Stochastic modelling of variable renewables in long-term energy models: Dataset, scenario generation & quality of results

Pernille Seljom and Lisa Kvalbein and Lars Hellemo and Michal Kaut and Miguel Muñoz Ortiz
IFE, Kjeller, Norway
SINTEF, Trondheim, Norway

in: Energy. See also BibTeX entry below.

BibTeX:

```
@article{SeljomEtAl2021,
  author    = {Pernille Seljom and Lisa Kvalbein and Lars Hellemo and Michal Kaut
               and Miguel Mu{\~{n}}oz Ortiz},
  journal   = {Energy},
  title     = {Stochastic modelling of variable renewables in long-term energy models:
               {D}ataset, scenario generation {\&} quality of results},
  year      = {2021},
  pages     = {121415},
  volume    = {236},
  doi       = {10.1016/j.energy.2021.121415}
}
```

Stochastic modelling of variable renewables in long-term energy models: Dataset, scenario generation & quality of results

Pernille Seljom Lisa Kvalbein Lars Hellemo Michal Kaut
Miguel Muñoz Ortiz

May 2021

Variable electricity generation from wind and solar influences the design of a cost-efficient and reliable energy system. This paper presents a method that uses stochastic programming to represent variable renewable electricity generation in long-term energy system models, and demonstrates this on a Norwegian TIMES model. First, we derive hourly PV- and wind-generation data by modifying satellite-based data, based on a comparison with historical generation data. Second, the satellite-based dataset is transformed into a manageable set of scenarios that is used as an input to the stochastic energy-system model. This is done using six different scenario generation methods. Third, we solve the energy-system model with three of the scenario-generation methods and evaluate the quality of the corresponding model value by stability tests.

We demonstrate that scenarios generated from the six methods have significantly different moment-based and Wasserstein distance error relative to the dataset. Further, the energy system model results show that the number of scenarios needed to achieve stability differs between the three used scenario generation methods.

1 Introduction

Non-dispatchable variable renewable electricity generation (also called intermittent electricity generation) and electrification of end-use are considered as solutions to decarbonising the future energy system. However, an increased share of variable renewable electricity supply, and more unpredictable electricity use, requires flexibility of the energy system to ensure that the electricity demand is always met. There are several strategies that can reduce the mismatch between variable renewable electricity generation and the electricity demand. First, variation in variable supply can be counteracted by generation from flexible electricity technologies, such as thermal power and hydropower, and by use of batteries and by electricity trade. Second, demand can be changed over time with end-use flexibility by, e.g. fuel switch, demand response and use of local storage. Consequently, a decarbonised future requires highly integrated energy system between energy supply, energy infrastructure and end-use sectors.

Long-term energy models can systematize the complexity of a future energy system and their analysis can thus provide decision support on future decarbonisation pathways. To ensure reliable model decisions that provides optimal investments, that explicitly values flexibility, it is necessary to take into account the variable weather dependency of renewable generation. Stochastic programming is identified in [1] as one modelling approach to improve the representation of variable renewable energy in long-term energy models, since it, e.g. captures the need for back-up capacity and flexible solutions. As demonstrated in, e.g. [2], a stochastic-programming approach can be used to provide investments that explicitly consider different operational situations that can occur due to outcomes of weather-dependent renewable generation.

This paper presents a three-step approach to representing variable electricity generation in long-term energy models, and demonstrates this approach on a TIMES model [3] of the Norwegian energy system [4]. The first step derives a dataset that describes the variable and uncertain characteristics of renewable generation. This is done by modifying simulated hourly satellite-based solar- and wind-power generation data, using historical generation data. The second step transforms the dataset into discrete scenarios which are used as an input to the energy-system model. Since the computational effort increases with number of scenarios used, it is desirable to use as few scenarios as possible, as long as the scenarios provides good quality model results.

The third step involves identifying the scenario-generation method that provides a good representation of the dataset, and evaluating how many scenarios is needed to provide model results of good quality. This is done first by calculating the distance between the scenario trees and the satellite-based dataset, and by applying stability tests on the model results. The motivation of the stability tests is to ensure that the results depend on the dataset rather than the scenario generation methodology itself, and to evaluate the number of scenarios that are needed to achieve stability.

The main contribution of this paper is the overall evaluation of the dataset, scenario generation method and the quality of results related to modelling of variable renewables. Thus, this paper is a suitable guideline on how to incorporate variable renewable generation into long-term energy models by using stochastic programming. A second contribution is the comparison of the distance-error and stability for different scenario-generation methods. The final contribution is that it is the first time the moment-based scenario-generation method by integer optimization, from [5], is applied and evaluated in a long-term energy-system model.

In the following, we give an overview of related literature to set our contribution in a context.

1.1 Related literature

Long-term energy models, with endogenous investments, are used to provide insights on the cost-optimal supply and demand side in future developments of the energy system [6]. BALMOREL [7, 8], MESSAGE [9], OseMoSYS [10] and TIMES [3] are well-known modelling frameworks that are designed to generate long-term energy models.

According to [11], a main challenge of long-term energy system models is to find the trade-off with the resolution in time, in space, in techno-economic detail and in sector-coupling to provide good decision support. Through a literature review, the authors conclude that none of the studied long-term energy model have a simultaneous high resolution in all these fields. This confirms that models have limitations, and their weaknesses should be acknowledged when used to provide decision support. In [6], it is emphasised that a major challenge of energy-system models with variable renewables is to balance the temporal and spatial resolution with data availability and computational tractability. This is, e.g. because a higher temporal resolution requires data at a high time resolution and increases the computational time of the models. The impact of the temporal resolution in energy planning models is demonstrated in [12], that concludes that a high temporal resolution is important when modelling variable renewable resources. The authors show how a model that does not fully consider the variability of demand and supply can overestimate the amount of demand that is met by variable electricity generation. This is supported by [13], that demonstrates that a coarse temporal resolution, when modelling the integration of renewable electricity generation into the Belgium electricity system, provide sub-optimal investments, overestimating the share of variable renewables and underestimate the future operating and maintenance costs.

One way of lowering the need for a high temporal resolution in long-term energy models, is by selecting representative parts, such as weeks or days, of a longer time-series. For example, [13] demonstrates that the accuracy of an energy-system model can be improved by using a statistical method to select a set of representative historical days from an hourly time series of a year. In [13], the selection of representative days is done by using the so-called Integral Method [14], that involves that the peaks of the representative days, corresponds to the average peak value of all the corresponding historical days. Another way of selecting representative days, is done in [15], by using the k-medoids approach, where an actual observed day, that minimises the distance of the selected day to the other members of its cluster, is selected. A conclusion of this work is that a clustering approach can give significant different model results on capacity needs and technology mix compared to using the Integral Method to select representative days. Further, [16] demonstrates a generalised clustering algorithm to select representative time periods, and compares it with three other selection methods, for a power system expansion model of Northern Europe. In this study, the authors conclude that the performance of the various methods depends on input data and the number of representative periods.

In the above literature, only hourly renewable generation data for one year is used to select representative time periods. Thus, these studies do not consider the variations in renewable generation over several years. Further, the representative time periods are used to define the model input of deterministic models, that make investment decisions based on only one realisation of weather-dependent parameters for each sub-annual time-slice. Consequently, these models do not explicitly consider the uncertainty of weather-dependent renewable generation in the optimization. In fact, in a review of modelling tools, [17] concludes that it is only a small share of the long-term energy-model literature that takes into account the weather-dependent uncertainty of renewable electricity generation.

Stochastic programming [18] is a mathematical framework that can be used to explicitly model short-term uncertainty of, e.g. variable renewables in optimization models. A two-stage stochastic model can be applied to provide investments that explicitly consider parameters that are exposed to short-term uncertainty. A stochastic approach to incorporate short-term uncertainty in TIMES was first introduced in [19], is demonstrated in a user guide [20], and was first applied to a TIMES model of the Danish energy system to consider the short-term uncertainty of wind-power generation [2]. Despite several later applications of stochastic modelling of short-term uncertainty in TIMES models, the majority of the TIMES literature uses a deterministic approach [21].

Studies using long-term energy models with a stochastic modelling of short-term uncertainty demonstrates that the model results differ to those from a deterministic model. For example, in a study addressing decarbonisation strategies of the Arctic settlement at Svalbard [22], the authors demonstrate that a deterministic investment strategy is not able to meet the energy demand in at least one of the operational scenarios, as the deterministic modelling approach overestimates the contribution from wind energy. Analyses of the impact of zero energy buildings on the Scandinavian energy system [23] concludes that a deterministic approach overestimates the competitiveness of variable electricity generation and underestimates investments in heat capacity in buildings.

A drawback of introducing stochasticity into long-term energy models is that the computational effort increases with the number of discrete scenarios. There is also a trade-off between increasing the temporal resolution and increasing the number of scenarios. This is demonstrated in [24] that compares the results of a TIMES model of the European electricity- and heat market using a deterministic and stochastic modelling approach, with various temporal resolutions. In this study, a stochastic model with 12 sub-annual time-slices and 15 scenarios, and a deterministic model with 672 time-slices, both used about 10 hours to solve, whereas a deterministic model with 2016 time-slices used about 299 hours. The authors conclude that for this model instance, the stochastic model provides better results, compared to both the deterministic models.

For stochastic long-term energy models, the dataset to describe the variable electricity generation is derived by using either generation statistics or satellite-based simulations. [2] derive the wind power availability, also called capacity factor, for both Danish spot price regions. However, using statistics can be cumbersome when covering a high spatial resolution or many countries/ regions. The analysis of [24], [25] and [26] are examples of studies that use Renewables.ninja [27] as a basis for wind- and PV availability for 30 years for most European countries, when analysing the development of the European power- and heat market towards 2050. Renewables.ninja is an open web application that is based on the Global Solar Energy Estimator (GSEE) model [28] and the Virtual Wind Farm model [29], using the MERRA reanalysis [30], biased-corrected using national generation data. Another openly available source of data for all European countries is EMHIRES [31], [32], [33]. It is not clear which of the dataset has the best quality, although a comparison for selected countries (though not Nordic countries) indicates that Renewables.ninja is better for wind and EMHIRES is better for PV [34]. Our paper generates wind time series based on the methodology of [35], applying adjustments to the Renewables.ninja data, to make it more suitable for Norwegian conditions.

