, as is knowledge as to how to manage the variations. These aspects are in focus in this paper.

Variability within an electricity system can arise from both the consumers and the producers of electricity (e.g., wind and solar power), and the outcomes of this variability are reflected in the net load (load minus variable generation). This net load must be supplied by dispatchable sources of power, such as thermal power and storage systems, so as to maintain balance in the grid. The characteristics of the net load variability change depending on the timescale, region, and power generation mix. Similarly, the cost-optimal flexibility measures vary depending on the characteristics of the net load variability (see, for example, [2]).

Energy system optimization modeling [and in particular, electricity capacity expansion (ECE) modeling] is an important tool for studying electricity system variability, flexibility measures, interactions between technologies, and other energy system aspects, as it generates a least-cost system that complies with a range of constraints. However, such

* Corresponding author.
E-mail address: jonathan.ullmark@hotmail.com (J. Ullmark).

https://doi.org/10.1016/j.energy.2024.134346
Received 12 December 2023; Received in revised form 27 December 2024; Accepted 30 December 2024
Available online 6 January 2025

models are limited because increasing the temporal, technological or geographic scope eventually takes the model complexity beyond the point at which it can be solved with modern methods and hardware. As a consequence, different studies and research groups have had to compromise by balancing the level of detail and the scope according to the research question investigated. This is evident in a review of energy and electricity system modeling tools [3], in which a wide variety of scopes and resolutions are identified in the literature. Ideally, to represent variability in an accurate way, the modeling in electricity system planning studies would have to be carried out with all available historical weather data, so that an electricity system could be designed that manages a wide range of net-load conditions. This need for weather data from multiple years was highlighted already in 2016 by Bloomfield et al. [4]. By analyzing the load duration curves and the estimated need for baseload and peak capacity for 35 years of climate data, those authors identified large interannual differences that increased in scenarios with higher penetration levels of wind power. Subsequently, modeling with multiple weather-years has been performed in several studies.

Collins et al. [5] have investigated the impacts of different weather-years on the European power system by running a continent-wide ECE model of Europe's future power system using 30 different weather-years (1985–2015). They have shown that there are differences in the system cost, emissions and cost-optimal composition of the electricity system between different weather-years. They have demonstrated that, in general, these differences increase as the share of electricity generated from renewables in Europe increases (from ca. 13 % in Year 2015 to around 30 % in Year 2030). Based on various indicators, they have also identified two profile years that best represent the 30-year period: 1989 and 2012.

A study conducted by Zeyringer et al. [6], which models Great Britain for weather-years 2001–2010 using 50 geographic nodes, has shown that many investments are sensitive to the weather-year modeled and that simultaneous modeling of multiple years increases investments in flexible, gas-based generation capacity.

Dowling et al. [7] have studied the role and value of long-term energy storages (in the form of hydrogen storage in salt caverns) in a modeled 100 % renewable US electricity system using 39 weather-years (1980–2018). When modeling up to six weather-years simultaneously, they found substantial differences in the cost-optimized investment levels depending on which years were selected. They also found that the long-term energy storage capacity levels, on average, increased as additional weather-years were considered.

Ruhnau and Qvist [8] have modeled investments and operation in the electricity system in Germany using 35 weather-years (1982–2016), and have found that a substantial level (56 TWh) of hydrogen storage is required for a renewable electricity system in Year 2050. They have also highlighted the risk of using individual weather-years, as their results indicate that there is significant variability in the investment levels between individual years and that the investments calculated from looking at individual years can differ greatly from the investments identified when using all weather-years simultaneously. In particular, the long-term storage investments have been found to be consistently underestimated for individual weather-years, except when using the specific year with the highest demand for long-term storage.

The abovementioned studies are designed to ensure that a wide range of weather events is captured. However, this extended temporal scope limits other aspects of the models by restricting the geographic scope or the set of technologies represented. Thus, the options available to manage variability (including trade between regions, demand-side flexibility, supply-side flexibility and storage technologies) have been limited in those previous studies ([5–8]). Therefore, there is a need to identify methods that account for the full range of variability across multiple years, while maintaining an adequate geographic scope and including a broad set of technologies. To model a more-representative time-series without having to make so many technological and/or geographic trade-offs, multiple methods to reduce the number of time-steps have been proposed. In the review of Teichgraeber and Brandt [9], many of these methods are presented and discussed in terms of their benefits and limitations. They divide the various approaches into: (i) quantitative methods, such as clustering; and (ii) qualitative methods, such as hand-picking periods based on a visual analysis of the load or renewable energy duration curves. They also highlight a common problem with methods that aggregate time-series into representative time-slices: the challenge of maintaining chronology in a way that operational constraints and long-term storage balances are included in the modeling. Two methods are used to solve this problem: (i) ordered and linked representative time-slices; and (ii) hierarchical time clustering (i.e., clustering of consecutive time-steps). As explained by Teichgraeber and Brandt, if the modeled time-series are divided into m periods that can be clustered and represented by a smaller number n periods, the first (i) method uses all m periods for the modeling of storage but only n periods for other aspects of the energy system. This causes the model solution time to scale with the number of represented years even when the number of representative periods is kept constant, although not to the same extent as without the use of representative periods. The other (ii) method, described in detail by Pineda and Morales [10], is akin to resolution down-sampling, albeit focused on periods of low variance. This method, referred to as Chronological Time-Period Clustering (CTPC), can reduce the number of modeled time-steps without directly affecting the chronology and, thus, the storage equations. However, CTPC cannot reduce the number of time-steps per year too much without drastically altering the profile of the net load. As such, it too has a limited impact when applied to multiple years (or decades) of weather data.

In a study conducted by Gonzato et al. [11], multiple time-series reduction methods were evaluated in terms of model solution time and error level, as compared with the full (1-year) time-series, when applied in an ECE model. They found that all the methods introduced significant errors in terms of short- and long-term storage capacity investments for both 32 and 128 representative days (out of 365), although some were found to perform better than others depending on the final number of time-steps. Following this benchmarking, Garcia-Cerezo et al. [12] improved the CTPC algorithm by explicitly representing local and global extremes in the resulting time-series. Using an ECE model, they have shown that the modified algorithm yields a lower relative system cost error using 720 time-steps than the original algorithm using 1440 time-steps (1.6 % versus 1.94 %). Domínguez and Vitali [13] have also improved the CTPC algorithm by introducing a multi-stage approach in which not only consecutive hours are merged but also consecutive days or weeks. By first merging similar days or weeks, a lower number of final time-steps can be achieved while maintaining both seasonal and daily variabilities. While this is found to reduce the error produced by an ECE model in terms of investment levels, total cost and unmet demand in most cases, it has a particularly strong impact on the running time of the algorithm, reducing it from circa 45 min to 1 min. Hoffmann et al. [14] have, instead, presented a way to combine representative time-slices (days) and CTPC in an optimal way. They have achieved this by iteratively reducing in small increments either the number of representative days or the number of time-segments within each day, whichever option minimizes distortion of the represented time-series (as judged by the root mean-square error).

An issue that arises when picking representative time-periods (using any aforementioned method) is that related to wind and solar generation, as the regional amplitudes of these sources depend on the design of the electricity generation system. In other words, the representativeness of a given period depends on unknown variables, such as wind and solar power investment levels. This issue is exacerbated when multiple wind sites or classes of wind are considered in each region. Using exogenously given wind capacity levels at the point of aggregating the time-series, Moradi-Sepahvand and Tindemans [15] have presented and evaluated a method to generate representative days that include the net-load extremes. Using an ECE model with 21 binary investment variables and

four candidate technologies, they have shown that the modified CTPC method [12] gives a total system cost error of 10.1 %, or 8.7 % when combined with the multi-stage approach presented in Ref. [13]. The method proposed by Moradi-Sepahvand and Tindemans has been found to reduce this total system cost error to 2.2 %, although that evaluation is based on a very small-scale scenario. Given that this method depends on exogenous variable renewable capacity levels, its benefits may be more-difficult to realize in a model that features both wind and solar investments across multiple regions. An alternative approach that does not require generation capacity levels as the input has been proposed by Hilbers et al. [16] and Zhang et al. [17]. This approach consists of two stages. In the first stage, an ECE model is applied to subsamples of the full time-series and in the second stage, time-periods are clustered using also their respective ECE model outputs. Wind and solar power investment levels, storage usage, net-load and electricity price are some of the ECE model variables that have been proposed to improve the clustering. A similar method has been presented by Liu et al. [18], although their algorithm places special emphasis on the representative extreme day and features a loop between the two stages, which is repeated until the same representative extreme day is found twice in a row. In a subsequent study [19], Hilbers et al. have combined their previous two-stage clustering method with ordered and linked representative days, to allow also for inter-day storage. Compared with single-stage clustering with the same number of representative days, they show that this two-stage approach reduces the error level, particularly in terms of the unmet demand when the resulting capacity mix is applied to a dispatch-only run for the full time-series. However, the addition of the first stage (i.e., modeling of random time-slice samples) increases the solution time relative to the number of years to be represented, even when the final number of representative days is kept constant. Moreover, the linking and ordering of representative days to enable long-term energy storage increase the solution time as the number of years is increased.

Notably, there are studies that take approaches other than time-series aggregation to identify a capacity investment mix based on decades of weather-years. With the aim of producing a near-optimal solution that is less-susceptible to infeasibility resulting from real-world uncertainties, while also not requiring that all weather-years are simultaneously optimized, Grochowicz et al. [20] have presented a method that is based on each weather-year's near-optimal solution space. Using an ECE model with 31 interconnected European countries, Grochowicz et al. [20] have analyzed the near-optimal solutions of 41 weather-years space to find a "robust" solution in the intersection. While this approach produces a capacity mix that is less-prone to infeasibility, it may not be suitable when studying aspects of the electricity system where the interest lies in the impact on the total system cost or the cost-optimal investment levels.

Despite the ongoing and collaborative improvements being made to time-reduction methods, they still present significant downsides when applying a high number of weather-years to ECE models with large amounts of variable electricity generation and multiple flexibility options. More specifically, these methods only partially decouple the model size, and thus the solution time, from the temporal scope represented when long-term storage is included in the modeling. This study tests the hypothesis that a few well-chosen weather-years are sufficient to represent the weather variability of multiple decades.

The weather-year selection method proposed here is applied to investigate how extreme net-load events affect the cost-optimized electricity system design. To the best of the authors' knowledge, no previous studies have investigated this question while accounting for multiple weather-years with high temporal resolution, international electricity trade, and multiple options regarding flexibility and electricity generation.

In summary, the aims of this work are.

1) To propose and evaluate a method that identifies representative sets of weather-years that have both typical and particularly challenging high net-load events.
2) To use the representative sets of weather-years to analyze the impacts on the least-cost system composition when considering extreme net-load events.

An extreme net-load event may be associated with extreme weather, which may impact the power system infrastructure. The impact of extreme weather on the power system infrastructure is outside the scope of this work.

2. Methodology

The methodology of this work consists of two parts: (i) a novel algorithm for uncovering representative sets of weather-years; and (ii) a linear ECE model that is used to evaluate the set-finding algorithm and investigate the impacts of challenging net-load events on the electricity system composition. Algorithm 1 gives the six steps taken to identify the sets of weather-years that represent the variations in net load for the entire period of 1980–2019. The period of 1980–2019 is chosen due to the availability of pertinent data, and the 40-year period is split into 39 weather-years (disregarding the first half-year of 1980 and the last half-year of 2019), which are then combined to construct the weather-year sets. As formulated in this work, the algorithm does not require exogenous capacity levels of non-dispatchable generation to calculate the net load; instead, these capacity levels are generated by applying the ECE model to a first guess of representative weather-year(s). If the non-dispatchable capacity levels are known beforehand, a representative set of weather-years can be generated by following Steps 2–5 only once.

Algorithm 1. Finding a representative set of weather-years

1. Let iterations $i = 0$ and create a subset S_i of one or more weather-years $y \in Y$, where Y is the complete set of weather-years that should be represented. For each $y \in Y$, the electricity consumption $d_{y,t}$ and generation profiles $g_{y,t,p}$ for each variable renewable energy (VRE) technology $p \in P^{VRE}$ are needed for all time-steps $t \in T$.
2. Apply the ECE model to the load and generation a time-series that belongs to the years in S_i, so as to find the capacity levels $c_{p,i}, \forall\, p \in P^{VRE}$ that minimize the total system cost for S_i.
3. Use capacities $c_{p,i}\ \forall\, p \in P^{VRE}$ to calculate the net load $n_{y,t,i}$, and let n_i^* denote the concatenated net-load for all $y \in Y$ in chronologic succession such that $|n_i^*| = |Y|^*|T|.\ n_{y,t,i} = d_{y,t} - \sum_{p \in P^{VRE}} \left(g_{y,t,p} {}^*c_{p,i}\right), \forall y, t \in Y, T$
4. Convert† the net load $n_{y,t,i}$ into a net-load recurrence matrix $M_{y,i}$ for each $y \in Y$, and convert the full-period net-load n_i^* into a reference recurrence matrix M_i^*. Both $M_{y,i}$ and M_i^* should be made into equal size $m \times n$.
5. Find the weights $w_{y,i}$ that minimize the error E_i in the following equations, where X is the size that the final set S should have. $\min(E_i) = \sum_{j=1}^{m} \sum_{k=1}^{n} \left|M_{j,k}^{diff}\right|$
   $M_{j,k}^{diff} = \sum_{y \in Y} (M_{y,i,j,k} {}^*w_y) - \frac{M_{i,j,k}^*}{40}, \forall\, i \in \{1,2,\ldots,m\}, j \in \{1,2,\ldots,n\}\ \sum_y w_{y,i} = 1$
   $0 \le w_{y,i} \le x_{y,i}, \forall y \in Y, x_y \in \{0,1\}$
   $\sum x_{y,i} \le X, x_y \in \{0,1\}$
6. Increment i and repeat Steps 2–5 for $S_{i+1} = \{y \in Y | x_y = 1\}$ until $S_{i+1} = S_i$ or $S_{i+1} = S_{i-1}$. In the latter case, Steps 3–5 are repeated one final time using $(c_{p,i+1} + c_{p,i})/2$

† The process of converting $n_{y,t,i}$ into $M_{y,i}$ is further described in the Step 4 subsection.