Similar to the literature on selecting representative days for deterministic models, the literature on stochastic long-term energy models applies several methods to generate scenarios from a dataset. Random sampling is applied to generate scenarios for a power-generation expansion model in [36], k-means clustering is used in transmission and generation planning in [37] and to a power-generation expansion model in [38], and an iterative moment-based sampling approach is used in, e.g. a TIMES energy-system models of the European electricity and district heat sector in [24], a TIMES energy-system model of the Scandinavian energy system [39] and to the EMPIRE European power-market model in [25].

Most of the stochastic energy literature lacks a discussion on the quality of the presented model results [21]. This is a weakness, since poor scenario representations can give results that depend on the scenario representation rather than the characteristics of the renewable generation. There are, however, studies that acknowledge this weakness. The authors of [40] acknowledge that the shortcomings of using few scenarios is that they do not cover the total range of fluctuations in variable generation, and [39] states that using a higher number of scenarios can improve the quality of the model results.

Executing stability tests on in-sample and out-of-sample stability [41] is a way of evaluating the solution quality. Stability tests indicate whether the optimal solution value is skewed by a misrepresentation of the true underlying stochasticity. Stability tests are used only in a small share of the literature on long-term energy models. [2] applies both in-sample and out-of-sample stability tests in a Danish TIMES model with uncertain wind-power production. Although not discussed explicitly, in-sample stability is tested in numerous studies that evaluate how the model solution value is effected by the number of scenarios used, see, e.g. the power expansion problem in [38] and [42], a bi-level investment problem for a wind power producer in [43] and a European energy-system model in [24]. If the scenarios are generated by random sampling, Sample Average Approximation (SAA) [44] can be used to evaluate the solution quality by providing a confidence intervals of the optimality gap for given model solutions. The use of SAA and stability tests of random sampled scenarios are demonstrated for an energy-system model of Denmark in [21]. If the scenario generation is based on a clustering method, such as k-means, the optimality gap of clustered scenarios can be derived. This is demonstrated in [37] in a power-expansion model covering the west of the Unites States, with uncertain hydropower, PV, wind generation and electricity demand.

2 Methodology

First, this section presents the models structure and assumptions of the TIMES energy-system model, including the stochastic modelling approach. Second, we describe how the dataset for wind power and PV is generated from satellite- and measured data. Third, we give an overview of six different scenario-generation methods that transform the dataset into discrete and manageable model input to the energy-system model. Fourth, we describe the in-sample and out-of-sample stability tests. Fifth, we introduce a conceptual flowchart of the methodology that is used in the paper.

2.1 Energy-system model

IFE-TIMES-Norway [4] is a TIMES [3] optimization model of the Norwegian energy system. TIMES is a bottom-up framework that provides a detailed techno-economic description of resources, energy carriers, conversion technologies and energy demand. It is mainly used for medium- and long-term analysis on the global, national and regional levels, including the IEA Energy Technology Perspective [45]. TIMES models minimize the total discounted cost of a given energy system to meet the demand for energy services for the regions over the period analysed. The total energy-system cost includes investment costs in both supply and demand technologies, operation and maintenance costs, and in-come/costs of electricity trade with countries outside Norway.

IFE-TIMES-Norway is a technology-rich model of the Norwegian energy system that is divided into five regions, corresponding to the current electricity market spot price areas, see Fig. 1. The model provides operational and investment decisions from the starting year, 2020, towards 2060, with five model periods within this model horizon. Each model period is divided into 96 sub-annual time slices, where four seasons are represented by 24 hours each. The model has a detailed description of end-use of energy, and the demand for energy services is divided into several end-use categories within industry, buildings and transport. Note that energy services refer to the services provided by consuming a fuel and not the fuel consumption itself. For example, the heating demand in buildings is an energy service while the fuel used to heat the building is not.

Each energy-service demand category can be met by existing and/or new technologies using different energy carriers, such as electricity, bio energy, district heating, hydrogen and fossil fuels. Consequently, in each time-slice, the use of energy carriers, such as electricity, is a model result and not a model input, making sector coupling a part of the optimization. There are, for example, endogenous investments in heat pumps and electric vehicles that couple the power sector with the heat and transport sector, respectively. In total, the demand for energy services can be met by using 400 end-use technologies within buildings, transport and industry. Other input data include fuel prices; electricity prices in countries with transmission capacity to Norway; renewable resources; and technology characteristics such as costs, efficiencies, and lifetime and learning curves.

The transmission capacity within and outside the Norwegian model region is a model input that is based on current capacity and ongoing transmission capacity expansion. Further, there it is an investment option to expand the transmission capacity, both

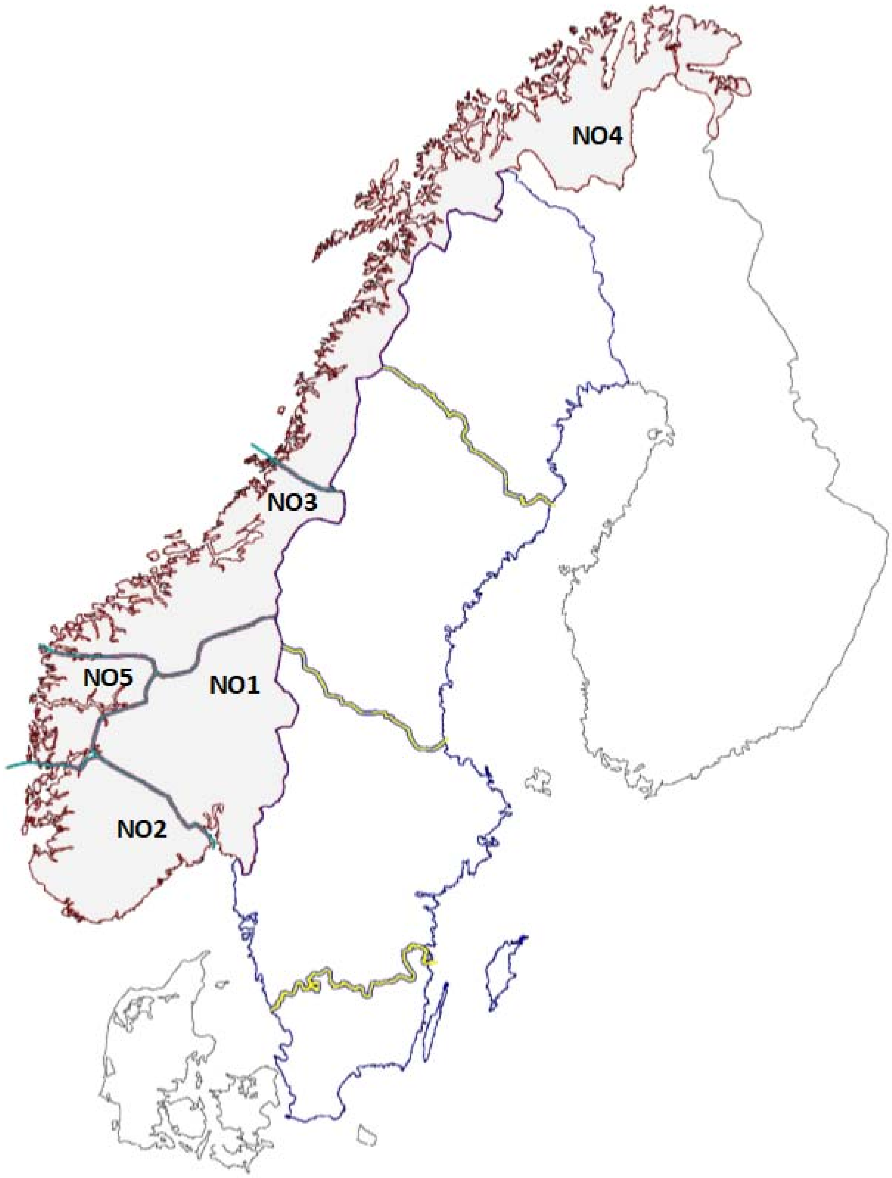

Figure 1: Illustration of model regions of IFE-TIMES-Norway, NO1 to NO5.

within Norway and to trading countries. The electricity prices in the model regions in Norway are endogenous, as they are the dual values of the electricity balance equation, while the electricity prices in the countries with trading capacity to Norway, including Denmark, Sweden, United Kingdom, Finland, Netherlands and Germany, is a model input. Further, we assume these electricity trade prices to be independent of the traded quantities to Norway.

The model includes two types of hydro plants, Reservoir plants and Run-of-the-river (ROR) plants that are divided into five different types; existing plants and four new plant

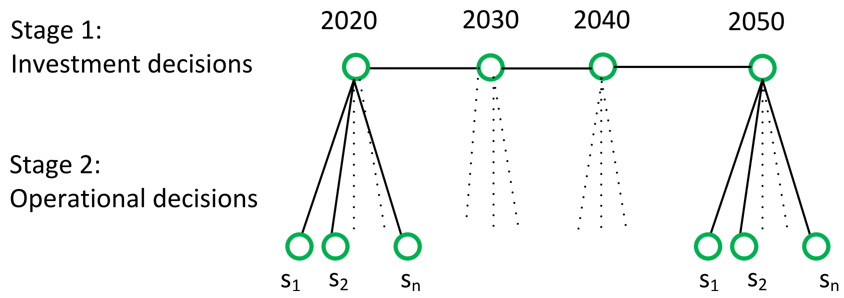

Figure 2: The information structure of a two-stage stochastic model with short-term uncertainty.

options with different investment costs. The capacity for the existing plants is a model input whereas the capacity for the new plant options requires endogenous investments. For the ROR plants, with seasonal dependent electricity generation. As opposed to the ROR plants, the Reservoir plants are flexible between the time-slices of the model. The annual electricity generation from the Reservoir plants is constrained by an annual capacity factor, the annual production over the maximum theoretical production over a year. In addition, we constrain the seasonal electricity generation from the Reservoir plants to a maximum and a minimum production level according to historical observations.

The existing wind power, and wind power that is under construction is an input to the model. Further, for each region, new wind power is modelled by ten technology types that has different costs, operational conditions and upper potentials in each region. Further, solar power is split into building applied PV on commercial and residential buildings.

A two-stage stochastic framework [18] is applied to provide investment decisions in IFE-TIMES-Norway that explicitly consider various operational situations that can occur due to short-term uncertainty in wind and PV generation. Each uncertain parameter is represented by a set of possible realizations, called scenarios, all having the same probability of occurring. Further, we apply a multi-horizon model structure, with no dependency on the operational decisions between the model periods [46]. Figure 2 illustrates the resulting information structure. The first-stage decisions, here investment decisions for a specific period, are made under uncertainty, that is, without knowing the realisation of the variable electricity generation for the period considered. The second-stage decisions, here operational decisions, are made after the uncertain parameters of the period considered are revealed. Hence, there is one operational decision for each of the scenarios. Consequently, investments are identical for all scenarios, whereas operational decisions are scenario dependent.

To consider the different operational situations in the optimization, TIMES minimises the investment costs and the average of the operational costs for all scenarios. This gives investment decisions that recognise the expected operational cost, and that are feasible for all the model-specified realisations of the uncertain parameters. Note that the investment and operational model decisions are made simultaneously for all model periods, with operational decisions that are contingent on the realisations of short-term uncertainty. In this paper, it is assumed that the uncertain parameters related to the availability of wind and PV generation are the same for all model periods.

2.2 Dataset

The dataset used to describe the short-term uncertainty includes representative capacity-factor time series of the five different spot price regions in Norway, NO1 to NO5, for wind and PV generation. Capacity factor is described as the ratio of net electricity generated for a given time period, to the total energy that could have been generated if that plant would have operated at full-power during the same period of time.

These time series are based on data from Renewables.ninja [47]. Capacity Factor (CF) series obtained for PV were evaluated as sufficiently accurate in comparison with individual PV panels and are used as provided. However, local wind conditions are not accurately represented in Renewables.ninja database due to the interpolation from large grid units of the database. Therefore, wind speed series obtained from Renewables.ninja will be adjusted to better represent the local wind speeds in the chosen wind parks representing the aforementioned regions, and then the adjusted wind speeds will be used to calculate CF. Individual wind farms have been selected due to the importance of the specific locations of these power plants. For PV, we assume that locations in the centre of the main Norwegian cities in the spot-price regions are representative for the whole area.

2.2.1 Adjustments average wind speed from Renewables.ninja

The resolution provided by Renewables.ninja is based on MERRA-2 Wind data: a 0.625 by 0.5 degrees grid (approximately $50\,\mathrm{km} \times 50\,\mathrm{km}$), which can be inexact to represent local weather phenomena or local geography. The time resolution is 1h and the dataset is available for 19 years (2000-2018). To improve accuracy for wind parks, new time series are calculated from the wind profiles obtained from Renewables.ninja, adjusting the average wind speed with values from Global Wind Atlas 3.0 that has higher spatial resolution of $3\,\mathrm{km} \times 3\,\mathrm{km}$ [48].

To escalate the new average speeds to the average hub height of the wind parks, the logarithmic dependence of wind speed with height is used, also used in [35] and described in [49].

The new adjusted wind speed profile $v(t)$ is obtained as

$$v(t) = v_{ninja}(t) \cdot \frac{v_{GWA}^{h_{hub}}}{\overline{v_{ninja}}}\,, \tag{1}$$

where $v_{ninja}(t)$ is the wind speed obtained from the MERRA-2 data set from the Renewables.ninja API, $\overline{v_{ninja}}$ is the average wind speed of MERRA-2 data set (considering wind speed data from 2000 to 2018) and $v_{GWA}^{h_{hub}}$ is the escalated wind speed, from $50\,\mathrm{m}$ (where v_{GWA}^{50} is given) to the specific average hub height of each individual wind park.

To obtain production time series from the wind-speed time series, a turbine power curve is used, depending on the installed technology and model. The curves were obtained from

Table 1: Input data used for the Renewables.ninja API for the selected wind parks and wind speed scaling parameters. C_{wp} is the installed wind park capacity in MW, h_{hub} is the hub height (in m), v_{GWA}^{50} is the average wind speed from GWA at 50 m in m/s and z is the roughness length in m.

Wind park	Area	Lat.	Long.	Turbine model	C_{wp}	h_{hub}	v_{GWA}^{50}	z
Marker	NO1	59.499 76	11.732 501	Vestas V136 3.45	54	142	5.61	1.5
Raskiftet	NO1	61.208 037	11.758 786	Vestas V126 3.6	111.6	117	5.84	0.5
Lista	NO2	58.157 056	6.711 444	Siemens SWT 2.3 93	71.3	80	7.52	0.2
Høg Jæren	NO2	58.642 712	5.763 826	Siemens SWT 2.3 93	73.6	80	7.96	0.2
Ytre Vikna	NO3	64.900 572	10.891 87	Enercon E70 2300	39.1	64	7.95	0.05
Valsneset	NO3	63.818 975	9.623 031	Enercon E70 2300	11.5	64	8.13	1.5
Flakken	NO4	70.100 411	20.106 254	Vestas V90 3000	54	80	6.46	0.5
Kjøllefjord	NO4	70.922 16	27.281 894	Siemens SWT 2.3 82	39.1	70	7.34	0.05
Raggovidda	NO4	70.763 249	29.084 506	Siemens SWT 3.0 101	45	80	9.32	0.05
Midtfjellet 1	NO5	59.930 537	5.372 79	Nordex N90 2500	85	100	7.31	0.2
Midtfjellet 2	NO5	59.930 537	5.372 79	Nordex N100 2500	25	100	7.31	0.2
Midtfjellet 3	NO5	59.930 537	5.372 79	Nordex N100 3300	39.6	100	7.31	0.2

[50] and [51]. In order to represent a wind park, the previous power curve is smoothed using the methodology and parameter values in [52].

2.2.2 Representative wind generation

Wind parks selected to represent each spot-price region are listed in Table 1. The longitude and latitude for the locations is given with all decimals to allow a replication of the results, as the MERRA-2 grid is interpolated to the exact given location (see [28]).

Finally, despite the adjustments described above, there is significantly differences in the average satellite-based CF and the historical CF, as shown in Table 2. To take this to account, all the hourly satellite-based CF values were adjusted, by multiplying with one value, such that the average CF of the dataset corresponds to the annual CF values of the historical generation.

2.2.3 PV time series

In order to get representative PV capacity-factor time series for the different Norwegian trading regions, representative orientations and tilt angles have been selected. CF time series have been obtained from Renewables.ninja database through the API. The time series provided an acceptable yearly correlation when comparing to four PV installations in southern Norway (three in NO1 and one in NO2) and one in the very north, in the city of Tromsø (NO4), between 0.72 and 0.93. Thus, no adjustments were performed. In Table 3, the input data for the Renewables.ninja API is summarised for the chosen locations (the most representative cities Oslo, Bergen, Trondheim, Kristiansand, Tromsø).

Table 2: Final reference capacity factor (average capacity factor in the area provided by NVE [53]) used to define representative wind profiles for each of the regions. Average CF for NO5 is based on concessions exclusively, since no wind park has been installed yet

Area	Average CF	Generated CF
NO1	0.39	0.47
NO2	0.37	0.46
NO3	0.36	0.35
NO4	0.35	0.26
NO5	0.39	0.42

Table 3: Different solar system configuration considered (above) and the main cities representative of the Norwegian electricity trading regions (below)

Installation type	Orientation	Tilt(°)
Roof (flat)	East-West	10
Roof (inclined)	South	30

Location	Region	Lat.(°)	Long.(°)
Oslo	NO1	59.91406	10.75208
Kristiansand	NO2	58.159912	8.01822
Trondheim	NO3	63.430519	10.39507
Tromsø	NO4	69.649204	18.95532
Bergen	NO5	60.39126	5.322064

Additionally, the system losses have been assumed to be 10 % for all cases and all systems are fixed (no tracking).

2.3 Scenario generation

In this section, we present several methods for representing the data series of wind and solar capacity factors as scenarios that can be used by optimization models. All methods come from [5] and are based on the same idea: select S historical days, such that the multivariate distribution of capacity factors across the S sequences is as close to the overall empirical distribution as possible. The reason we want to use whole sequence, consisting of 24 hours of a day, instead of the usual approach of sampling from an estimated stochastic process, is to ensure that the sequences are realistic both in terms of inter-temporal dependencies and dependencies/correlations between parameters and regions.

Since there are significant seasonal differences in distributions of the capacity factors, we generate one set of scenarios per season. Note that this implies assumption of independence between seasons. Moreover, the scenario-tree structure of TIMES (Fig. 2) dictates that corresponding scenarios have the same probability in the four seasons. We achieve this is by generating equiprobable scenarios, i.e., require that all scenarios have probability $1/S$.

2.3.1 Random and Iterative sampling

The easiest option is to simply sample (randomly select) S sequences for each of the four seasons. We will refer to this approach as *Random sampling*. It is very fast, taking only a fraction of a second. The main weakness of this method is that we can end up with a highly non-representative sample, in particular for smaller S.

One way to improve this is to repeat the sampling N times and then use the 'best' sample—we refer to this variant as *Iterative Sampling*. To select the 'best' sample, we need some measure of the sample 'quality'. One option is use some distance between the scenario distribution and the distribution of the whole data set. In this paper, we follow [2] and use the difference between the first four marginal moments plus correlations as the distance. Since stochastic programs typically react stronger to errors in means and weaker for the higher moments [54], we use the following weights for the four moments and correlations: 10, 5, 2, 1; 2.