In addition to the definition given in Algorithm 1, the process is illustrated in Fig. 1 and further explained in the following subsections. As shown in Fig. 1, the algorithm interfaces with the ECE model by giving a set of weather-years to be modeled as input to the ECE model and taking the resulting capacity mix provided by the ECE model as the input to the next step of the algorithm. The capacity mix provided by the ECE model is used in Steps 3 and 4 of the algorithm to generate the net load for all 39 weather-years. In Step 5, a set containing a limited number of weather-years is selected to represent the net load of all 39

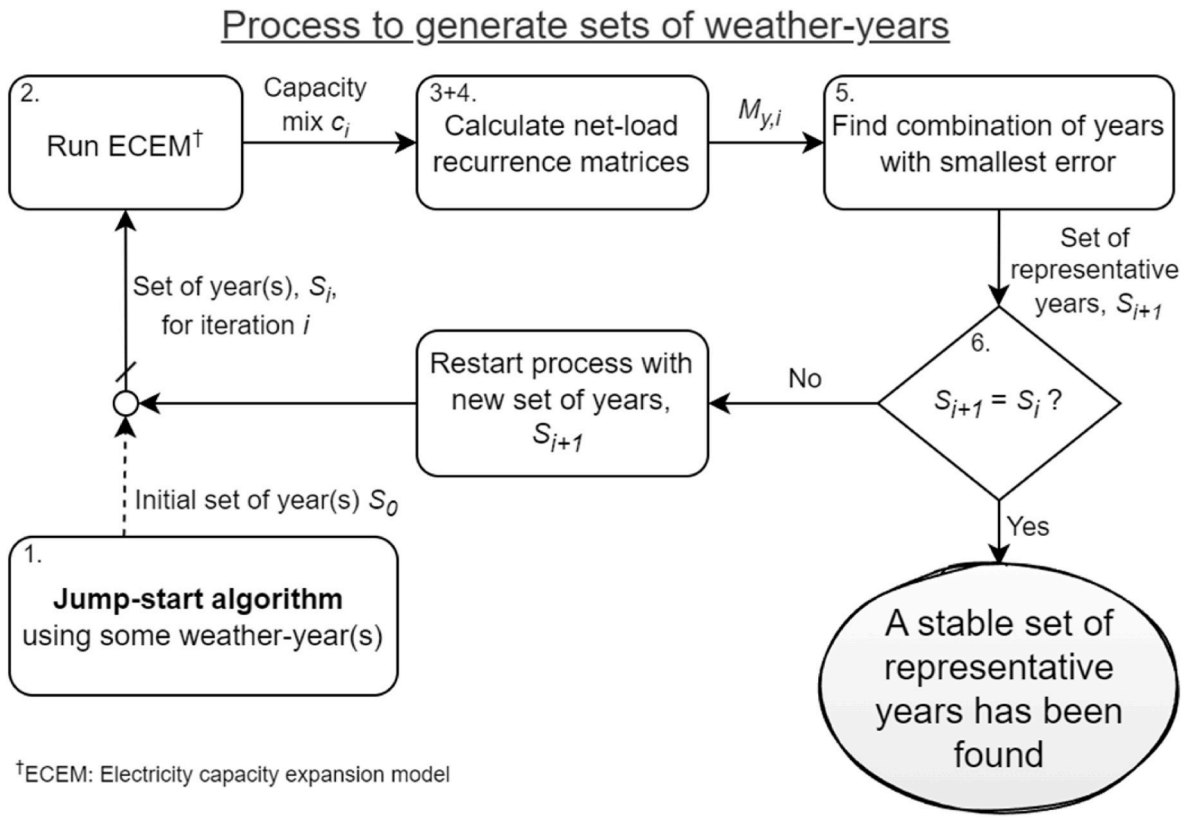


Fig. 1. Flowchart illustrating the algorithm for finding a set of weather-years that represents the net-load variations in the 40-year period of 1980–2019.

weather-years. If this set of weather-years is the same as the set originally given to the ECE model, a set of representative years has been identified. Otherwise, the new set of weather-years is provided to the ECE model to calculate a new capacity mix and the process is reiterated. The features of the ECE model are summarized in the Step 2 subsection, and a mathematical summary is provided in Appendix A.

It should be noted that, absent exogenously given capacity levels, the ideal application of the algorithm would be to let the initial set defined in Step 1 include the complete period Y for which load and weather data are available. In that situation, the set of weather-years S generated in Step 5 would be the set that most closely resembles the full period Y in terms of net-load characteristics for the capacity mix that gives the minimal system cost for the full period Y. However, the algorithm is herein started using single years as S_0 to test the usefulness of the algorithm in those cases in which modeling the entire available time-series is impossible. As such, at the point of convergence in Step 6, the resulting technology mix is cost-optimal for the set of representative weather-years used to generate it. Algorithm 1 is executed for two groups of weather-year sets and for a progressively increasing number of years within the sets. The first and main group of sets combines hand-picked extreme-weather years with years that are identified through the fingerprint-matching optimization, while the second group only uses years that are discovered through the optimization. The initial weather-year used to jump-start the algorithm (S_0 in Fig. 1) should be as close to a representative year as possible, to expedite the iterative process. For example, the initial weather-year may be selected for having close to the average potential yearly wind output, thereby representing a "normal" year. However, single years are herein used to test the usefulness of the algorithm in those cases in which modeling the entire available time-series is impossible. The algorithm is, in this study, jump-started twice using different weather-years each time, i.e., 2012 (Jan 1, 2012 to Dec 31, 2012) and 2016–2017 (July 1, 2016 to June 30, 2017). Two starting points are used to test the robustness of the algorithm with regards to the specific starting point used. The weather-year 2012 was selected, since it has been used in other studies (see, for example, [21,21,22,22–24]) and it has been recommended as a representative year for the European power system [5]. The weather-year of 2016–2017 was selected as having close-to-average wind full-load hours. The scripts used in this work for selecting years, reading the model output data, and running Steps 3–5 are available in a public GitHub repository [23].

Step 2. Electricity capacity expansion modeling

The model used in this work is a linear ECE model based on the model used in a previous study by the authors [26]. Modifications made to the model allowed for multiple weighed weather-years to be modeled simultaneously. Each weather-year has an hourly time resolution, and the model features six interconnected regions, covering Sweden, Finland, Norway, Denmark, Germany and The Netherlands, as illustrated in Fig. 2. The model includes exogenous generation capacities only for nuclear power and hydropower, where the installed capacity is less-flexible and the lifetime is expected to extend beyond Year 2050. A wide range of technologies is available for investment (for a full list, see Table A1 in Appendix A), including nuclear power, bio-based base-load, mid-load and peak-load power, and wind and solar power at multiple sites for each region, as well as batteries and hydrogen storage systems (including optional fuel cells). The ECE model also enables a linearized representation of thermal cycling with start-up and part-load costs, based on Weber [25] and described in detail in a study carried out by Göransson [26]. Transmission capacity between regions is exogenously set as a combination of the real transmission capacity across the regional borders (illustrated in Fig. 2) and estimates of future expansion projects, as projected by ENTSO-E. A mathematical description of the ECE model and a list of the technologies available for investment can be found in Appendix A.

In addition to the traditional electricity demand, the ECE model features the hourly demand profiles for future heating and electrified transportation systems, as well as the yearly electricity and hydrogen

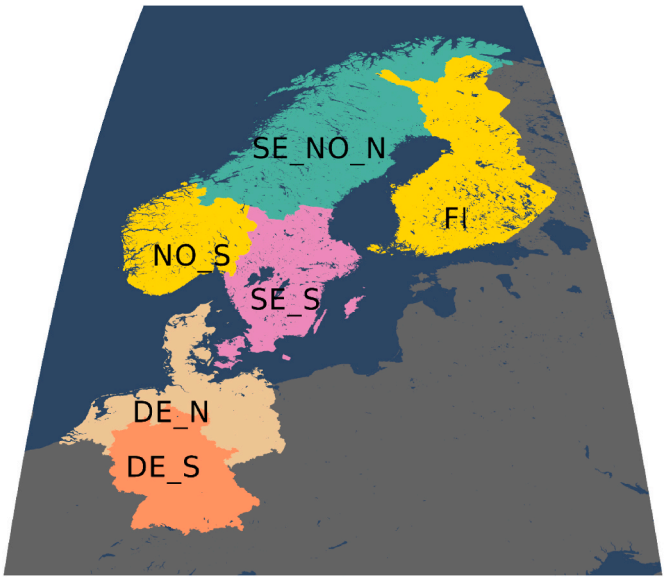


Fig. 2. Map showing the geographic scope and level of resolution of the present study. Note that some regions (SE_S and DE_N) include countries that are not included in the region label.

demands for a future electrified industry. The yearly loads are listed in Table 1, and the profiles are available elsewhere [27]. While the industry demands are evenly spread across all hours of the year, the use of batteries and hydrogen storage units can disconnect the generation of electricity from the consumption of electricity. For the heating demand, it is assumed that district heating (DH) networks will deliver the same amount of heat as they do today. The non-DH portion of the hourly heating demand is instead assumed to be fully electrified (using heat pumps that produce three units of heat per unit of electricity consumed), so this is added to the hourly electricity demand. The implementation of the electrified transport sector is based on the study of Taljegard et al. [28], and includes both the electricity and hydrogen demands for transportation to replace the fossil fuel demand. This implementation uses an aggregated representation of the vehicle fleet and assumes that 30 % of the electric car fleet is charged strategically (included in the optimization), while the remainder of the fleet is charged when the car arrives at home.

The profiles of the demand, wind and solar power used in this work are based on the ERA5 reanalysis data [31], and are generated by scripts available in the GlobalEnergyGIS repository at GitHub [32] and developed for a study carried out by Mattsson et al. [33]. This method uses machine learning to enable the generation of demand profiles for years and regions for which the measured electricity demand is missing but data for the temperature, wind and solar irradiation levels are available. For this work, the machine training for the electricity demand is performed on the demand profiles for 44 regions during Year 2015, and the training for the heating demand is performed on the demand profiles for 28 regions during the period of 2008–2015. However, high and low peaks tend to be underestimated using this approach. This is evident when the synthetic electricity load analyzed in isolation, albeit to a much lower degree when combining a traditional (synthetic) electricity load with a future projected electrified heating demand (which is highly temperature-dependent). It should also be noted that the profiles only capture already existing demand patterns, which means that they can only represent traditional loads. The settings used in the creation of the profiles for this work, along with the profiles, are publicly accessible in GitHub [27].

2.1. Step 4. calculating net-load recurrence matrices

One way to represent the net-load variability of an electricity system is to enumerate the recurrences of net-load events for each amplitude and duration. The resulting matrix can be illustrated using a heatmap or used as a fingerprint of that period's net load variability. This process was originally described in a study by Göransson [34] (see *Algorithm 1*), and is herein modified to create a matrix that is better suited for use as a fingerprint. This modification is carried out, in Step 5 of *Algorithm 1* in Ref. [34], by counting each net-load event as both its maximum duration (as in Ref. [34]) and as each discrete duration between zero and the maximum. In this step, a 12-h rolling average is also applied to the hourly net load, so as to reduce the low-energy net-load fluctuations from solar photovoltaics (PV). It is assumed that short-duration variations ($\leq$12 h) are similar for all the years and not relevant in terms of the selection of representative weather-years. If this assumption is not valid, the rolling average can be applied for a shorter duration (or removed) to represent more accurately short-duration events. By doing so, the demand for rapid power cycling from batteries or thermal power plants is better represented, albeit at the risk of long-duration energy deficiencies being counted as multiple events of shorter duration. While transmission bottlenecks between regions are accounted for in the electricity system optimization, the net loads of all the regions are aggregated when selecting weather-years. The benefits and drawbacks of doing so are further covered in the *Discussion* section. In Fig. 3, a recurrence heatmap for the 1980–2019 reference period is shown. The heatmap shows that net-load recurrence is highest for low durations and low amplitudes, with the peak net load being 360 GW and the longest positive net-load duration being 127 days. However, the 127-day event has a very low amplitude and can easily be covered by the base-load thermal or hydropower capacity. In the reference system shown in Fig. 3, there are approximately 32 GW of reservoir hydropower and 51 GW of nuclear power. Above this amplitude (at around 83 GW), the longest duration is 56 days, which is annotated in Fig. 3 as belonging to weather-year 2002–2003 (i.e., July 1, 2002 to June 30, 2003). The contours of other weather-years that are on, or lie close to, the heatmap outline are also highlighted in Fig. 3.

From an energy system analysis point-of-view, for a set of weather-

Table 1
Electricity and hydrogen demands (separated by forward slash) for each sector and industry considered in the ECE model. For heating, the total energy demand is listed.

	Traditional load	Heating	Transport [28]	Industry			
				Steel + iron [29]	Cement[a]	Ammonia[a]	Battery production [30]
SE_NO_N	41.3	15.0	4.1/2.1	1.8/31.7	1.2	0.0/0.5	3.8
NO_S	150.5	43.3	12.2/7.2	0/0	1.2	0.2/2.5	5.4
SE_S	201.6	92.5	23.7/9.9	1.4/16.9	3.3	0/0	2.3
FI	117.2	63.6	22.3/13.9	2.1/4.4	1.5	0/0	1.9
DE_N	419.2	278.3	77.0/24.2	11.2/23.3	11.0	1.6/21.3	10.2
DE_S	566.9	443.2	127.8/42.1	18.6/38.7	26.3	1.2/15.5	13.0
Conversion factors [kWh/kg]				0.82/1.7	0.96/0	0.45/5.98	47.23[b]/0

[a] Production is assumed to remain at current levels.
[b] kWh of electricity consumed per kWh of battery capacity produced.

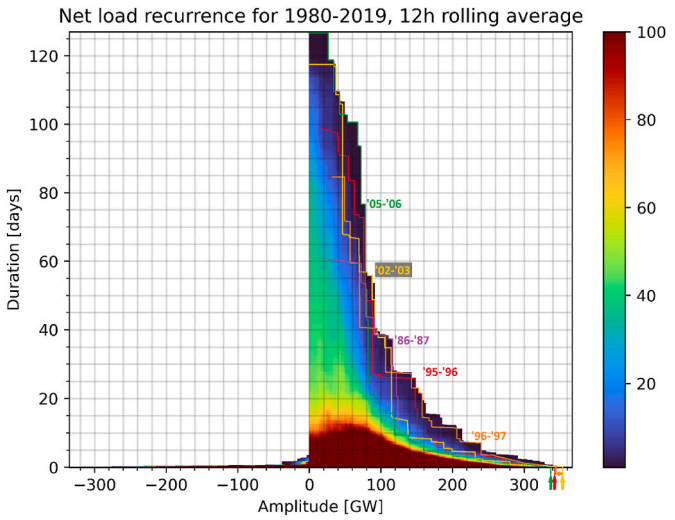


Fig. 3. Heatmap of the recurrence of each combination of amplitude and duration of the net load between Year 1980 and Year 2019, for the capacity mix with 66 % wind and solar power, as obtained from the modeling in this work. Years that make up the outline of the heatmap are contoured and annotated (e.g., "05–06") by hand. Note that the heatmap color scale saturates at 100 recurrences, in order to enhance clarity in the medium-to-low-recurrence regions.

years to be representative it should include both the medium-to-high-probability periods of various seasons and the outlier low-probability periods (particularly the challenging ones). Challenging events that are of interest include both the highest peak in net load and the longest durations of positive net load, of various levels. Both events can be found in Fig. 3 by identifying the years at the contour. These two components – highest peak and longest duration of positive net load – ensure that the electricity system can deal with both the highest power demand and the largest energy deficiencies. In the example shown in Fig. 3, there are multiple years with extraordinary peak net loads (after applying a 12-h rolling average), while the weather-year 1985–1986 stands out as having the highest peak net load (366 GW), followed by the weather-year 2002–2003 (363 GW). Similarly, while there are multiple years with long-duration, positive net-load events, the weather-years 2005–2006, 2002–2003, 1996–1997 and 1986–1987 have especially long duration events at various amplitudes. If one or more of these years at the contour can be visually identified and hand-picked for inclusion in the representative set of weather-years (Y_{i+1} in Fig. 1), at least the representation of extreme net-load events found in the hand-picked years can be guaranteed. While the exact criteria for the best years to pick are difficult to quantify, it is clear from Fig. 3 that (with 83 GW of dispatchable nuclear power and hydropower) a combination of weather-years 2002–2003 and 1996–1997 would cover a large portion of the contour. Regardless of whether the extreme weather-years are hand-picked in this way, the base/normal weather-years must be selected in a different way, as described in the next subsection.