2.3.2 k-means clustering, standard and constrained

Another option is to use k-means clustering [55, 56] to divide the data into S clusters, and select a representative scenario from each cluster. For the latter, we simply use the scenario closest to the cluster's mean. The advantage of k-means is that it is a standard method with many available implementations and that the method is fast. On the other hand, there is no guarantee that the resulting scenarios constitute a good approximation of the distribution. This is further aggravated by our requirement of equiprobable scenarios. The last issue can be addressed using the constrained k-means clustering method from [57], which allows for constraints on cluster sizes.

2.3.3 Optimizing distance in moments and correlations

Instead of using Iterative sampling, where the best of N samples are selected, we can use mixed-integer linear programming to select the historical days, by minimizing the distance in moments and correlation, see [5]; we refer to this method as *Optimization*. Binary variables are used for selecting the days, and we express all moments and correlations as linear combinations of $E[X^k]$. These, in turn, can be written as a linear combination of the binary selector variables and corresponding powers of the data, giving a linear model, see [5] for details. Unfortunately, the resulting models cannot be solved to optimality for realistic problems with data spanning more than a couple of years, since we need one binary variable per day. With our data set, this means between 1625 and 1656 binary variables, depending on the season.

On the other hand, since the Optimization algorithm is doing an 'intelligent search', it should find good solutions faster than using 'blind' sampling. In theory, it also has the advantage of providing bounds on the distance from the best possible selection, but this bound is often very loose, and thus unusable, in practical cases.[1]

2.3.4 Wasserstein distance

Instead of using moments and correlations, one can measure the distance of the scenarios from data using the Wasserstein distance [58, 59, 60]. Since exact minimization of Wasserstein distance is not possible for real-life problems, we have implemented a variant of Algorithm 2 from [61], which we initialize by Voronoi sets implied by a randomly-generated sample. See [5] for details.

2.4 Stability tests

We evaluate the quality of the scenarios using stability tests from [41], since scenarios represent an approximation of the dataset. Moreover, any scenario-generation method involving randomness will add variability to the solutions of the problem, as the trees will differ between consecutive runs. We want both the approximation error and the variability from the scenario-generation to be as small as possible.

[41] introduces two tests: the *in-sample stability* test investigates whether a scenario generation method creates a set of scenarios that consistently result in approximately the same objective function, when used to solve the problem with different scenario sets generated from the same input. Furthermore, there should not be too large discrepancies in the optimal value when the number of scenarios, S, increases.

The goal of the *out-of-sample stability* test is to measure both stability and quality of the solutions obtained when using scenarios from a given method. In principle, the test requires the use of the whole dataset, though in practice one commonly uses an approximation [41]. We will refer to this approximation as a *benchmark tree*. We solve the optimization problem on the benchmark tree with first-stage decisions fixed to the values from a given scenario-based solution; the resulting objective value is an approximation of the quality of the solution.

As with the in-sample tests, we want small variability from scenario sets generated with the same input. If we can solve the optimization problem on the benchmark tree, we can also use the difference between the scenario-based solutions and the optimal value as a measure of quality of the scenario-based solution.

2.5 Methodology flowchart

Fig. 3 shows a conceptual flowchart of the methodology used in this paper to illustrate the interactions between the dataset, energy system model, scenario generation methods and stability tests.

First, the dataset, consisting of 19 years of hourly capacity factors of wind and PV is the input to six different scenario generation methods. Each scenario methods generate a set of scenarios consisting of 3, 9, 15, 21 and 30 number of scenarios. Then,

[1] In general, the optimization-based method has an additional advantage of being able to optimize the probabilities as well, allowing for a better match.

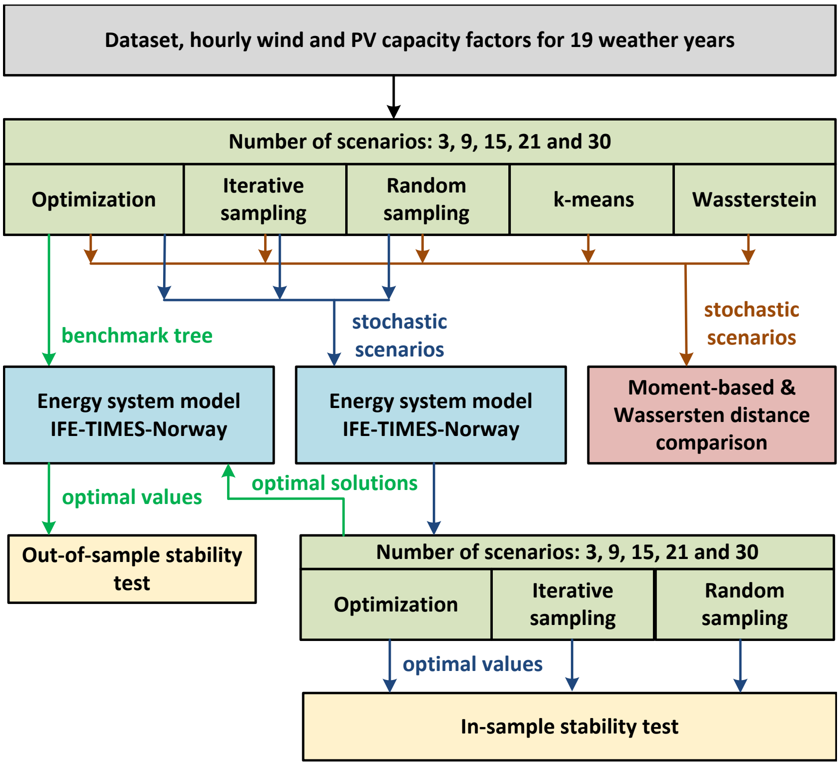

Figure 3: Conceptual flowchart of paper methodology

the resulting scenario trees are compared based on a moment-based and a Wasserstein distance. Thereafter, we use three of the six scenario generation methods as an input to the energy-system model. Finally, the corresponding optimal values of the model results is used to test for in-sample stability whereas the optimal solutions is used to test for out-of-sample stability on a bench-mark scenario tree from the Optimization scenario generation algorithm.

3 Results and discussions

First, this section compares the scenario generation methods in terms of runtime and moment-based and Wassersten distance. Thereafter, in-sample and out-of-sample stability are used to evaluate the quality of the results.

3.1 Scenario generation

We generate 24-hour scenarios from the data presented in Section 2.2 for 15 capacity factor series; hourly capacity factors for wind power, residential PV and commercial PV for the five model region of IFE-TIMES-Norway.

One way to handle the time dimension is to consider values for each hour as a separate time series, resulting in $24 \times 15 = 360$ series. This means that the clustering methods will

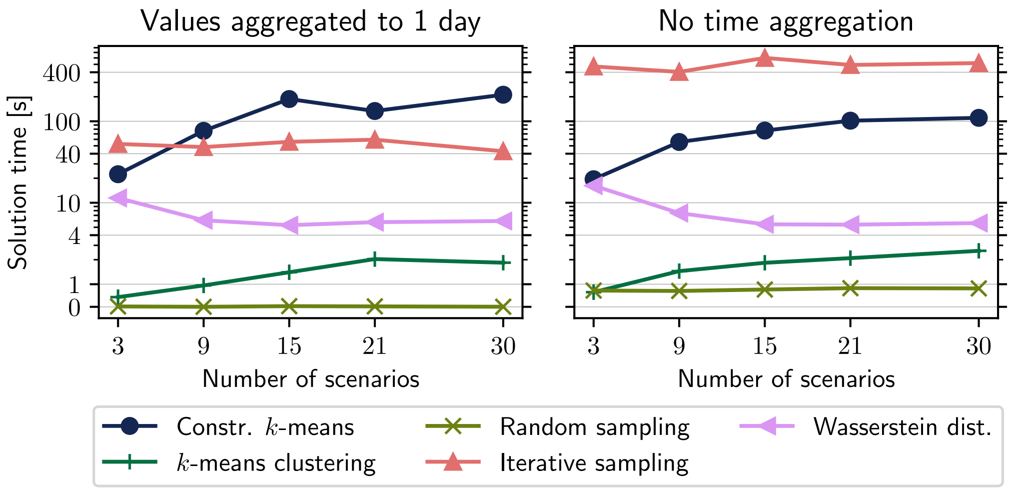

Figure 4: Runtime of the different scenario-generation methods, with (left) and without (right) time aggregation. Note the logarithmic scaling for times over 1 second. The 'Iterative sampling' method was run with 2500 iterations. Times for Optimization are not shown, as it was run with a time limit of 900 seconds that was reached in almost all cases.

work in dimension 360 and all moment-based methods need to evaluate $4 \times 360 = 1440$ moments and $360 \times 359/2 = 64\,620$ correlations, and to match all this with just 3–30 equiprobable scenarios.

An alternative way to dealing with the time dimension is to aggregate the values and use daily averages, reducing the dimension to 15 and the number of moments and correlations to 60 and 105, respectively. While we give up control over the intra-day distribution, the lower dimension should allow for a better match, so it is not clear which approach will produce better scenarios. Moreover, some of the methods run significantly faster in the lower dimension.

The scenario-generation was implemented in Python: k-means clustering uses `KMeans` class from scikit-learn, the constrained k-means method uses https://adared.ch/constrained-k-means-implementation-in-python/, and the optimization model was implemented in Pyomo [62] and solved using FICO™ Xpress solver version 8.8.0.

3.1.1 Runtime comparison

Scenario-generation runtimes on a PC with 2.8 GHz Intel CPU and 16 GB RAM are presented in Fig. 4, for $S \in \{3, 9, 15, 21, 30\}$. All values are average of 10 runs. In each run, we generate scenarios for all four seasons separately and report run-time of each generation. Hence the results are an average of 40 values, and the time to generate scenarios for all four seasons is four times the reported values.

The figure does not include the runtime for Optimization, as it was run with a time limit of 900 s. In the aggregated case, it reached the time limit in all but two instances,

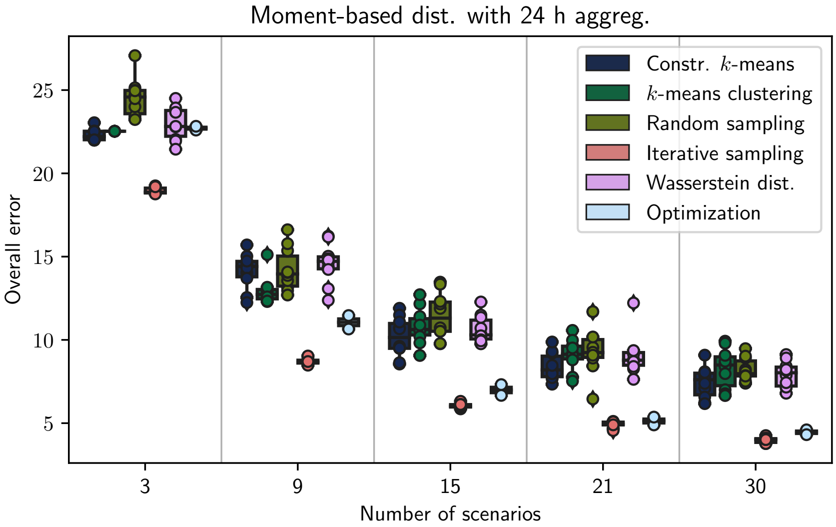

Figure 5: Moment-based distance of scenarios with aggregated data

so the reported results are not optimal solutions. In the case without time aggregation, the model gets so large that we do not even get a feasible solution within the time limit. In the rest of the paper, we therefore use scenarios from the aggregated case.

As depicted in Fig. 4, random sampling, k-means and Wasserstein-based methods are both fast and unaffected by the dimension of the input vector and the number of scenarios. The runtime of Wasserstein-based method is actually faster with more scenarios, possibly because it gets easier to obtain a good match. The constrained-k-means method is slower and its run-time increases with the number of scenarios. Run-time of the iterative sampling depends on the selected number of iterations, N. In addition, it increases with dimension, from under $50/2500 = 0.02$ seconds per iteration in the aggregated case (dim. 15) to $500/2500 = 0.2$ seconds per iteration in the full case (dim. 360). This is expected, as the moment-based distance evaluation is the most time-consuming part of the method.

Note that the requirement for a fast solution time to generate stochastic scenarios are less important for long-term energy models that is used for planning purposes, compared to operational models that generate new scenarios on a more frequent basis. Also, the stochastic assumptions on wind and PV capacity factors in energy-system models is a static model assumption, that does not need to be updated if the data basis is unchanged.

3.1.2 Comparison of measured distribution error

Next, we compare the generated scenarios in terms of the computed distance in moments and correlations. We run each method 10 times, $M = 10$.

Since Optimization gives the same solution on each run, we generate an additional set of scenarios for Optimization on a different machine, with a different version of Xpress (v. 8.9.2).

The results are presented in Figs. 5 and 6 for scenarios with and without aggregation,

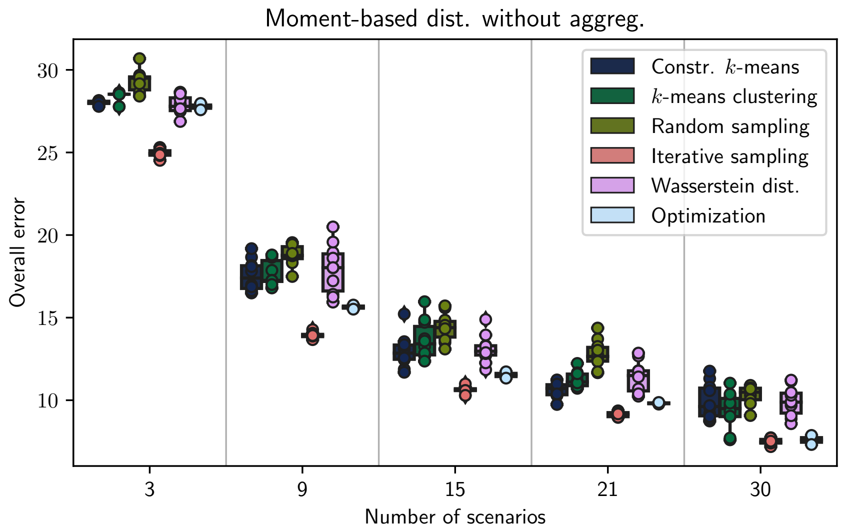

Figure 6: Moment-based distance of scenarios with non-aggregated data

respectively. There, we can see that:

- Iterative sampling is best among the analysed scenario generation methods in all the tested cases.
- Optimization is the next-best method for all variants with at least 9 scenarios. It is very close to Iterative sampling for 21 and 30 scenarios with aggregation and 30 scenarios without aggregation.
- There is very little difference between standard and constraint k-means, so the extra runtime of the constrained method is probably not justified.
- The Wasserstein-distance-based method is not much better than Random sampling. However, we have to remember that this method optimizes a different distance measure, so it is not a fair comparison.

Since the evaluation uses the same distance that the optimization model tries to minimize, one could ask why it is not best every tested variant. There are three reasons for this:

- In most of the runs, Xpress reached the time limit before finding the optimal solution.
- As pointed out above, Optimization becomes too difficult to solve without aggregation, so we use the results with aggregation for both variants. This means that in all non-aggregated cases, the evaluation distance differs from the optimized one.
- The model uses an LP approximation of the distance [5]. In particular, in all the moment formulas including $(X - \hat{\mu})^k/\hat{\sigma}^k$, we replaced the actual sample values $\hat{\mu}$,

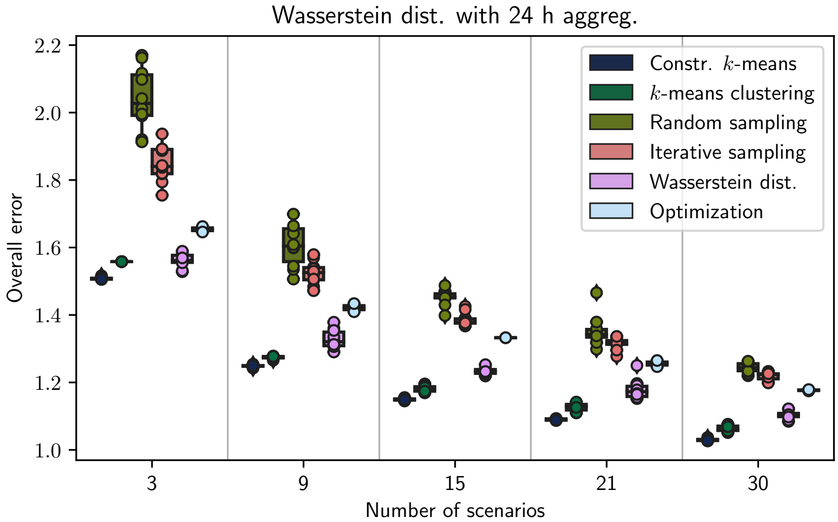

Figure 7: Wasserstein distance of scenarios with aggregated data

$\hat{\sigma}$ (which would be variables in the model) by their target values (which are parameters of the model), to keep the model linear. This implies that the model's evaluation of the higher moments and correlations is exact only if the mean and standard deviation is matched exactly—yet this is impossible with only few scenarios. This explains why the optimization method gets better, relative to the other methods, as the number of scenarios increases.

If we use the Wasserstein metric to measure the discretization error, the order of the methods changes, as illustrated in Figs. 7 and 8. This distance prefers both variants of k-means, followed by the Wasserstein-based heuristic. This is to be expected, since k-means can be seen [5] as a heuristic for minimizing Wasserstein distance of order 2 with an Euclidean metric—which is what we use.

This illustrates that one should not select a scenario-generation method based solely on the statistical properties of the generated scenarios, because we do not know upfront which properties are important for the model at hand. Instead, we need to solve the model on the trees and see which method produces trees that give the 'best' results.

3.2 Energy-system model results

In this section, we use the scenarios to solve the TIMES energy-system model. Unfortunately, the TIMES model takes up to several days to solve for the largest instances, so we had to limit the scenario-method selection. For this reason, we test only the following three methods:

- Iterative sampling (with 2500 samples), as this is a method tested with TIMES earlier.
- Optimization, since it is an attempt to improve Iterative sampling.

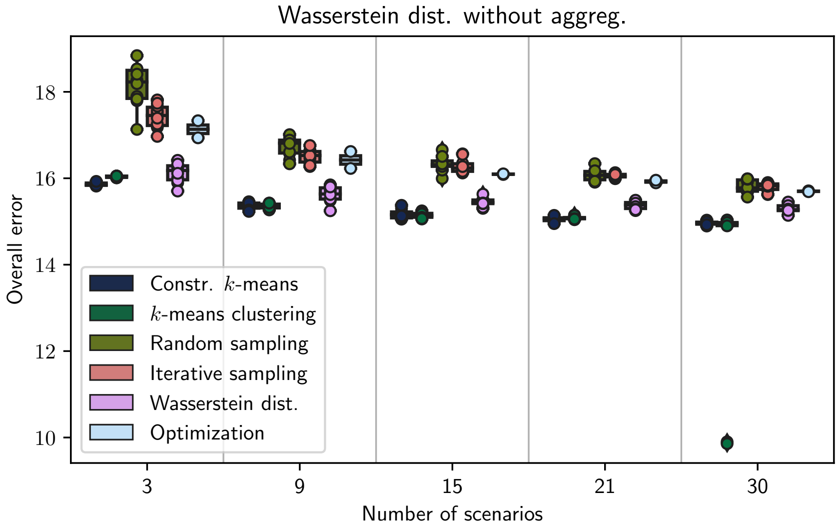

Figure 8: Wasserstein distance of scenarios with non-aggregated data

- Random sampling, as a simple benchmark.

However, this does not necessarily indicate that the three methods are the best suited scenario generation methods for long-term energy models.

We use the trees generated in the previous section, which gives us $2 \times 10 + 2 = 22$ scenario trees. With five scenario sizes, $S \in \{3, 9, 15, 21, 30\}$, this means $22 \times 5 = 110$ different candidate instances, in addition to a deterministic model consisting of one scenario with average values.