2.2. Step 5. weather-year selection optimization

Using the BlackBoxOptim and JuMP software packages in Julia, sets of weather-years are optimized to find the optimal way to combine them to represent the reference 1980–2019 matrix (after being divided by 40). The optimization minimizes the error found through Eqs. (1) and (2) by changing the weights (w_i) of a number of years given by the set size. As the weights represent probability, their sum must equal one. If some years are qualitatively identified as extremes that define the longest duration or highest peak, such as Years 2002–2003 and 1996–1997 in Fig. 3, it may be appropriate to lock their weights at 1/40 (or however many years of profiles are used in the analysis), so as to represent correctly the probability of the extremes that they contain. In Eq. (1), the difference between the reference matrix and the combined individual years is calculated as:

$$M_{diff} = \sum_i M_i{}^* w_i - M_{ref} \quad (1)$$

The elements in the difference matrix, M_{diff}, are then summed to find the total error of the combination of years and weights. In Eq. (2), the absolute value of each element is summed, while the square root, square, and $\log_{10}$ of $|M_{diff}| + 1$ of each element are also used in the evaluation of this methodology.

$$E = \sum |M_{diff}| \quad (2)$$

The summation function changes the relative value of the different amplitudes in M_{diff} but also makes E more computationally demanding to calculate. In particular, the square and $\log_{10}$ values may be of interest, as they accentuate large and small $|M_{diff}|$ elements, respectively. Accentuating the smaller elements makes the larger areas in the top part of the heatmap more-valued, thereby presenting a more-accurate representation of rare net-load events. Accentuating the larger elements makes the optimization focus more on high-recurrence events, thus assigning "normal" years a higher priority in the optimization. If the extreme net-load events are captured by explicitly including years that stand out in a heatmap such as that in Fig. 1, the priority when selecting the remaining year(s) should be to represent accurately the more-regular net-load events, i.e., using the square or absolute summation. Further details of the fingerprint-matching step are provided in Appendix B.

2.3. Electricity composition sensitivity

The ECE model is run for two extra sets of weather-years. The first extra set has the same years as one of the sets with hand-picked years, although with equal weights for all years. This set is used to investigate how sensitive the electricity system composition that results from the modeling is to the weights used. In particular, this will give a higher representation of the extreme (hand-picked) years. The second extra weather-year set includes all available weather years and is used as the reference (labeled *All years*) when evaluating the results. *All years* is modeled with a time resolution of 2920 h (every third hour) instead of 8760 h, so as to limit the size of the ECE model.

3. Results

This section is divided into three parts. First, the set of weather-years found using the proposed algorithm is presented. Second, an evaluation of the weather-year selection methodology, in which the selected sets are applied in an ECE model, is described. Third, the ECE model results are used to identify cost-efficient electricity system compositions, with accounting for unusually challenging net-load events.

3.1. Weather-year set selection

Table 2 presents the representative set of weather-years identified (using the algorithm in Fig. 1) and analyzed in this work, which includes two groups of weather-year sets: sets with and sets without two hand-picked (2 HP) extreme years (listed in the first two columns). Each group features multiple sets with an increasing number of years found through optimization, denoted as *X opt.* To investigate the impact of the start year ("Initial year Y_0" in Fig. 1), the algorithm was run twice for *2 HP + 1 opt.*, *2 HP + 2 opt.*, and *2 HP + 3 opt.*, once starting with Year 2012 (Jan 1, 2012 to Jan 31, 2013) and once starting with Year 2016–2017 (July 1, 2016 to June 30, 2017). The two initial years result in the same final set for *2 HP + 2 opt.* and *2 HP + 3 opt.*, whereas the algorithm generates different sets for *2 HP + 1 opt.* This indicates that, in some cases, the choice of initial year affects the outcome of the

Table 2
The years and their weights for all representative weather-year sets analyzed in this work. The difference between the two groups lies in whether or not hand-picked sets are included in the set.

2 HP + 1 opt.	Years	2002–2003	1996–1997	2001–2002					
	Weights	0.025	0.025	0.950					
2 HP + 2 opt.	Years	2002–2003	1996–1997	2000–2001	2016–2017				
	Weights	0.025	0.025	0.360	0.590				
2 HP + 3 opt.	Years	1995–1996	1996–1997	1993–1994	2012–2013	2013–2014			
	Weights	0.025	0.025	0.276	0.343	0.331			
2 HP + 4 opt.	Years	2002–2003	1996–1997	1990–1991	1994–1995	2008–2009	2015–2016		
	Weights	0.025	0.025	0.242	0.208	0.252	0.248		
2 HP + 5 opt.	Years	2002–2003	1996–1997	1993–1994	2000–2001	2003–2004	2010–2011	2011–2012	
	Weights	0.025	0.025	0.136	0.189	0.199	0.238	0.188	
2 HP + 6 opt.	Years	2002–2003	1996–1997	1993–1994	1999–2000	2000–2001	2003–2004	2010–2011	2011–2012
	Weights	0.025	0.025	0.117	0.115	0.182	0.201	0.191	0.144
2 HP + 2 opt., eq. w.	Years	2002–2003	1996–1997	2000–2001	2016–2017				
	Weights	0.25	0.25	0.25	0.25				
2 opt.	Years			2000–2001	2016–2017				
	Weights			0.376	0.624				
3 opt.	Years			1981–1982	1992–1993	1999–2000			
	Weights			0.317	0.246	0.437			
4 opt.	Years			1990–1991	1994–1995	2008–2009	2015–2016		
	Weights			0.280	0.223	0.272	0.225		

algorithm. A comparison of the resulting electricity system capacity mixes of the two versions of *2 HP + 1 opt.* is shown in Appendix C.

In Tables 3 and 4, the iterative process of the algorithm is shown for *2 HP + 1 opt.* and *2 HP + 2 opt.*, respectively. Each row, representing one step in the iteration, shows the hand-picked and optimized years that were identified using the net-load matrix of the set in the previous row, with iteration 0 representing the starting point used. However, as can be seen for both cases, the choice of hand-picked extreme weather-years may change as the wind and solar power mix changes between iterations. In Table 4, the breaking of a back-and-forth loop is also shown, i.e., the average capacity mix from iterations 4 and 5 is used in the final iteration. With the set selection method suggested in this work, the final sets are generally established after 3–5 iterations.

3.2. Method evaluation

For each weather-year or set of weather-years, the ECE model generates the outputs of a technology capacity mix and the hourly operation of the technology mix. The technologies include both electricity-generating technologies and storage technologies. In the results, hydrogen storage units are deployed to decouple the operation of the electrolyzer from the hydrogen demand of industry, and while hydrogen-driven, electricity-generating technologies (fuel cells) are available for investments, no such investments are made. The battery storage capacities are paired with separate (dis)charging capacities, following the storage capacity at a ratio of 4.8–7.5 GWh/GW (the lower values are for model runs with lower battery storage capacities).

3.2.1. Evaluating individual weather-years

A comparison of the installed capacities (generation and storage) for 11 individually modeled weather-years is shown in Fig. 4. The weather-years shown include the two years with the highest total thermal capacities (1989–1990 and 1994–1995) as well as the two extreme weather-years (1996–1997 and 2002–2003) that appear repeatedly in Table 2, and the two weather-years used to start the algorithm (2012 and 2016–2017). The technology mix from a model run featuring all the weather-years simultaneously (*All years*) is also shown. Among the individual weather-years in Fig. 4, large variability is seen with regards to the thermal, solar PV and battery storage capacities. The wind and hydrogen storage capacities also vary, albeit to lesser extents. In Fig. 4, neither 2016–2017, which was a weather-year with average wind full-load hours, nor 2012, a weather-year that was found to be representative of Europe [5], show capacity levels similar to that of *All years*. However, 1989–1990 and 1994–1995 capture the high level of investment in thermal capacity seen in *All years*, and also resemble *All years* in terms of solar PV, battery and hydrogen storage capacities. For 1989–1990, the wind power capacity also resembles that of *All years*. Two years with annual full-load hours for wind and solar power similar to that of 1989–1990 are included in Fig. 4: 1991–1992 and 1992–1993. Although 1991–1992 and 1992–1993 resemble 1989–1990 in terms of wind power capacity, they differ from 1989 to 1990 and *All years* in terms of their thermal and solar PV capacities. These examples show that years with similar average annual full-load hours of wind and solar power may still differ with regards to their cost-optimal levels of thermal generation and storage capacity. Thus, the yearly full-load hours of wind and solar power do not appear to be a good predictor of how well a year represents *All years*. The aggregated recurrence matrices shown in Fig. 3 and the net-load duration curves in Appendix D also fail to explain why the individual weather-years of 1989–1990 and 1994–1995 have a cost-optimal capacity mix that is similar to that of *All years*.

To investigate further the representativeness of weather-years 1989–1990 and 1994–1995, which were found in Fig. 4 to resemble *All years* in terms of installed capacities, the use of biogas is compared for multiple years in Table 5. The weather-years listed in Table 5 include

Table 3
Resulting set of weather-years for each iteration to find 2 HP + 1 opt., starting from weather-year 2016–2017. The hand-picked weather-years were selected to cover the peak net-load demand, as well as to cover the longest duration events involving 80–200 GW of net load.

Iteration	Opt. year 1	Hand-picked 1	Hand-picked 2
0		2016–2017	
1	2001–2002	1986–1987	1989–1990
2	2001–2002	1996–1997	2002–2003
3	2001–2002	1996–1997	2002–2003

Table 4
Resulting set of weather-years for each iteration starting from Year 2016–2017. The hand-picked years were selected to cover the peak net-load demand, as well as to cover the longest duration events entailing 80–200 GW of net load.

Iteration	Opt. year 1	Opt. year 2	Hand-picked 1	Hand-picked 2
0			2016–2017	
1	1982–1983	2010–2011	1986–1987	1989–1990
2	1985–1986	2013–2014	1996–1997	2002–2003
3	1982–1983	2010–2011	1996–1997	2002–2003
4	2000–2001	2016–2017	1996–1997	2002–2003
5	1982–1983	2010–2011	1996–1997	2002–2003
Tiebreaker	2000–2001	2016–2017	1996–1997	2002–2003

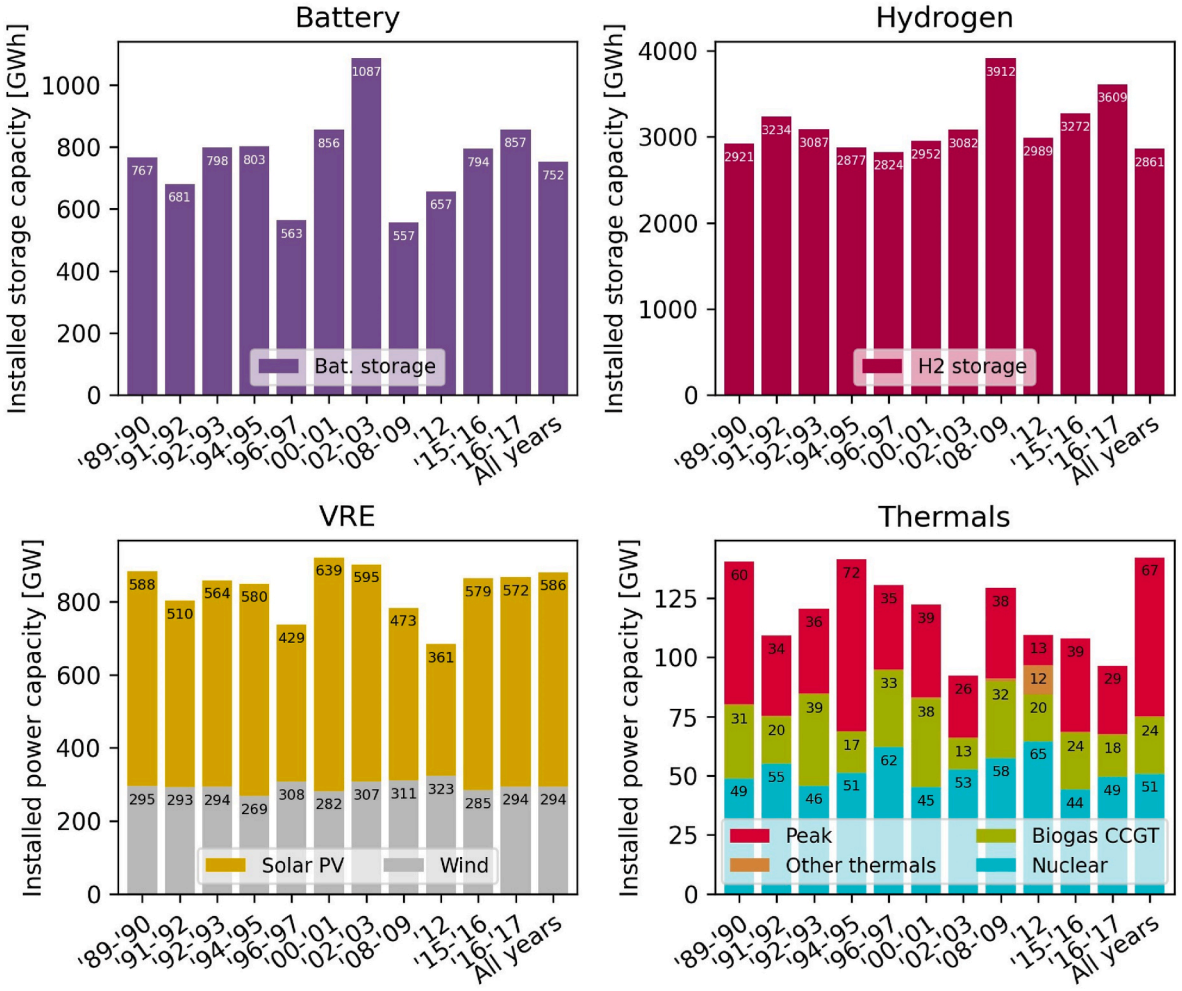


Fig. 4. Installed capacities in batteries, hydrogen storage units, VRE, and thermal technologies for a selection of single years, as well as the All years case. The shared capacity in "Other thermals" is pre-existing fossil-fueled power plants with technical lifetimes that extend beyond Year 2050.

Table 5
Biogas usage levels for the weather-years featured in Fig. 4, as well as the mean biogas usage level in the All years model run. Weather-years 1980–1981 and 2007–2008, the years with the lowest and highest levels of biogas use, respectively, are also included for reference. The years 1989–1990 and 1994–1995 are highlighted in bold.

	Biogas use [TWh]	Difference relative to *All years* [%]
Mean of *All years*	68	
1989–1990	**79**	**+16**
1991–1992	53	−22
1992–1993	65	−5
1994–1995	**52**	**−23**
1996–1997	82	+21
2000–2001	76	+12
2002–2003	37	−46
2008–2009	83	+21
2012	71	+4
2015–2016	57	−16
2016–2017	48	−30
1980–1981 (min)	25	−64
2007–2008 (max)	93	+36

those shown in Fig. 4, as well as 1980–1981 and 2007–2008, which have the lowest and highest biogas consumption levels of all the years, respectively. The mean biogas usage in *All years* (68 TWh) and the difference in biogas usage between *All years* and each year are also shown. The table shows that the biogas usage levels of individually modeled weather-years varies between 25 TWh and 93 TWh, with the levels in 1989–1990 and 1994–1995 being 79 TWh and 52 TWh, respectively. These differences between *All years*, 1989–1990 and 1994–1995 indicate that similarities in terms of overall installed capacity do not necessarily translate into similar biogas usage levels, though the differences are limited when compared to the extremes (1980–1981 and 2007–2008). A comparison of the biogas usage levels listed in Table 5 and the biogas CCGT capacities shown in Fig. 4 reveals that they generally follow each other, albeit more closely in some cases (e.g., 1996–1997 and 2008–2009) than in others (e.g., 1992–1993 and 2000–2001). Therefore, a good representation of the installed biogas capacity does not guarantee a good representation of the yearly biogas usage level. Indeed, the installed biogas CCGT capacity for 2007–2008 (the weather-year with the highest biogas consumption) is lower than the capacities of 1992–1993 and 2000–2001, at 35 GW.

It should be noted that the differences in biogas usage between individually modeled weather-years are smaller than the differences in biogas usage between years in the *All years* case, where a single capacity mix is deployed for all weather-years. In *All years*, the lowest yearly biogas usage is 25 TWh for 2013–2014, and the highest yearly biogas usage is 136 TWh for 1995–1996.

3.2.2. *Evaluating weather-year sets*

Figs. 5 and 6 show the capacity mixes for the sets with and without

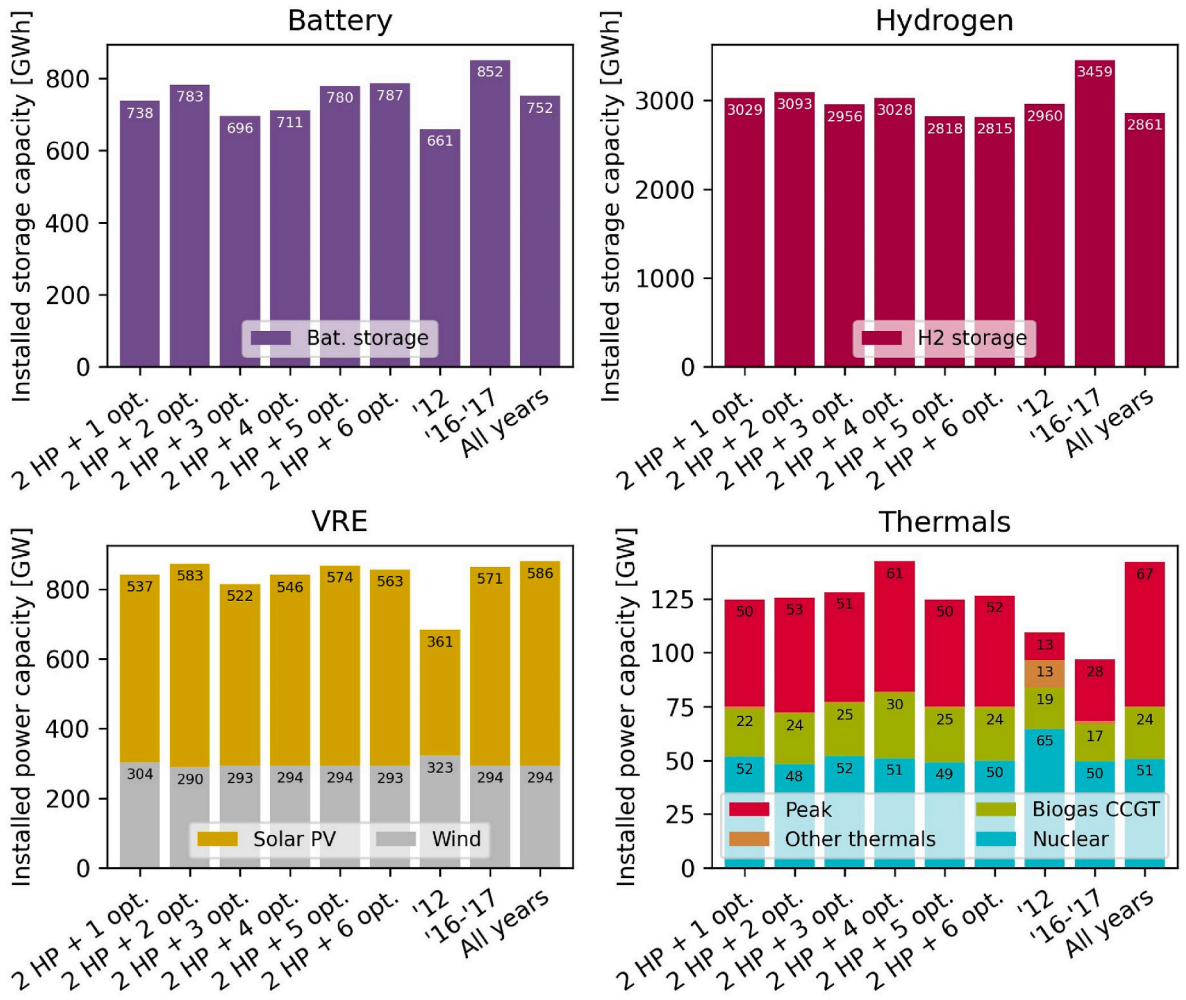


Fig. 5. Installed capacities in batteries, hydrogen storage units, VRE, and thermal power plants as obtained from the modeling in Step 1 (see Fig. 1). The category bars for "Other thermals" represent combined heat and power plants and other pre-existing power plants with technical lifetimes that extend beyond Year 2050.

hand-picked years, respectively. The two weather-years used to initiate the set selection algorithm, 2012 and 2016–2017, as well as *All years*, are shown for reference. The installed capacity mixes given by the sets of weather-years generally vary to a lesser extent than those for individual years, and they bear a closer resemblance to the capacity mix of *All years*. There is no trend or improvement apparent in Fig. 5 when additional years are added (in addition to the three years in *2 HP + 1 opt.*) to the sets, and the main difference observed between the weather-year sets is the higher capacities of peak and biogas CCGT generation in *2 HP + 4 opt.* (closer to that of *All years*). Closer inspection of the years included in *2 HP + 4 opt.* reveals that the difference is due to weather-year 1994–1995, which also features a high thermal capacity when investigated in isolation (as illustrated in Fig. 4). This increased thermal capacity is not seen for *2 HP + 5 opt.* or *2 HP + 6 opt.*

The installed capacities for the sets without hand-picked extreme years (all-optimized sets) are shown in Fig. 6, along with *2 HP + 1 opt.*, *2 HP + 2 opt.*, *2 HP + 3 opt., and 2 HP + 4 opt.*, and *All years* for reference. As shown in the figure, the differences between the all-optimized sets and the sets with hand-picked years are negligible. Especially similar are the sets with the same number of optimized years (e.g., *2 opt.* and *2 HP + 2 opt.*), which indicates that there may be little value in including hand-picked years, at least in terms of the resulting energy system composition.

In Table 6, the yearly mean biogas usage levels are listed for all the weather-year sets that contain five or fewer years discovered through optimization. The mean biogas usage level in *All years* is also shown, along with the difference in percent between *All years* and each weather-year set. This shows that while there are differences between *All years* and the sets, the differences are largest for *2 HP + 1 opt.*, which features only one weather-year that was not hand-picked for its challenging net-load events, and for the sets *4 opt.* and *2 HP + 4 opt.*, both of which feature the same optimized weather-years. The optimized weather-years in *4 opt.* and *2 HP + 4 opt.* include 2008–2009, which is the weather-year with the third-highest biogas usage when modeled in isolation, as well as three weather-years that lie closer to the average. While the combination of these weather-years obviously does not result in biogas usage that is equal to the mean for *All years*, the deviation is limited by the inclusion of multiple weather-years.

Additional methodological evaluation of the results in Appendix C indicates that the set-finding algorithm results in nuclear power-favoring sets (compared to *All years*) when using three or fewer weather-years identified through optimization. However, the method used in this work for resolving back-and-forth looping between candidate weather-year sets successfully picks the candidate set that most closely resembles *All years*. Appendix C also includes the results for an alternate version of *2 HP + 2 opt.*, with equal weights assigned to the weather-years, showing that while this affects the system composition, the exact weights used (when including four or more weather-years) do not appear to be of great importance.

Fig. 7 gives the distributions of the normalized installed capacities for individual weather years (top), random sets of 3–6 years, and sets of 3–6 weather years that were selected using the algorithm proposed in

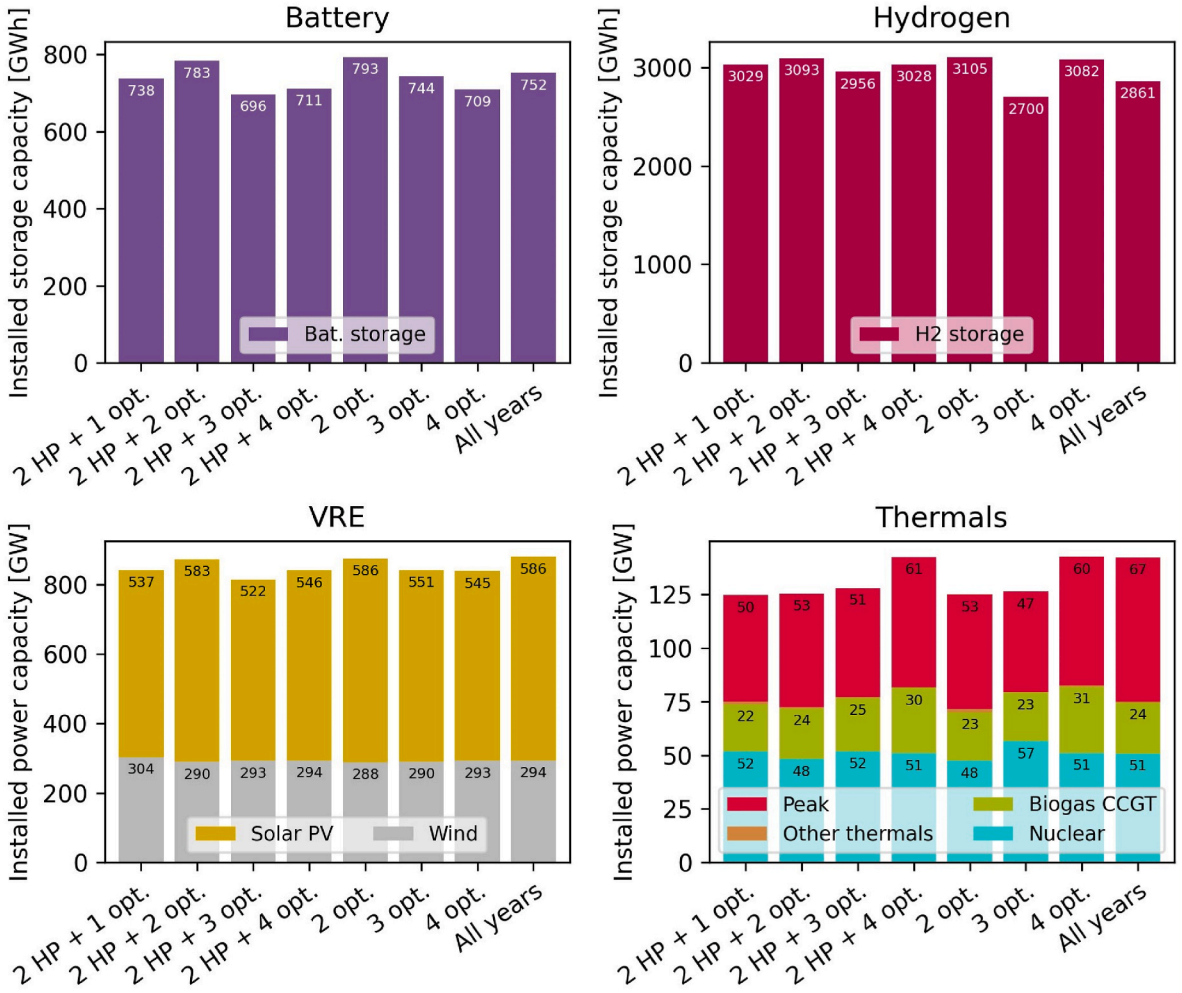


Fig. 6. Installed capacities in batteries, hydrogen storage units, VRE, and thermal technologies. Four sets with hand-picked years are shown, together with three all-optimized sets and the All years case. The shared capacity in "Other thermals" is for pre-existing fossil-fueled power plants with technical lifetimes that extend beyond Year 2050.

Table 6
Yearly mean biogas usage levels for a selection of weather-year sets, as well as the yearly mean biogas usage in the All years model run.

	Biogas usage [TWh]	Difference relative to *All years* [%]
Mean of *All years*	68	
2 HP + 1 opt.	57	−16
2 HP + 2 opt.	68	0
2 HP + 3 opt.	64	−6
2 HP + 4 opt.	82	+21
2 HP + 5 opt.	64	−6
2 opt.	66	−3
3 opt.	65	−4
4 opt.	82	+21

this work (represented by triangles). As shown in the figure, the distribution of the installed capacities for individual weather-years is significant. Random sets of three weather-years typically give a much more homogenous result, i.e., the box representing the span of the 25th and 75th percentiles is much smaller, although outliers remain. Thus, there is a risk that the randomly selected set of years will give installed capacities that deviate significantly from those of the average three weather-year set. With sets of five or more random weather-years, the worst-performing sets yield installed capacities that are close to the mean. Using the algorithm proposed in the present study, sets of weather years can be identified that perform close to the mean of the randomly picked sets. Using the algorithm reduces the risk of choosing sets that result in installed capacities that differ significantly from that of *All years*. Moreover, a set of three weather-years is sufficient to obtain a good estimate of cost-optimal generation and storage capacities, peak power excepted. The algorithm and the sets of randomly selected years typically perform poorly when assessing installed peak capacity. To estimate accurately cost-efficient investments in peak capacity, the year with the highest net-load event must be included. This year can be identified in the net-load matrix proposed in Algorithm 1 if the net load is not subject to a 12-h rolling average.

3.3. Cost-efficient measures to manage challenging net-load events

In Fig. 8, the installed capacities are shown for *2 HP + 2 opt., 2 HP + 3 opt., 2 HP + 4 opt.*, as well as for the two commonly hand-picked extreme weather-years, 1996–1997 and 2002–2003, and the two starting-point weather-years used (2012 and 2016–2017). Fig. 8 shows that when 1996–1997 and 2002–2003 are modeled in isolation the measures taken to manage their challenging net-load events are fundamentally different than when 1996–1997 and 2002–2003 are part of the weather-year sets. In isolation, the weather-years 1996–1997 and 2002–2003 stand out as those with the second-lowest and highest battery capacities of all the weather-years, respectively. They also diverge in terms of generating capacity, with weather-year 1996–1997 having

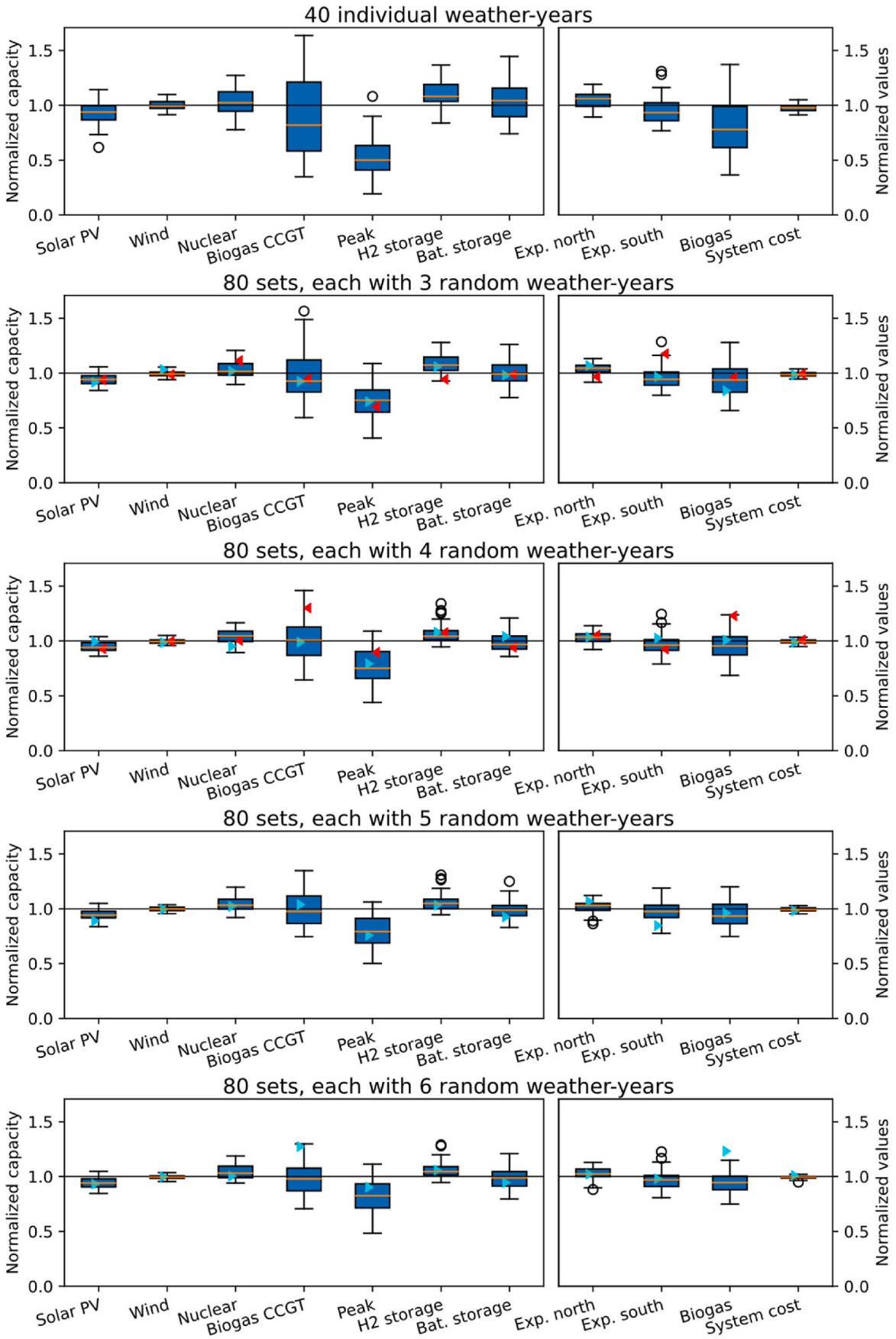


Fig. 7. Distributions of normalized installed capacities for individual weather-years (top), random sets of 3–6 years, and sets of 3–6 weather years, selected using the algorithm proposed in this work (cyan-colored triangles with hand-picked years and red-colored triangles without hand-picked years).