The optimization of the TIMES model was carried out with a 2.2 GHz Xeon Silver 4114 Intel processor with 96 GB RAM, using the CPLEX solver with the Barrier solution method. All but one thread, i.e. 19 threads, were available for CPLEX.

3.2.1 Energy-system model runtime

The median runtime of the energy-system model for the candidate instances are presented in Fig. 9. Note that the runtime depends on the selection and customisation of solver and the number of threads used.

Clearly, a deterministic model, with only one scenario, has a significantly lower runtime than the stochastic candidates. For example, the deterministic model uses 2.4 minutes, whereas the runtime for $S = 15$ mediates on 694 minutes and $S = 30$ mediates on 5382 minutes (corresponding to about 3.7 days).

The primary reason for limiting the number of scenarios to $S = 30$ in this study is that we consider the runtime of this case to be on the limit of what is practical for an energy-system modeller. We also experienced that our computer did not have enough memory when we ran the model with $S = 60$.

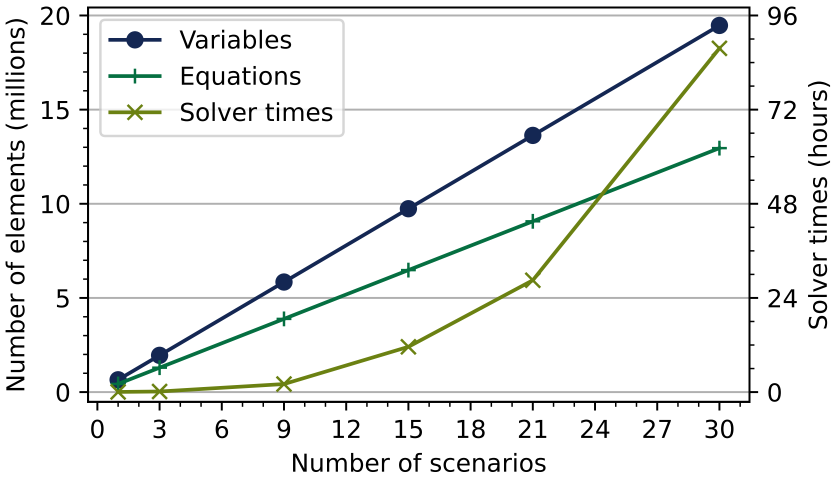

Figure 9: Number of equations and variables in the model and CPLEX solution times, as a function of the number of scenarios.

3.2.2 In-sample stability

The in-sample stability is illustrated in Fig. 10, where the optimal objective function value, represented by total system cost in billion EUR, is plotted for all candidate instances.[2]

The percentage difference among the instances is relatively small since the objective function includes a significant share of costs that are not related to wind and PV power. Nevertheless, the absolute difference between the instances is considerable, where the largest difference between the instances is 4 655 million EUR, observed between one instance with $S = 3$ and another instance with $S = 9$, both generated with Random sampling.

The in-sample stability is improved with a higher number of scenarios for all the three scenario generation methods. Optimization has the lowest deviation of the system cost among the candidate instances and has relative stable solution from 15 scenarios or more. The system costs for Optimization, in million EUR (MEUR) is 951 892 and 952 054, 951 788 and 951 970, and 951 978 and 952 013, respectively for $S = 15$, 21 and 30. Due to that we consider the performance of Optimization to be best among the three methods, we will use the first instance using Optimization with $S = 30$ as our benchmark solution in the following. This implies that we here assume that this benchmark tree is a satisfactory representation of the true distribution.

As expected, random sampling has the highest deviation of system cost among the ten instances of a given scenario size, where the deviation is significantly reduced when the number of scenarios is increased. The difference in energy-system cost between the highest and lowest value, in MEUR, is for example 4232, 2154, and 967 respectively for $S = 3$, 15 and 30. For comparison, the differences between the highest and lowest value

[2]Four of the 111 instances were reported by the solver as optimal, but with infeasibilities after unscaling. We have included these instances in figures and description below.

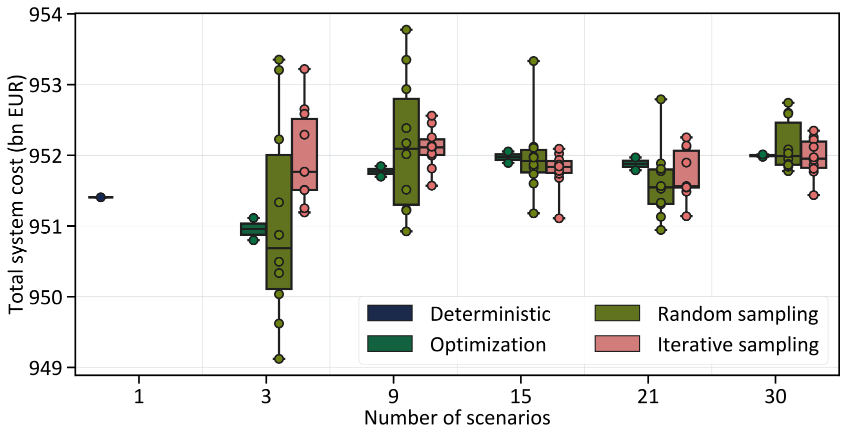

Figure 10: Illustration of in-sample-stability. Total system costs in billion EUR for all candidate instances. The boxes illustrates the 25-75 percentile with median marked as a line within each box for the scenarios with the same S created with the same method. The dots represent the candidate scenario trees, M.

for Iterative sampling is MEUR 2029, 984, and 913 for the same S.

3.2.3 Out-of-sample stability

For the out-of-sample stability test, we fix the investment decisions from each of the candidate instances and solve the model on the benchmark scenario tree ($S = 30$ and *Optimization*).

The out-of-sample stability is depicted in Fig. 11 and Fig. 12 as the optimality gap. The gap is the difference between the total system cost of the benchmark and the candidate system cost. Consequently, the benchmark solution has a gap of zero and is therefore not included in the figures.[3] One of the candidate solutions ($S = 3$ and Random sampling) was infeasible on the benchmark tree. This implies that the investment decisions from this instance provides a solution that is not able to meet the energy-service demand.

Note that the optimality gap is above zero for all instances. This implies that there will be an additional energy-system cost if the investment decisions are based on another scenario-generation method than the benchmark.

For all S, investments made with the scenarios generated with Optimization give the lowest optimality gap. Also, going from $S = 3$ to $S = 9$ gives a notable reduction in the gap for all generation methods. The results demonstrate that using Optimization gives good model results with nine or more scenarios.

3.2.4 Renewable investment strategy

The stability tests are based on evaluating the optimal objective function value and not on the model solution itself. This is because a model can provide almost identical

[3]In 24 of the 111 instances, the solver reported finding the optimal solution, but with infeasibilities after unscaling. These instances are included in the reported results.

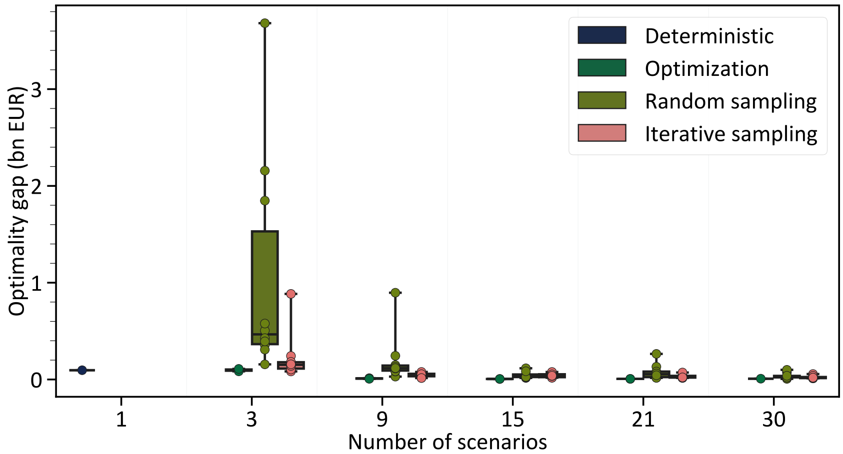

Figure 11: Illustration of out-of-sample-stability. Optimality gap in billion EUR between the benchmark and the benchmark scenario tree with fixed capacities from the candidate instances. The boxes illustrates the 25-75 percentile with median as a line within each box for the scenarios with same S created with the same method. The dots represent the candidate scenario trees, M.

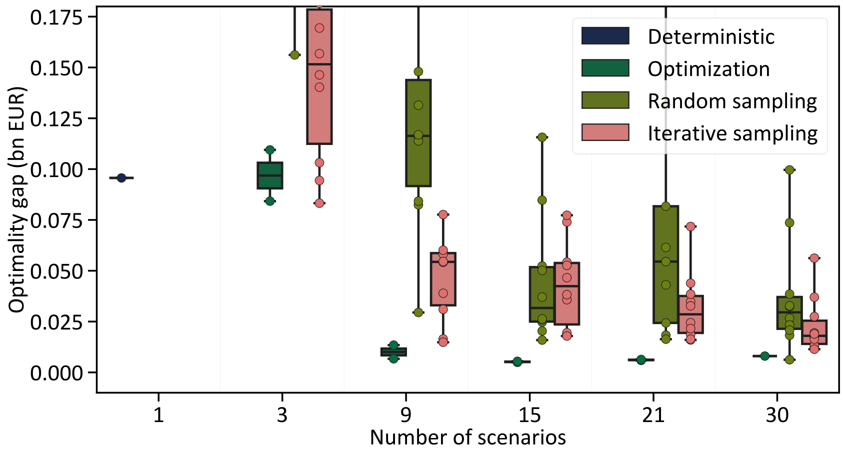

Figure 12: Detail of Fig. 11 in the y-axis between 0 and 0.18 billion EUR.

objective function values and have two different investments strategies (if the objective function is flat). Nevertheless, model insights from energy-system models are primarily based on model solution and not on the corresponding optimal energy system cost.