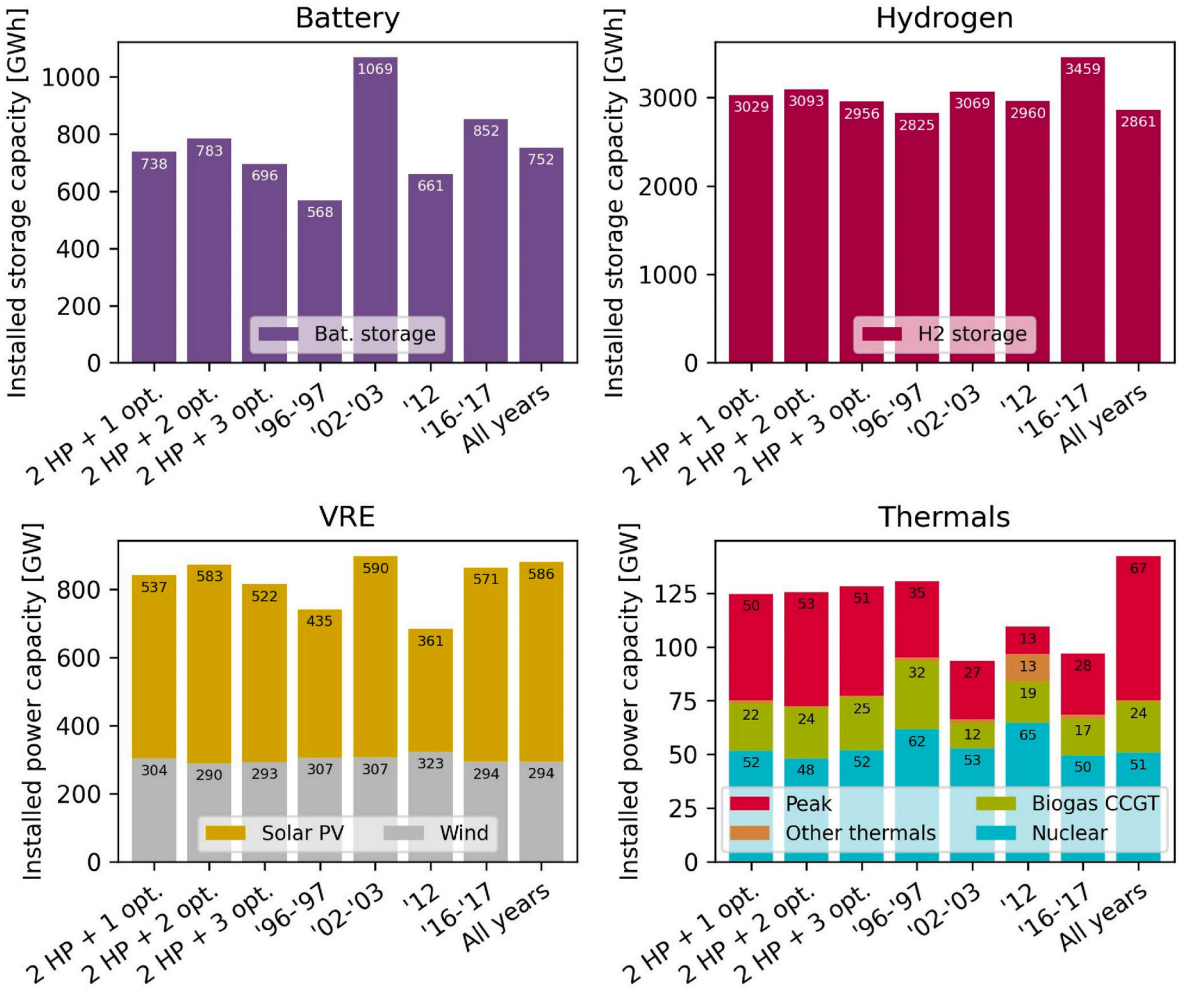


Fig. 8. Installed capacities in batteries, hydrogen storage units, VRE, and thermal technologies for three weather-year sets, two extreme hand-picked years, and the two used starting-point years. The shared capacity in "Other thermals" is for pre-existing fossil-fueled power plants with technical lifetimes that extend beyond Year 2050.

more nuclear power and mid-load thermal power, while weather-year 2002–2003 has more solar power compared to *All years*. The weather-year 2002–2003, which has one of the highest durations of 80–110 GW net load (the combined hydropower and nuclear capacity in *All years* being 83 GW), as well as one of the highest net-load peaks, is found to have the largest battery storage, battery power and solar PV capacities observed among the sets and weather-years included in Fig. 8. When modeled in isolation, the cheapest solution for the long periods of low-to-mid net load appears to be additional wind and solar power (coupled with increased battery storage). Similarly, the particularly high peak net load appears to be met not by additional thermal capacity but by additional battery capacity, which also complements the high solar PV capacity. For weather-year 1996–1997, which stands out as the year with the longest or almost longest durations of net-load events in the range of 120–240 GW (for a reference capacity mix; see Fig. 3), the ECE model gives the lowest battery investments and the second-lowest total VRE capacity. Instead, its technology mix contains one of the highest levels of thermal power of all the single weather-year model runs. However, the hydrogen storage level is about the same for the extreme years as for all the other model runs, which suggests that its dimensioning is driven by other factors or more commonly occurring variations.

The most cost-effective technology mixes during the extreme net-load events in weather-years 2002–2003 and 1996–1997 are, however, not the same when these extreme events are part of the representative sets. Fig. 8 shows that when the extreme events are integrated into the weather-year sets, and are assumed to occur only once every 40 years (each), their impacts on the system shift. Capacity-wise, the large battery capacity seen for 2002–2003 and the base-load and mid-load thermal capacities noted for 1996–1997 are replaced by peak turbines. As only the variable costs scale with the probability of the net-load event, low-probability events are expected to favor solutions that are associated with a low investment cost.

In Table 7, the average yearly system cost for each set of years is shown, along with the variable costs and biomass usage levels for a few selected weather-years. The biomass is used partially as solid biomass, and partially to produce biogas (with an assumed gasification efficiency of 70 %). The profile weather-years shown in Table 7 include: the two extreme weather-years (1996–1997 and 2002–2003); the two years used as initial years (2012 and 2016–2017); and the weather-year 1993–1994, which is part of multiple sets after being identified through the optimization process. While these indicators say nothing about the representativeness of the sets, they confirm that when weather-years with extreme net-load events are modeled in isolation (see 'Single year' rows highlighted in bold in Table 4) the variable costs and biomass usage levels are significantly lower than for the same weather-years when part of a weather-year set. This is because the solution for these extreme net-load events, when their low rate of recurrence is taken into consideration, relies on the use of biogas in gas turbines rather than investments in nuclear power or VRE.

Table 7
Yearly costs and biomass usage levels for each profile year in the investigated sets. The average yearly system cost, which is the minimization objective of the ECE model, consists of both the CAPEX and yearly OPEX costs, weighted for each included profile year. The variable costs for each year are given without the corresponding weights.

Weather-year	Set	Weight	Average yearly system cost [G€/yr]	Variable costs [G€]	Biomass usage [TWh]
1993–1994	**Single year**	100.0 %	80.2	8.7	82.4
	2 HP* + *3 opt.	27.6 %	81.7	10.3	104.1
	2 HP* + *5 opt.	13.6 %	83.9	9.5	100.2
	2 HP* + *6 opt.	11.7 %	85.6	9.8	103.6
1996–1997 Long mid-load durations	**Single year**	100.0 %	84.0	12.2	120.2
	2 HP* + *1 opt.	2.5 %	81.7	13.5	165.1
	2 HP* + *2 opt.	2.5 %	82.3	14.4	192.6
	2 HP* + *3 opt.	2.5 %	81.7	14.8	188.9
	2 HP* + *4 opt.	2.5 %	84.8	14.4	183.8
	2 HP* + *5 opt.	2.5 %	83.9	13.8	180.3
	2 HP* + *6 opt.	2.5 %	85.6	14.2	185.8
2002–2003 Long low-load durations + peak	**Single year**	100.0 %	81.9	7.1	55.1
	2 HP* + *1 opt.	2.5 %	81.7	16.3	214.4
	2 HP* + *2 opt.	2.5 %	82.3	15.6	206.1
	2 HP* + *4 opt.	2.5 %	84.8	16.1	210.0
	2 HP* + *5 opt.	2.5 %	83.9	13.7	170.6
	2 HP* + *6 opt.	2.5 %	85.6	14.2	177.8
2012	**Single year**	100.0 %	84.6	10.3	109.8
2016–2017	**Single year**	100.0 %	78.6	7.9	72.2
	2 HP* + *2 opt.	59.0 %	82.3	8.4	85.3

With gas turbines being the preferred strategy to manage net-load events of long duration that occur rarely, it is relevant to investigate the size of the fuel storage required to supply these gas turbines. As Table 7 shows, there is large variability in biomass usage between the modeled weather-years. While the amounts of hydrogen and battery storage are explicitly found in the ECE model results, the cost-optimal biogas storage and production capacities are not. Analyzing the results for *All years* reveals an average annual biogas consumption of 68 TWh, with the minimum being 25 TWh in 2013–2014 and the maximum being 136 TWh in 1995–1996. Assuming that the biogas production capacity is expensive and that the storage capacity is available at a relatively low cost, the production of biogas can be estimated based on the average biogas consumption rate (68 TWh/year). This constant biogas production (or storage refill) rate can then be combined with the hourly consumption to produce a theoretical storage level curve. In Fig. 9, the storage level curve reveals that a 235-TWh storage facility would be sufficient for the modeled electricity system in northern Europe if all the studied weather-years were connected in consecutive order. It is noteworthy that this is lower than the biogas demands for the two winters of Years 1995–1997, which total 262 TWh, due to refilling of the storage. However, the storage level curve also shows that 70 TWh of storage

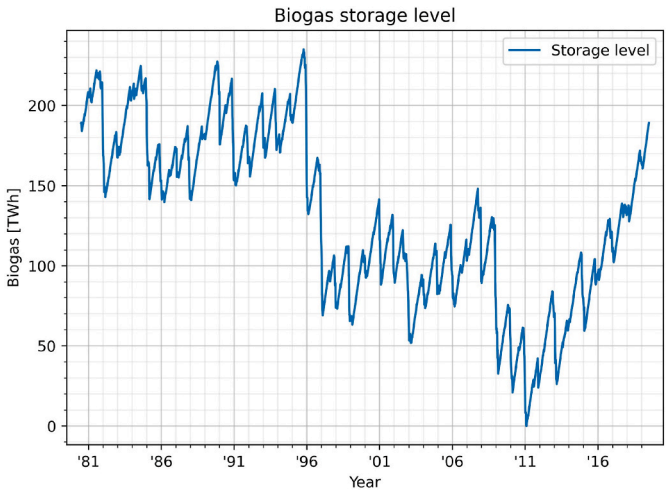


Fig. 9. A biogas storage level curve that assumes a constant refill rate of 68 TWh per year.

capacity can be avoided if the refill rate is increased between Years 1997 and 2008 (by 6.4 TWh/year or 9 %). This would reduce the total storage capacity requirement to 165 TWh for the electricity and heat supply in northern Europe for the 39 weather-years investigated. For reference, the current (as of March 2023) natural gas storage capacities are 250 TWh in Germany and 139 TWh in The Netherlands [35].

4. Discussion

The usefulness of simultaneously modeling a set of multiple years is not limited to predicting more accurately a generation and storage capacity mix. When studying a specific dynamic interaction or technology value, the use of multiple modeled years means that a broader span of net-load conditions can be tested. Moreover, this is done such that the effects experienced in one year affect the system for all years. Extreme net-load events can also be represented with their appropriate probabilities, which may be crucial for solving some research questions. The results of the present work show that modeling single years may significantly misrepresent the cost-optimal investments for nuclear power, wind power, solar PV, biogas OCGT and CCGT, batteries and hydrogen storage.

In the analysis of the ECE model results, it is found that annual variations, including those represented by the most-extreme weather-years, do not negatively influence the share of electricity coming from renewable technologies. The primary observable effects when comparing the representative sets with the individual years are increases in the peak-load and mid-load thermal capacities. The investment levels seen in the sets for other technologies remain close to the average of the individual years (although not close to any single year in all aspects). The technologies that contribute to periods of energy shortages, when modeling extreme years in isolation, are found to be additional wind, solar, or base-load and mid-load power capacities. However, when these unfavorable years form part of a set and their low rate of recurrence is accounted for, there is a shift towards increased utilization of thermal generation, which has a low investment cost but high running cost, to cover the shortages. While this suggests that increased utilization of wind, solar or nuclear power may not be a cost-efficient solution for interannual variations of wind and solar resources, it also suggests that fuel storage management is crucial to ensuring a reliable and stable carbon-neutral electricity system. Still, estimates of the required biogas storage (not considering the transportation of biogas) indicate that the current natural gas storage facilities in northern Europe are sufficient to cover the future projected biogas storage demand. However, these estimates were derived under the assumption that future weather-years

will follow the same chronologic order as past weather-years, while the chronology between weather-years is not accounted for in the algorithm proposed in this work. Thus, this estimate should be regarded with caution and further validated in future work.

As mentioned in the *Introduction* section, Ruhnau and Qvist [8] have investigated the need for long-term storage units in a future renewable electricity system in Germany based on 35 weather-years. These authors have reported that a large hydrogen storage unit (36 TWh of electricity output) is required to manage the largest energy deficiencies during the 35-year period. They have also shown that modeling using individual years generally underestimates the required storage capacity. For comparison, the hydrogen storage capacities for northern Europe found in the present work are in the range of 2.8–3.5 TWh. However, there are substantial differences between the studies with respect to the representation of flexibility. By excluding consumer flexibility and international trade, while adopting a limited bioenergy capacity, Ruhnau and Qvist have discarded many of the potential alternatives to hydrogen storage systems. While the use of hydrogen storage instead of biogas storage does not affect the conclusion drawn by Ruhnau and Qvist that the role of storage may include smoothening of interannual variability, it would increase the costs if the existing natural gas infrastructure was not to be used where possible. Still, both their study and the present work show that focusing on single years can lead to misrepresentation of the storage (and generating) capacities.

4.1. Further reflections on the methodology

An important aspect of the algorithm proposed in this work to identify weather-year sets is the qualities that it seeks to represent in its weather-year selection. The algorithm minimizes the difference in net-load characteristics, although this difference is evaluated in terms of the technology mix obtained when the set is applied in an ECE model. However, since the net-load characteristics are vital for dimensioning the technology mix and the value of flexibly dispatchable generation, representation of the former (net-load characteristics) should lead to a good, if not optimal, representation of the latter. As the results show for the individual weather-year 1989–1990, specific individual years may yield storage wind, solar PV and total thermal capacities that are closer to those of *All years* than those generally found for the weather-year sets. This does not, however, mean that an individual weather-year also represents all weather-years with respect to other aspects of interest in the ECE model. The results show, for the example of biogas usage, that the yearly consumption level for weather-year 1989–1990 is close to that of *All years* (79 TWh versus 68 TWh), whereas modeling of only one year gives no indication as to the biogas usage variability between years. In addition, it is not clear by which metrics a representative individual weather-year (such as 1989–1990) can be identified without access to a model run that features all years simultaneously. As the results regarding the installed capacities of individually modeled years show, neither weather-year 2012 nor weather-year 2016–2017 yields capacity levels similar to those of *All years* despite Collins et al. [5] finding 2012 to be one of the most-representative weather-years for the European power system.

While it seems intuitive that including hand-picked extreme years would improve the representativeness of the weather-year sets, the results (in Fig. 6) do not show any significant improvement in the capacity mix, as compared with *All years* when including hand-picked years. In addition, the selection of handpicked years adds subjectivity to the algorithm as it entails the analyst deciding which combination of net-load duration and amplitude is extreme and warrants inclusion. Still, there are some qualitative arguments that should be considered. From an energy system modeler's point-of-view, explicitly ensuring that years with extreme net-load events are included may be of importance. Another reason to hand-pick extreme years lies in the assigned weights. If they are not hand-picked, and instead covered according to the years chosen by the optimization, the weights for the weather-years with extreme net-load events will be based also on the non-extreme parts of the year, which will likely result in an over-representation of the extreme events included. Extreme events may of course occur more or less often than once every 40 years. The important thing is, however, to capture (as accurately as possible) that these events are in fact unusual. On the one hand, as shown in Appendix C, increasing the weights of the extreme years from 1/40 to 1/4 shifts the balance in terms of cost-optimal electricity generation sources towards favoring nuclear power over solar power. On the other hand, if extreme weather-years are not hand-picked, an error summation function that overly emphasizes high values (such as the squared error summation) may completely disregard the low-recurrence regions. While this work does not compare weather-year sets discovered using different error summation functions without hand-picked extreme years, a comparison is made between the sets with and without hand-picked years (in Fig. 6). This comparison indicates that absolute value summation (without hand-picked years) results in weather-year sets with similar technology mixes as that for *All years,* though the nuclear power capacity is overestimated in *3 opt.* and the peak power capacity is generally underestimated regardless of the error summation function. This similarity in investment levels with and without hand-picked extreme events suggests that extreme events may not be critical for a time-series to be representative. However, it should be stressed that the importance of these events depends on the research question in focus. If, for example, access to biogas is limited on a yearly basis, the inclusion of extreme net-load events becomes more relevant.

One weakness of the algorithm used here to identify sets of weather-years is the need to break the loop when two or more different candidates end up in a cycle and the algorithm fails to find a set that is self-representative (i.e., representative of its own net-load fingerprint). While the method chosen in the implementation in this work results in sets with capacity mixes that resemble each other and that of the *All years* scenario, there is no guarantee that this would turn out to be the case. Similarly, while the algorithm finds sets that are self-representative, there is no guarantee that these sets are the best representative sets of all weather-years. Yet, the results of this study indicate that the sets are highly representative of the set that consists of all 39 weather-years, with the exception of a peak power capacity shortage. If modeling all available weather-years is possible in some reference scenario, the aforementioned potential weaknesses can be avoided. By using the capacity mix of the model run with all weather-years as input into Step 2 of the algorithm (as illustrated in Fig. 1), the set found in Step 3 is, by definition, the set that is most-representative (in terms of net load amplitude, duration and number of occurrences) of all weather-years. The potential peak power shortage can also be alleviated by identifying the hour with the highest net load (for a given capacity mix) and simply adding peak capacity until sufficient dispatchable capacity exists to supply the worst hour. For example, performing this comparison for *2 HP + 3 opt.* indicates that 17 GW of peak capacity should be added, which would bring its capacity in line with that of *All years* (though some dispatchable capacity differences remain due to the differences in VRE capacity and, thus, the net load).

Another potential weakness of the proposed set-finding algorithm is that it neither explicitly includes nor necessarily values representation of weather-years with good wind and solar production levels. In the duration-amplitude-recurrence matrix presented in Fig. 3, years with average and poor wind and solar resources will have values with higher amplitudes and durations than years with good VRE resources. In the same way, years with good VRE resources show a higher recurrence of net loads with low amplitudes and durations. As such, while there is no explicit mechanism for including years with good VRE resources, the objective of minimizing the difference in net-load matrices should indirectly also value normal, if not good, weather-years. In the sets identified in this work, those with at least three optimized years include at least one year with above average wind full-load hours. Weather-years with below average wind full-load hours are found in all sets, both in those with hand-picked years and in those in which the years are

found through optimization.

A notable detail of the algorithm as implemented in this work is the use of a single net-load time-series per weather-year. The algorithm, as is, aggregates all the regional loads and non-dispatchable generation when building the net-load recurrence matrix, thereby neglecting all transmission bottlenecks. This is done to avoid an additional regional dimension in the matrix, without which the visual analysis and set-building optimization are considerably easier tasks. However, since it is possible that hours of dimensioning regional net-load are hidden by the regional aggregation, regional net-load recurrence matrices may better capture the need for peak thermal capacity. However, this may instead fail to capture net-load events that span multiple regions.

While not shown in this work, the weather-year selection algorithm proposed herein could be used together with other methods to reduce the temporal scope, such as representative days or chronologic time clustering to reduce the upscaling of ECE problems capturing interannual variability.

5. Conclusions

This work investigates the impacts of net-load events that have unusually high amplitudes and long durations on the modeled cost-optimal electricity composition in northern Europe. It is found that when the probability of occurrence of such events is considered, net loads of extreme amplitude and duration mainly lead to increases in the thermal peak load capacity (in this work, represented by open-cycle biogas turbines). The size of the biogas fuel storage required to feed the turbines is also dimensioned by interannual weather variability. Post-modeling calculations indicate that current natural gas storage units in Germany and The Netherlands may cover the modeled future demand for biogas storage in northern Europe.

In order to include net-load events of unusually high amplitude and duration in an electricity system capacity expansion model, this work proposes and evaluates an algorithm that identifies combinations (sets) of weather-years that represent a broader range of weather-years. In the set-finding algorithm, the representativeness of the sets of weather-years is evaluated in terms of the net-load variations that an electricity generation system would experience. The algorithm is applied to find weather-year sets that contain 3–8 of the 40 years of weather data. The weather-year sets include weather-years that are identified using an error-minimizing optimization system, and some sets also include years that are hand-picked for their extreme net-load amplitudes and durations.

Evaluation of the obtained weather-year sets is carried out by comparing the results from an ECE model using the algorithm-based sets, sets of random years, individual years or all years at once (*All years*). By examining the cost-optimized generation and storage capacities, this study finds that while there are large differences in generation levels and storage capacities between individually modeled years, the differences between the weather-year sets are small in comparison. Furthermore, the capacity mixes for the weather-year sets are found to resemble closely the capacity mix of *All years*, with the exception of a lack of peak thermal power capacity.

Based on the results of this study, it is concluded that by applying the proposed weather-year set-finding algorithm, the risk associated with choosing a few weather-years to represent the variety of net-load events (including the most-challenging ones) experienced over a 40-year period is reduced, as compared to a random selection of weather years. This effect is found to be strongest for sets that consist of fewer than five weather-years. Furthermore, the authors recommend using the set-finding algorithm to identify sets with (at least) three optimized years and two hand-picked years. If a smaller set is necessary, removing the two hand-picked years may be preferable to using only one or two optimized years.

CRediT authorship contribution statement

Jonathan Ullmark: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lisa Göransson:** Writing – review & editing, Supervision, Conceptualization. **Filip Johnsson:** Writing – review & editing, Supervision, Funding acquisition.

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 study has been funded by the Swedish Energy Agency, Project number 44986-1.

Appendix A. Model equations

This section lists the core constraints of the ECE model used in this work, as well as the modifications that were made to allow for simultaneous optimization of multiple weather-years. The generating and storage technologies available for investment are also listed, along with their technical lifetimes, efficiency levels, investment costs, and running costs (where applicable). These techno-economic parameters are assumed to be valid for Year 2050. For thermal power plants, additional costs are applied at start-up and when operating at part-load. These costs are reflected in the objective function, and the underlying equations can be found elsewhere [26]. For most of the technologies, the techno-economic parameters are sourced from technology catalogues published by the Danish Energy Agency [36]. Since nuclear power is not included in the technology catalogues, its parameters are taken from the World Energy Outlook 2018 of the IEA [37]. As the ECE model is built to support various future model-years, the investment costs and efficiencies are further fitted to a polynomial S-curve, reflecting the learning-curve effect over time.

Table A1
List of technologies available for investment, as well as their investment and running costs, technical lifetimes and efficiency levels, in the ECE model. The running costs apply only when running at full load, as reduced efficiencies at part load increase the running costs. While the running costs already include the efficiencies, they are listed for transparency.

	Investment cost [€/kW(h)]	Fixed O&M* [€/kW/yr]	Running cost [€/MWh]	Lifetime [yr]	Efficiency [%]
Nuclear ST	3985	123	12	60	33
Biomass ST	1980	52	116	40	35
Biomass CHP	3260	106	136	40	30
Biogas CHP	1211	32	157	40	49
Biogas CCGT	900	17	127	30	61

(*continued on next page*)

Table A1 (*continued*)

	Investment cost [€/kW(h)]	Fixed O&M* [€/kW/yr]	Running cost [€/MWh]	Lifetime [yr]	Efficiency [%]
Biogas OCGT	450	15	184	30	42
Biogas CCS	1571	38	93	30	54
Heat pump	1000	7.9	–	25	300
Fuel cell	841	55	–	10	50
H_2 electrolyzer	395	18	–	20	79
Electric boiler	50	0	–	20	99
H_2 storage	11	0	–	40	88
Battery storage	79	0	–	25	–
Battery power	68	1	–	25	98
Stand-alone PV	266	7	–	40	–
Rooftop PV	377	11	–	40	–
Offshore wind	1531	36	–	30	–
Onshore wind	1149	13	–	30	–

*O&M: operation and maintenance.

The ECE model features a range of equations, sets, parameters and variables, of which the core ones are listed in Table A2 and featured in Eqs. (3)–(7) below. The main variables are investments in ($i_{r,p}$) and operation of ($g_{r,t,p,y}$) each technology (p), as well as electricity trade ($e^{netexport}_{r,r',t,y}$) between each interconnected region. The operation variable ($g_{r,t,p,y}$) represents the electricity output for the generating technologies and the storage level for the storage technologies. For the VRE technologies, the electricity generation is added directly as the capacity level multiplied by the generation profile ($X_{t,p,y}$). In this work, the model has been extended by adding the weather-year set (Y) to the operation-related variables and some parameters. In the objective function, the costs for each year are multiplied by that year's weight and then summed.

Table A2
Descriptions of the sets, variables and parameters used in the mathematical description of the electricity system optimization model used in this work.

Sets		
R	Subregions	
T	Time-step, {1,..,8784}	
P	Technology	
P^{VRE}	Variable renewable technologies (wind, solar and run-of-river)	
P^{ESS}	Energy storage technologies (Li-ion batteries, hydrogen storage)	
P^{charge}	Energy storage charging technologies	
$P^{discharge}$	Energy storage discharging technologies	
$P^{thermal}$	Thermal power plant technologies	
P^{gen}	Electricity-generating technologies (excluding VRE)	
Q	$P^{ESS} \times P^{charge} \times P^{discharge}$ matrix connecting energy storage units with their respective charging capacity technologies (e.g., electrolyzer, inverter) and discharge capacities	
Y	Weather-year	
Variables		
$e^{netexport}_{r,r',t,y}$	Electricity net export from region r to region r' during time-step t and year y	[GWh/h]
$i_{r,p}$	Investment in technology p in region r	[GW]
$g_{r,t,p,y}$	Generation, or storage level, for technology p at time-step t in region r for year y	[GWh/h]
$g^{start}_{r,t,p,y}$	Start-up of thermal technology p at time-step t in region r for year y	[GW]
$g^{active}_{r,t,p,y}$	Online capacity of thermal technology p at time-step t in region r for year y	[GW]
$s^{charge}_{r,t,p,y}$	Charging of storage p in region r at time-step t for year y	[GWh/h]
$s^{discharge}_{r,t,p,y}$	Discharging of storage p in region r at time-step t for year y	[GWh/h]
Parameters		
η^{ESS}_{p}	Charging and discharging efficiency of technology p	[−]
C^{inv}_{p}	Investment cost for technology p	[k€/GW]
C^{OPEX}_{p}	Running cost (including fuel, CO_2 and variable O&M costs) for technology p	[k€/GWh]
C^{fixOM}_{p}	Fixed yearly O&M cost for technology p	
C^{start}_{p}	Start-up cost for technology p	[k€/GW]
C^{part}_{p}	Part-load cost for technology p	[k€/GW]
S^{rate}_{p}	Storage (dis)charge rate as a fraction of storage per hour	[−]
$W_{r,t,p,y}$	Hourly profile for VRE p at time t in region r for year y	
X_y	Weight/probability of weather-year y	[−]

The objective function to be minimized is the sum of the costs for investments (C^{inv}_p) and operation for each technology. The operation-related costs consist of fuel use and variable O&M costs (C^{OPEX}_p), part-load costs (C^{part}_p), and start-up costs (C^{start}_p).

$$\text{Min} \sum_{r,y \in R,Y} W_y * \left(\sum_{p \in P} i_{r,p} * C^{inv}_p + \sum_{p \in P} i_{r,p} * C^{fixOM}_p + \sum_{t,p \in T,P^{gen}} \left(g_{r,t,p,y} * C^{OPEX}_p \right) + \sum_{t,p \in T,P^{gen}} \left(\left(g^{active}_{r,t,p,y} - g_{r,t,p,y} \right) * C^{part}_p \right) + \sum_{t,p \in T,P^{gen}} \left(g^{start}_{r,t,p,y} * C^{start}_p \right) \right) \quad (3)$$