Fig. 13 shows the PV capacity of all instances in 2040. The results show larger deviation among the instances for the Random sampling generation method compared to the other methods. For example, the standard deviation of PV capacity for Random sampling is 3.9 GW, 0.73 GW , 1.3 GW for $S = 3$, 15 and 30 respectively. There are also significant differences in the expected capacities between Random sampling and the other two scenario generation methods when using few scenarios ($S = 3$ and 9) whereas the differences becomes smaller when using more scenarios. For $S = 30$, the capacity

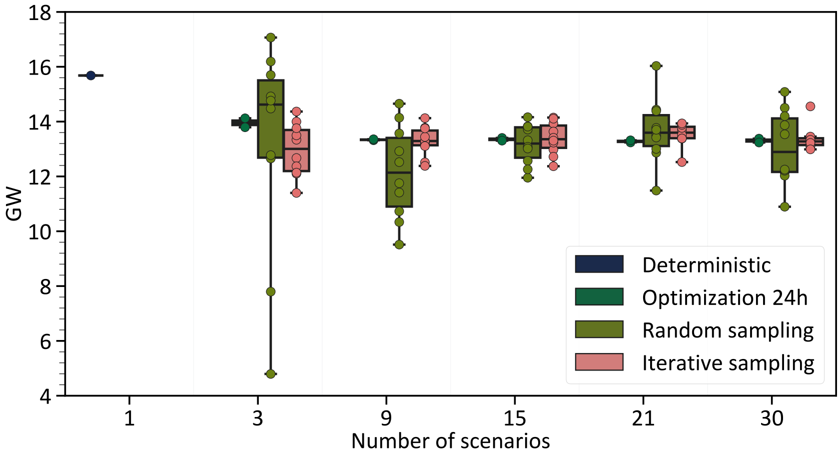

Figure 13: PV capacity in 2040 for all instances. The boxes illustrates the 25-75 percentile with median marked as a line within each box, whereas the dots represent the candidate instances.

ranges from 10.9 GW to 15.1 GW, 13.0 GW to 14.6 GW and 13.2 GW to 13.4 GW for Random sampling, Iterative sampling and Optimization among the scenario trees respectively. The corresponding figure of the wind capacities for the candidate instances are demonstrated in Fig. 14.

The results demonstrate that the design of the stochastic scenarios has larger impact on the investments in PV than for wind power. For example, the difference between the highest and lowest wind capacities is 2.03 GW, both from model solutions considering instances with 3 scenarios created with Random sampling. For PV, the same difference is 12.3 GW, also found in model solutions considering instances with 3 scenarios created with Random sampling. Another notable difference between the capacities for PV and wind is that for all S, the difference between the standard deviation of capacities between Random sampling and Iterative sampling is larger for PV than for wind. For example, for $S = 21$, the standard deviation for wind and scenarios created with Random sampling and Iterative sampling is 0.220 GW and 0.171 GW respectively. The same standard deviations for PV is 1.19 GW and 0.411 GW.

Solution of the deterministic model has less wind power and more PV capacity, compared to the stochastic model instances. For example, using Optimization and 30 scenarios, the deterministic model results has 18.5 % higher PV capacity and 2.7 % lower wind power capacity in 2040 compared to our benchmark solution.

4 Conclusions

This paper presents an approach to analyze variable wind and PV generation by stochastic modelling in long-term energy models, and demonstrates this approach on a TIMES model of the Norwegian energy system.

We observe a satisfactory match when we compare satellite-based data with historical

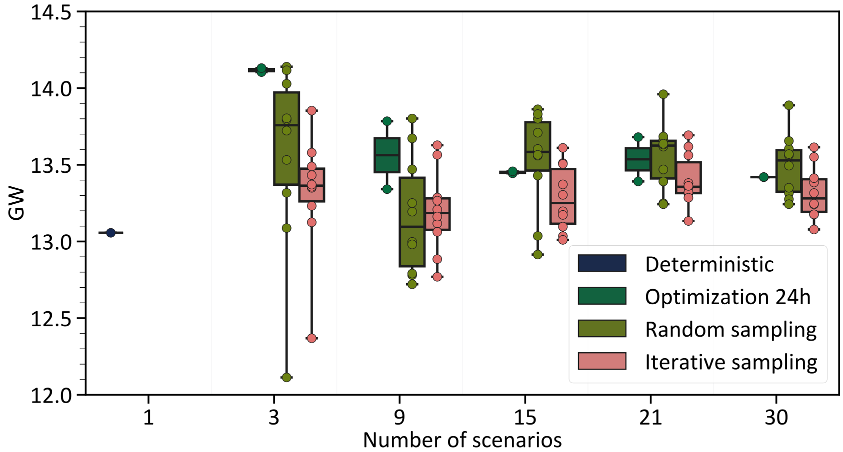

Figure 14: Wind capacity in 2040 for all instances. The boxes illustrates the 25-75 percentile with median marked as a line within each box, whereas the dots represent the candidate instances.

generation for PV but not for wind power. To take this mismatch into account, we modify the data series for wind generation to better consider local conditions for Norwegian wind parks. This implies that using satellite based renewable generation data uncritically in energy-system models, without comparing with production statistics, can give misleading model insights.

The distribution error between the satellite-based dataset and scenarios generated by six different scenario generation methods is evaluated. We observe that the performance ranking among the scenario generation methods differs if a moment-based or a Wasserstein distance is used. For our case, the moment-based iterative sampling method, Iterative sampling, performs best related to moment-based distance, whereas k-means performs best related to the Wasserstein distance. Since the distribution error depends on the method it is measured on, it demonstrates that it is challenging to select the best performing scenario generation method by solely investigating the statistical properties of the scenarios.

To evaluate the performance of the scenario generation methods, related to the quality of the model results, we test in-sample and out-of-sample stability by solving the energy-system model with scenarios generated by three different scenario generation methods. We demonstrate that the runtime of the energy-system model highly depends on number of scenarios, and conclude that the optimal scenario generation method is the method that provides stability by using as few scenarios as possible. Among the scenario generation methods used, the moment-based scenario generation method by integer optimization, Optimization, has the best stability, providing relatively good in-sample and out-of-sample stability with 15 scenarios. Further, we observe that model results with scenarios generated by Iterative sampling is considerably more stable than by using random sampled scenarios.

There are opportunities to further improve the performance of the tested scenario generation methods. Iterative sampling can be improved if the number of samples increases, and since Optimization can control the scenario-probabilities, allowing for scenarios to have different probabilities, would most likely improve the scenario quality from this method.

To increase the knowledge on what is best performing scenario generation method, we propose that future research perform stability tests for more scenario generation methods than what is done in this paper. Further, to provide more robust conclusions on the quality of each scenario generation method, we suggest testing for out-of-sample on several alternative benchmark trees.

The number of scenarios that is required to achieve stability, and the best performing scenario generation method can also depend on the characteristics of the energy system that is modelled. To investigate this assertion, the method of this paper can be applied to other models that cover energy systems that has different characteristics than the Norwegian energy system, that has flexibility through the large hydro reservoirs. Furthermore, the number of scenarios that is required to achieve stability can be influenced by the temporal resolution of the energy-system model. The trade-of between a higher temporal resolution and number of stochastic scenarios that is needed to achieve stability is therefore an interesting topic for further research.

There are several limitations of the stochastic energy system analysis of this paper. First, we only included a stochastic modelling of the short-term uncertainty of wind and PV generation, despite that other types of uncertainty can have an equally large impact on the optimal energy system design. For a Norwegian energy system perspective, the short-term uncertainty of hydro inflow, heat demand and European power prices can have a great impact. On the other hand, it is straight forward to include other uncertain parameters to the methodology that is presented in this paper. However, a challenge by including more uncertain parameters, is to derive a consistent dataset that ensures correlations. A second limitation of the analysis is that the modelling does not include end-use flexibility options, such as stationary batteries, thermal storage and flexible charging of electric vehicles. Suggestions for future research is therefore to evaluate how including end-use flexibility influences the scenario characteristics that is required to provide a solution of good quality.

5 Acknowledgements

The authors are grateful to the Research Council of Norway for funding this work through the following project: "Assessment of the Value of Flexibility Services from the Norwegian Energy System" (Grant number: 268097; `https://prosjektbanken.forskningsradet.no/#/project/NFR/268097`).

References

[1] Seán Collins, John Paul Deane, Kris Poncelet, Evangelos Panos, Robert C Pietzcker, Erik Delarue, and Brian Pádraig Ó Gallachóir. Integrating short term variations of

the power system into integrated energy system models: A methodological review. *Renewable and Sustainable Energy Reviews*, 76:839–856, 2017.

[2] Pernille Seljom and Asgeir Tomasgard. Short-term uncertainty in long-term energy system models—a case study of wind power in denmark. *Energy Economics*, 49: 157–167, 2015.

[3] IEA-ETSAP. Iea-etsap optimization modeling documentation, 2020. URL `https://iea-etsap.org/index.php/documentation`.

[4] Eva Rosenberg, Janis Danebergs, Pernille Merethe Sire Seljom, and Lisa Kvalbein. Documentation of ife-times-norway v1. Report, Institutt for Energiteknikk, 2020. URL `https://hdl.handle.net/11250/2681685`.

[5] Michal Kaut. Scenario generation by selection from historical data. *Computational Management Science*, Online, 2021. doi: 10.1007/s10287-021-00399-4.

[6] Stefan Pfenninger, Adam Hawkes, and James Keirstead. Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews*, 33:74–86, 2014.

[7] Hans F Ravn. The balmorel model: Theoretical background, 2001.

[8] H Ravn. Dynamic power system investment modeling and analysis. In *Ris ø International Energy Conference*, pages 151–155, 2011.

[9] David L McCollum, Volker Krey, Keywan Riahi, Peter Kolp, Marek Makowski, and Bettina Schreck. The IIASA Energy-multi criteria analysis tool (ENE-MCA). Technical report, IIASA, 2012. URL `http://pure.iiasa.ac.at/id/eprint/10142/`.