Storage balance is maintained through the charge ($s^{charge}_{r,t,p}$) and discharge ($s^{discharge}_{r,t,p}$) variables featured in the storage balance equation [Eq. (4)]. The charge and discharge variables are, in turn, limited by the (dis)charge capacity investment ($i_{r,q^{in/out}}$) and the technical (dis)charge rate limit (S^{rate}_p) of the storage. It should be noted that for hydrogen storage units, the charge and discharge variables are the same as the generation variables for electrolyzers and fuel cells, respectively.

$$g_{r,t+1,p,y} = g_{r,t,p,y} + s^{charge}_{r,t,p,y} * \eta^{ESS}_p - s^{discharge}_{r,t,p,y} \Big/ \eta^{ESS}_p, \forall r,t,p,y \in R,T,P^{ESS},Y \quad (4)$$

$$-\left(i_{r,p}\right)^{*}S_{p}^{rate} \leq -i_{r,q^{in}} \leq s_{r,t,p,y}^{discharge} - s_{r,t,p,y}^{charge} \leq i_{r,q^{out}} \leq \left(i_{r,p}\right)^{*}S_{p}^{rate}, \forall r,t,\left(p,q^{in},q^{out}\right),\mathrm{y} \in R,T,Q,Y \tag{5}$$

The variables $g_{r,t,p,y}^{active}$ and $g_{r,t,p,y}^{start}$, used to incur thermal cycling costs in Eq. (3), are connected to the operation of thermal power plants through Eqs. (6) and (7):

$$g_{r,t,p,y} \leq g_{r,t,p,y}^{active}, \forall r,t,p,y \in R,T,P^{thermal},Y \tag{6}$$

$$g_{r,t,p,y}^{start} \geq g_{r,t,p,y}^{active} - g_{r,t-1,p,y}^{active}, \forall r,t,\mathrm{p},\mathrm{y} \in R,T,P^{thermal},Y \tag{7}$$

The ECE model features the direct demands for electricity, DH and hydrogen. The demand for hydrogen is implemented as a drain from the storage equation, while the demands for electricity and are implemented as separate balances [see Eqs. (8) and (9)]. In the DH balance, generation and demand are aggregated across 26 2-week periods (O), representing an assumption that there is some flexibility in the DH network from storage (not explicitly modeled).

$$d_{r,t,y}^{elec,traditional} + d_{r,t}^{elec,heat} + d_{r,t}^{elec,transport} + d_{r,t}^{elec,steel} + d_{r,t}^{elec,cement} + d_{r,t}^{elec,ammonia} + d_{r,t}^{elec,batprod} \leq \sum_{p^{gen}} g_{r,t,p,y} + \sum_{p^{VRE}} i_{r,p}{}^{*}W_{r,t,p} - \sum_{p^{PtH}} g_{r,t,p,y} \Big/ \eta_{p} + \sum_{p^{store}} \left(s_{r,t,p,y}^{discharge} - s_{r,t,p,y}^{charge}\right)$$

$$+ \sum_{r'} e_{r,r',t,y}^{netexport}, \forall r,t,y \in R,T,Y \tag{8}$$

$$\sum_{t \in o} d_{r,t,y}^{DH,heat} \leq \sum_{t \in o} \left(\sum_{p^{CHP}} \left(g_{r,t,p,y}{}^{*}\alpha_{p}\right) + \sum_{p^{PtH}} g_{r,t,p,y} \right), \forall r,o,y \in R,O,Y \tag{9}$$

Appendix B. Additional results from the set-finding algorithm

In Figure A1, the combination with the least error ($\sum_i M_i{}^{*}w_i$), the resulting difference matrix (M_{diff}), and the error matrix ($\sum\left(\left|M_{diff}\right|\right)$) are illustrated as heatmaps, to indicate what the error and difference matrices look like. In the case shown, the weather-years 2002–2003 and 1996–1997 have been predetermined as extreme weather-years and, thus, each is given a weight of 1/40. The third weather-year (2014–2015) is that with the least squared error when combined with the two hand-picked years.

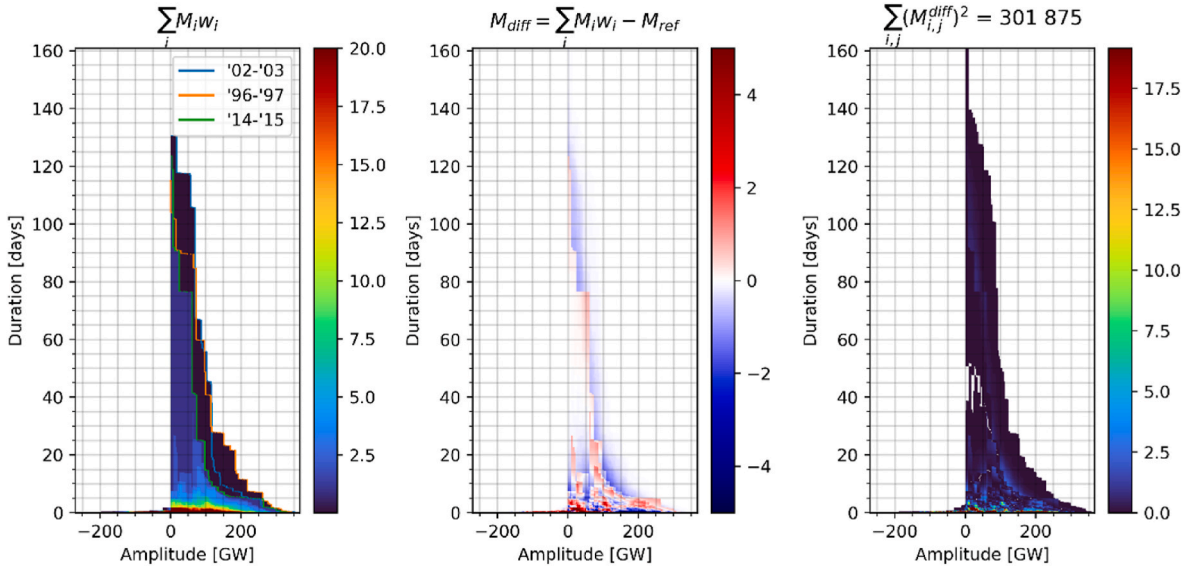


Fig. A1. Three heatmaps showing, for the combination of weather-years 1996–1997, 2002–2003 and 2014–2015, in order from left to right: the combined net-load recurrence matrix; the difference matrix M_{diff} [defined in Eq. (1)]; and the error matrix whose sum is to be minimized.

As the weights represent only a fraction of the 40 years evaluated, the weights must be between 0 and 1, and the sum of the weights must equal 1. Thus, the optimization has a very limited solution space but is non-linear, and the large size of M_{diff} is a complicating factor. The solution space for three variables between 0 and 1 that when summed must equal 1 can be represented by a triangle (part of a plane in a 3D space), with the corners of the triangle at the coordinates: (1,0,0), (0,1,0) and (0,0,1). This solution space, and the error at each point of the space, is illustrated in Figure A2 using a heatmap for an arbitrary set of weather-years. It shows what the non-linear solution space looks like. As can be seen in the figure, while the problem may be non-linear, there is only one minimum in the case with three weather-years to optimize. While this does not prove that there is not more than one local optimum for the other cases, it suggests how the solution space looks and how it differs between error summation functions. It should also be noted that the squared error version for single combinations is a quadratically constrained problem, which means that it has only one optimum. For the case in Figure A2, the choice of error summation function has only a weak impact on the position of the optimal position in the solution space, with the squared error function showing the greatest divergence. The same result has been found for all cases analyzed in this way by the authors.

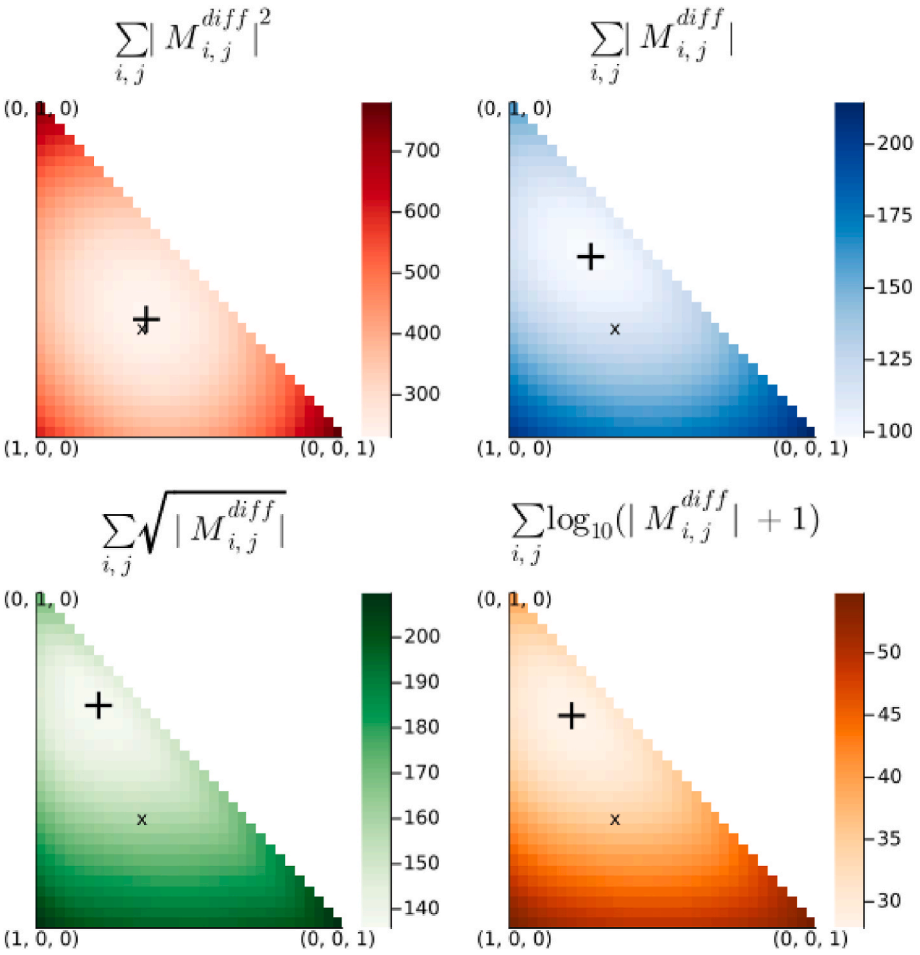


Fig. A2. Heatmaps showing four different strategies for calculating the error for each point in the solution space for combinations of three arbitrary years. The large '+' symbol indicates where the lowest error is found, and the middle-point at which all three years are weighed equally is indicated by an 'x'. Note that the error has been scaled by a factor of 1/1000 for better readability.

In this work, the optimization is run with multi-threading across four threads for 2–6 min per combination, depending on the numbers of undetermined weights and combinations being considered. For the seven heuristic optimization algorithms tested, particle swarm optimization, probabilistic descent, and variants of DE/rand/1/bin are found to be the most-effective at finding the minimum. To reduce the total time needed to find the best weights and best combination, the script is set up to evaluate initially only a few (5–16, depending on the size of the set) predetermined points in the solution space. These predetermined points include the midpoint of the solution space, as well as a few symmetrically spread positions. As the optimization objective improves very little (<5 %) between the initial search and the longer optimization, the vast majority of combinations can be eliminated using a brief search. The combination-optimizing step has also been formulated as a MIQCP (Mixed Integer Quadratically Constrained Program) for when the error is calculated using squared error summation. This MIQCP model finds the optimal solution for even sets of six optimized years in only minutes.

Appendix C. Extra methodology evaluation results

In Figure C1, the installed capacities are shown for *2 HP + 1 opt.*, *2 HP + 1 opt. (2012 start)*, the two versions of *2 HP + 2 opt.* that were shown to alternate back-and-forth in the iterations in Table 4 [the version that did not win the tie-breaker labeled as *2 HP + 2 opt. (b)*], as well as the extra set *2 HP + 2 opt., eq. w.* The two weather-years used as starting points in the fingerprinting algorithm are also included for reference. The figure shows substantial differences between *2 HP + 1 opt.* and *2 HP + 1 opt. (2012 start)*, with the latter having 12 GW more generation from nuclear power (all in the border-region DE_S) and less from VRE. A similar difference is seen between *2 HP + 2 opt.* and *2 HP + 2 opt. (b)*, with the latter again having more generation from nuclear power and less from VRE. The similarities between *2 HP + 1 opt. (2012 start)* and *2 HP + 2 opt. (b)* are observed also for Year 2012, and this indicates that the set-finding algorithm results in nuclear power-favored sets (compared to *All years*), depending on the specific weather-years contained in the set. When the weights for the years in *2 HP + 2 opt.* are equal, the representation of extreme weather-years increases, as does the amount of nuclear power, bringing the capacity mix closer to that of *2 HP + 2 opt. (b)*. However, the amount of wind power is also increased, and the change in overall system composition is not drastically different from the change in weights alone. As such, the exact weights used when including four or more weather-years does not appear to be of great importance, although using too few weather-years may result in sets that overestimate the value of nuclear power.

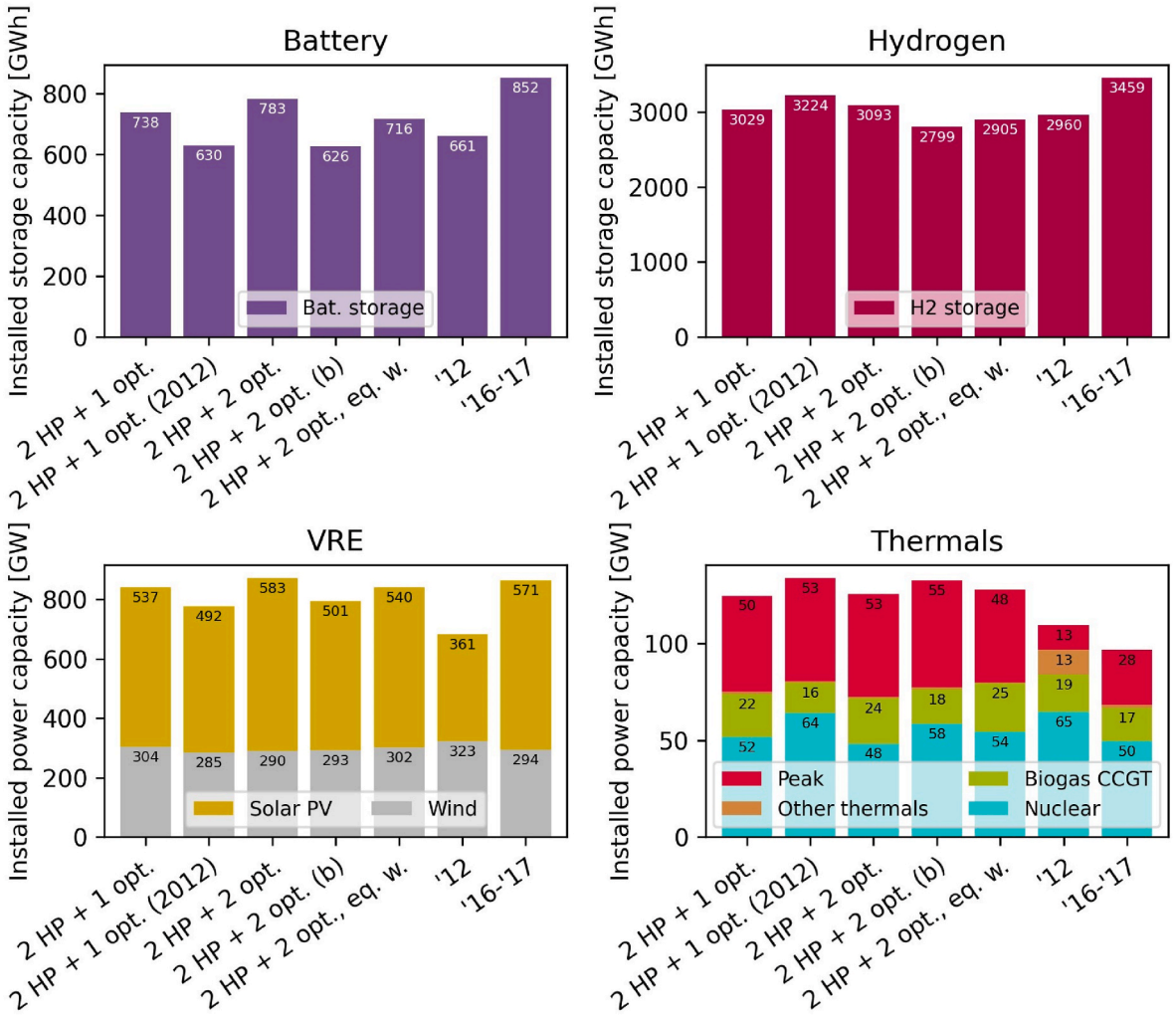


Fig. C1. Installed capacities in batteries, hydrogen storage units, VRE, and thermal technologies. The sets and years shown include those shown in Fig. 5, and additionally include a set with equal weights: 6 yr, eq. weights. The shared capacity in "Other thermals" represents pre-existing fossil-fueled power plants with technical lifetimes that extend beyond Year 2050.

Figure C2 shows the yearly electricity generation levels per technology for a number of weather-years (as annotated above the bars, e.g., 1993–1994). For each year, there are multiple columns labeled (on the *x*-axis) with either a set name, *All years* or *Single year*, depending on the model run from which the electricity generation data originate. A comparison of the results for the single weather-years of 1993–1994 and 2016–2017 with those for *All years* reveals little difference. However, for the extreme weather-years of 1996–1997 and 2002–2003, a difference can be seen in both bioenergy use (*Peak* + *Other thermals*) and the balance between electricity supplied from nuclear power and from VRE. This indicates that the differences from *All years* in terms of electricity supply are greatest for the years with more-challenging weather profiles. Some variance is also noted between the different sets, although the total difference relative to the *All years* generation level is small.

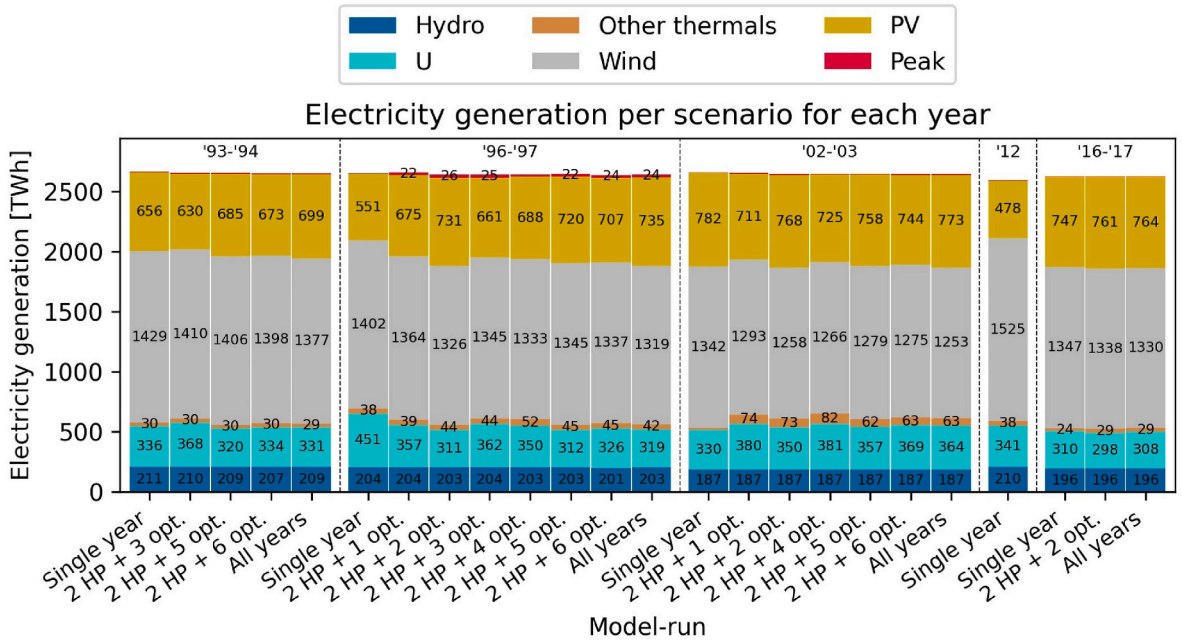


Fig. C2. Electricity generation levels for each of the weather-years in the optimized sets of weather-years. Each cluster of stacks represents 1 year. Each stack shows the electricity generation for that year in either a model run for only that year (labeled "Single year") or for a model run with a set of multiple years (labeled as "Set .").

Appendix D. Net-load duration profiles

In Figure D1, net-load duration curves are shown for weather-years 2012 and 2016–2017, as well as for weather-years 1989–1990 and 1994–1995, which are the years identified in Fig. 4 as those having the highest total thermal capacities when modeled in isolation, along with the average net-load duration for all the years. Again, the net load for each year is calculated using the *All years* capacity mix. Nonetheless, a large difference is seen for 1994–1995, in which the number of hours with ≥130 GW of net load is lower than for both the other years and the average. In particular, Figure D1 reveals that the net-load duration curve for 1989–1990, which was found in Fig. 4 to resemble that of *All years* not only in terms of storage, total thermal power and solar PV capacity (similar to 1994–1995), but also in terms of wind power capacity, shows a stronger resemblance to weather-year 2012 than to 1994–1995. This indicates that the net-load duration curve is a poor predictor of how well a weather-year will represent *All years* in terms of the technology mix.

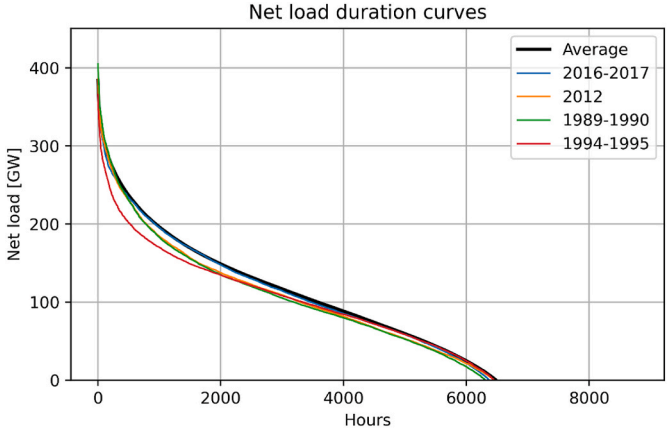


Fig. D1. Hourly net loads sorted in descending order for weather-years 1989–1990, 1994–1995, 2012 and 2016–2017, as well as for the average for all the years. The y-axis is cut off at 0 GW to highlight the differences that occur above 0 GW.

Data availability

I have made most input data publicly available. If there is data not available in the tables or the provided data repositories, I may be able to share it upon request.

References

[1] Mosca C, et al. Mitigation of frequency stability issues in low inertia power systems using synchronous compensators and battery energy storage systems. IET Gener Transm Distrib Sep. 2019;13(17):3951–9. https://doi.org/10.1049/iet-gtd.2018.7008.

[2] Göransson L, Johnsson F. A comparison of variation management strategies for wind power integration in different electricity system contexts. Wind Energy Oct. 2018;21(10):837–54. https://doi.org/10.1002/WE.2198.

[3] Ringkjøb HK, Haugan PM, Solbrekke IM. A review of modelling tools for energy and electricity systems with large shares of variable renewables. Renew Sustain Energy Rev Nov. 01, 2018;96:440–59. https://doi.org/10.1016/j.rser.2018.08.002. Pergamon.
[4] Bloomfield HC, Brayshaw DJ, Shaffrey LC, Coker PJ, Thornton HE. Quantifying the increasing sensitivity of power systems to climate variability. Environ Res Lett Dec. 2016;11(12):124025. https://doi.org/10.1088/1748-9326/11/12/124025.
[5] Collins S, Deane P, Ó Gallachóir B, Pfenninger S, Staffell I. Impacts of inter-annual wind and solar variations on the European power system. Joule Oct. 2018;2(10):2076–90. https://doi.org/10.1016/j.joule.2018.06.020.
[6] Zeyringer M, Price J, Fais B, Li P-H, Sharp E. Designing low-carbon power systems for Great Britain in 2050 that are robust to the spatiotemporal and inter-annual variability of weather. Nat Energy Apr. 2018;3(5):395–403. https://doi.org/10.1038/s41560-018-0128-x.
[7] Dowling JA, et al. Role of long-duration energy storage in variable renewable electricity systems. Joule Sep. 2020;4(9):1907–28. https://doi.org/10.1016/J.JOULE.2020.07.007/ATTACHMENT/CE7FB308-32A7-4A52-82DE-2E9757A3431B/MMC1.PDF.
[8] Ruhnau O, Qvist S. Storage requirements in a 100% renewable electricity system: extreme events and inter-annual variability. Environ Res Lett Mar. 2022;17(4):044018. https://doi.org/10.1088/1748-9326/AC4DC8.
[9] Teichgraeber H, Brandt AR. Time-series aggregation for the optimization of energy systems: goals, challenges, approaches, and opportunities. Renew Sustain Energy Rev Apr. 2022;157:111984. https://doi.org/10.1016/J.RSER.2021.111984.
[10] Pineda S, Morales JM. Chronological Time-Period Clustering for optimal capacity expansion planning with storage. IEEE Trans Power Syst Nov. 2018;33(6):7162–70. https://doi.org/10.1109/TPWRS.2018.2842093.
[11] Gonzato S, Bruninx K, Delarue E. Long term storage in generation expansion planning models with a reduced temporal scope. Appl Energy Sep. 2021;298:117168. https://doi.org/10.1016/j.apenergy.2021.117168.
[12] Garcia-Cerezo A, Garcia-Bertrand R, Baringo L. Priority chronological Time-Period Clustering for generation and transmission expansion planning problems with long-term dynamics. IEEE Trans Power Syst Nov. 2022;37(6):4325–39. https://doi.org/10.1109/TPWRS.2022.3151062.
[13] Domínguez R, Vitali S. Multi-chronological hierarchical clustering to solve capacity expansion problems with renewable sources. Energy Jul. 2021;227:120491. https://doi.org/10.1016/j.energy.2021.120491.
[14] Hoffmann M, Kotzur L, Stolten D. The Pareto-optimal temporal aggregation of energy system models. Appl Energy Jun. 2022;315:119029. https://doi.org/10.1016/j.apenergy.2022.119029.
[15] Moradi-Sepahvand M, Tindemans SH. Capturing chronology and extreme values of representative days for planning of transmission lines and long-term energy storage systems. In: 2023 IEEE belgrade PowerTech; Jun. 2023. p. 1–6. https://doi.org/10.1109/PowerTech55446.2023.10202993.
[16] Hilbers AP, Brayshaw DJ, Gandy A. Importance subsampling: improving power system planning under climate-based uncertainty. Appl Energy Oct. 2019;251:113114. https://doi.org/10.1016/j.apenergy.2019.04.110.
[17] Zhang Y, Cheng V, Mallapragada DS, Song J, He G. A model-adaptive clustering-based time aggregation method for low-carbon energy system optimization. IEEE Trans Sustain Energy Jan. 2023;14(1):55–64. https://doi.org/10.1109/TSTE.2022.3199571.
[18] Liu H, Li H, Chen J, Guo J, Tian R. A representative day selection method based on forward–backward sweep in generation expansion planning. Energy Rep Sep. 2023;9:1557–68. https://doi.org/10.1016/j.egyr.2023.04.335.
[19] Hilbers AP, Brayshaw DJ, Gandy A. Reducing climate risk in energy system planning: a posteriori time series aggregation for models with storage. Appl Energy Mar. 2023;334:120624. https://doi.org/10.1016/j.apenergy.2022.120624.
[20] Grochowicz A, van Greevenbroek K, Benth FE, Zeyringer M. Intersecting near-optimal spaces: European power systems with more resilience to weather variability. Energy Econ Feb. 2023;118:106496. https://doi.org/10.1016/J.ENECO.2022.106496.
[21] Söder L, et al. A review of demand side flexibility potential in Northern Europe. Renew Sustain Energy Rev Aug. 2018;91:654–64. https://doi.org/10.1016/J.RSER.2018.03.104.
[22] Taljegard M, Göransson L, Odenberger M, Johnsson F. Electric vehicles as flexibility management strategy for the electricity system—a comparison between different regions of Europe. Energies Jul. 2019;12(13):2597. https://doi.org/10.3390/en12132597.
[23] J. Ullmark, "PhD project scripts," GitHub. https://github.com/jonull93/PhD. accessed: Jan 12, 2024.
[24] Ullmark J, Göransson L, Johnsson F. Frequency reserves and inertia in the transition to future electricity systems. Energy Syst. Feb. 2023. https://doi.org/10.1007/s12667-023-00568-1.
[25] Weber C. Uncertainty in the electric power industry, vol. 77. New York, NY: Springer New York; 2005.
[26] Göransson L, Goop J, Odenberger M, Johnsson F. Impact of thermal plant cycling on the cost-optimal composition of a regional electricity generation system. Appl Energy Jul. 2017;197:230–40. https://doi.org/10.1016/J.APENERGY.2017.04.018.
[27] J. Ullmark, "Modelling-profiles," GitHub. https://github.com/jonull93/modelling-profiles. accessed: Jan 12, 2024.
[28] Taljegard M, Göransson L, Odenberger M, Johnsson F. To represent electric vehicles in electricity systems modelling—aggregated vehicle representation vs. Individual driving profiles. Energies Jan. 2021;14(3):539. https://doi.org/10.3390/en14030539.
[29] World Steel Association. World steel in figures. 2019. Brussels, https://worldsteel.org/wp-content/uploads/2019-World-Steel-in-Figures.pdf. [Accessed 3 November 2023].
[30] Transport & Environment. How not to lose it all: two-thirds of Europe's battery gigafactories at risk without further action. 2023. Brussels, https://www.transportenvironment.org/wp-content/uploads/2023/03/2023_03_Battery_risk_How_not_to_lose_it_all_report.pdf. [Accessed 3 November 2023].
[31] Hersbach H, et al. Complete ERA5 from 1940: fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service (C3S) Data Store (CDS) 2017. accessed: Jan 12, 2024.
[32] N. Mattsson, "GlobalEnergyGIS," GitHub. https://github.com/niclasmattsson/GlobalEnergyGIS. accessed: Jan 12, 2024.
[33] Mattsson N, Verendel V, Hedenus F, Reichenberg L. An autopilot for energy models – automatic generation of renewable supply curves, hourly capacity factors and hourly synthetic electricity demand for arbitrary world regions. Energy Strateg. Rev. Jan. 2021;33:100606. https://doi.org/10.1016/J.ESR.2020.100606.
[34] Göransson L. Balancing electricity supply and demand in a carbon-neutral northern Europe. Energies Apr. 2023;16(8):3548. https://doi.org/10.3390/EN16083548. 2023, Vol. 16, Page 3548.
[35] GIE. Gas infrastructure Europe - agsi. https://agsi.gie.eu/. [Accessed 4 October 2023].
[36] Danish Energy Agency. Technology data-energy plants for electricity and district heating generation. 2022 [Online]. Available: http://www.ens.dk/teknologikatalog. [Accessed 15 November 2022].
[37] IEA. World energy Outlook 2018. 2018. Paris, https://www.iea.org/reports/world-energy-outlook-2018.