[10] Mark Howells, Holger Rogner, Neil Strachan, Charles Heaps, Hillard Huntington, Socrates Kypreos, Alison Hughes, Semida Silveira, Joe DeCarolis, Morgan Bazillian, et al. Osemosys: the open source energy modeling system: an introduction to its ethos, structure and development. *Energy Policy*, 39(10):5850–5870, 2011.

[11] Matteo Giacomo Prina, Giampaolo Manzolini, David Moser, Benedetto Nastasi, and Wolfram Sparber. Classification and challenges of bottom-up energy system models - a review. *Renewable and Sustainable Energy Reviews*, 129:109917, 2020.

[12] Gustavo Haydt, Vítor Leal, André Pina, and Carlos A Silva. The relevance of the energy resource dynamics in the mid/long-term energy planning models. *Renewable energy*, 36(11):3068–3074, 2011.

[13] Kris Poncelet, Erik Delarue, Daan Six, Jan Duerinck, and William D'haeseleer. Impact of the level of temporal and operational detail in energy-system planning models. *Applied Energy*, 162:631–643, 2016.

[14] Gustavo Haydt, Vítor Leal, André Pina, and Carlos A. Silva. The relevance of the energy resource dynamics in the mid/long-term energy planning models. *Renewable Energy*, 36(11):3068–3074, 2014.

[15] Ian J. Scott, Pedro M. S. Carvalho, Audun Botterud, and Carlos A.. Silva. Clustering representative days for power systems generation expansion planning: Capturing the effects of variable renewables and energy storage. *Applied Energy*, 253:113603, 2019.

[16] Niina Helistö, Juha Kiviluoma, and Hannu Reittu. Selection of representative slices for generation expansion planning using regular decomposition. *Energy*, 211:118585, 2020.

[17] Hans-Kristian Ringkjøb, Peter M Haugan, and Ida Marie Solbrekke. A review of modelling tools for energy and electricity systems with large shares of variable renewables. *Renewable and Sustainable Energy Reviews*, 96:440–459, 2018.

[18] Peter Kall, Stein W Wallace, and Peter Kall. *Stochastic programming*. Springer, 1994.

[19] Richard Loulou, Gary Goldstein, Amit Kanudia, Antti Lehtila, and Uwe Remme. Documentation for the times model – Part I: TIMES concepts and theory. Technical report, IEA, 2016.

[20] Wouter Nijs and Kris Poncelet. Integrating recurring uncertainties in etsap energy system models. *VITO*, 2016.

[21] Pernille Seljom and Asgeir Tomasgard. Sample average approximation and stability tests applied to energy system design. *Energy Systems*, pages 1–25, 2019.

[22] Hans-Kristian Ringkjøb, Peter M Haugan, and Astrid Nybø. Transitioning remote arctic settlements to renewable energy systems–a modelling study of longyearbyen, svalbard. *Applied Energy*, 258:114079, 2020.

[23] Pernille Seljom, Karen Byskov Lindberg, Asgeir Tomasgard, Gerard Doorman, and Igor Sartori. The impact of zero energy buildings on the scandinavian energy system. *Energy*, 118:284–296, 2017.

[24] Hans-Kristian Ringkjøb, Peter M. Haugan, Pernille Seljom, Arne Lind, Fabian Wagner, and Sennai Mesfun. Short-term solar and wind variability in long-term energy system models - a european case study. *Energy*, 209, 2020.

[25] Héctor Marañón-Ledesma and Asgeir Tomasgard. Analyzing demand response in a dynamic capacity expansion model for the european power market. *Energies*, 12: 2976, 2019.

[26] Stian Backe, Magnus Korpås, and Asgeir Tomasgard. Heat and electric vehicle flexibility in the european power system: A case study of norwegian energy communities. *International Journal of Electrical Power & Energy Systems*, 125:106479, 2021.

[27] Renewables.ninja. Renewables.ninja, 2019. URL `https://www.renewables.ninja/`.

[28] Stefan Pfenninger and Iain Staffell. Long-term patterns of european pv output using 30 years of validated hourly reanalysis and satellite data. *Energy*, 114:1251–1265, 2016.

[29] Iain Staffell and Stefan Pfenninger. Using bias-corrected reanalysis to simulate current and future wind power output. *Energy*, 114:1224–1239, 2016.

[30] Michele M. Rienecker, Max J. Suarez, Ronald Gelaro, Ricardo Todling, Julio Bacmeister, Emily Liu, Michael G. Bosilovich, Siegfried D. Schubert, Lawrence Takacs, Gi-Kong Kim, Stephen Bloom, Junye Chen, Douglas Collins, Austin Conaty, Arlindo da Silva, Wei Gu, Joanna Joiner, Randal D. Koster, Robert Lucchesi, Andrea Molod, Tommy Owens, Steven Pawson, Philip Pegion, Christopher R. Redder, Rolf Reichle, Franklin R. Robertson, Albert G. Ruddick, Meta Sienkiewicz, and Jack Woollen. Merra: Nasa's modern-era retrospective analysis for research and applications. *Journal of Climate*, 24:3624–3648, 2011.

[31] Iratxe González-Aparicio, Andreas Zucker, Francesco Careri, Fabio Monforti-Ferrario, Thomas Huld, and Jake Badger. Emhires dataset: Part i: wind power generation. Technical report, Joint Research Centre, 2016.

[32] Iratxe González-Aparicio, Thomas Huld, Francesco Careri, Fabio Monforti, and Andreas Zucker. Emhires dataset part ii: Solar power generation. Technical report, Joint Research Centre, 2017.

[33] Iratxe González-Aparicio, Fabio Monforti, Patrick Volker, Andreas Zucker, Francesco Careri, Thomas Huld, and Jake Badger. Simulating european wind power generation applying statistical downscaling to reanalysis data. *Applied Energy*, 199: 155–168, 2017.

[34] Luiz Moraes, Christian Bussar, Philipp Stöcker, Kevin Jacqué, Mokhi Chang, and Dirk Uwe Sauer. Comparison of long-term wind and photovoltaic power capacity factor datasets with open-license. *Applied Energy*, 225:209–220, 2018.

[35] David Severin Ryberg, Dilara Gulcin Caglayan, Sabrina Schmitt, Jochen Linßen, Detlef Stolten, and Martin Robinius. The future of european onshore wind energy potential: Detailed distribution and simulation of advanced turbine designs. *Energy*, 2019.

[36] Shan Jin, Jean-Paul Ryan, Sarah M.and Watson, and David L. Woodruff. Modeling and solving a large-scale generation expansion planning problem under uncertainty. *Energy Systems*, 2(3):209–242, 2011.

[37] Francisco D. Munoz and Jean-Paul. Watson. A scalable solution framework for stochastic transmission and generation planning problems. *Computational Management Science*, 12(4):491–518, 2015.

[38] Yonghan Feng and Sarah M. Ryan. Scenario construction and reduction applied to stochastic power generation expansion planning. *Computers & Operations Research*, 40(1):9–23, 2013.

[39] Pernille Seljom and Eva Rosenberg. A scandinavian transition towards a carbon-neutral energy system. In *Limiting Global Warming to Well Below 2° C: Energy System Modelling and Policy Development*, pages 105–121. Springer, 2018.

[40] Stephan Spiecker, Philip Vogel, and Christoph Weber. Evaluating interconnector investments in the north european electricity system considering fluctuating wind power penetration. *Energy Economics*, 37:114–127, 2013.

[41] Michal Kaut and Stein Wallace. Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization*, 3:257–271, 2007.

[42] Esteban Gil, Ignacio Aravena, and Raúl Cárdenas. Generation capacity expansion planning under hydro uncertainty using stochastic mixed integer programming and scenario reduction. *IEEE Transactions on Power Systems*, 30(4):1838–1847, 2015.

[43] Luis Baringo and Antonio. J.l Conejo. Risk-constrained multi-stage wind power investment. *IEEE Transactions on Power Systems*, 28(1):401–411, 2013.

[44] Alexander Shapiro and Tito Homem-de Mello. A simulation-based approach to two-stage stochastic programming with recourse. *Mathematical Programming*, 81(3): 301–325, 1998.

[45] IEA. Energy technology perspectives 2020 - catalysing energy technology transformation, 2020. URL `https://www.iea.org/reports/energy-technology-perspectives-2020`.

[46] Lars Hellemo, Kjetil Midthun, Asgeir Tomasgard, and Adrian Werner. Multi-stage stochastic programming for natural gas infrastructure design with a production perspective. *Stochastic Programming. Applications in Finance, Energy, Planning and Logistics*, 2013.

[47] Stefan Pfenninger and Staffell Iain. Renewables.ninja. `https://www.renewables.ninja/`, 2019. Accessed: 2021-01-06.

[48] DTU Wind Energy and The World Bank. Global wind atlas 3.0, 2019. URL `https://globalwindatlas.info/`.

[49] John D Holmes. *Wind loading of structures*. CRC press, 2018.

[50] Michaël Pierrot. The wind power, 2019. URL `www.thewindpower.net/`.

[51] Lucas Bauer and Silvio Matysik. wind-turbine-models.com, 2019. URL `www.en.wind-turbine-models.com/`.

[52] Per Norgaard and Hannele Holttinen. A multi-turbine power curve approach. In *Nordic wind power conference*, volume 1, pages 1–2. Chalmers, 2004.

[53] NVE. Vindkraftdata, 2021. URL `https://www.nve.no/energiforsyning/kraftproduksjon/vindkraft/vindkraftdata/`.

[54] V.K. Chopra and W.T. Ziemba. The effects of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19 (2):6–11, 1993.

[55] James MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman, editors, *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. University of California Press, 1967.

[56] Stuart P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982. doi: 10.1109/tit.1982.1056489. Prev. published as a Technical Report RR-5497, Bell Labs, 1957.

[57] K.P. Bennett, P.S. Bradley, and A. Demiriz. Constrained k-means clustering. resreport MSR-TR-2000-65, Microsoft, 2000. URL `https://www.microsoft.com/en-us/research/publication/constrained-k-means-clustering/`.

[58] G. C. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming,*, 89(2):251–271, 2001. doi: 10.1007/PL00011398.

[59] Georg Ch. Pflug and Alois Pichler. Approximations for probability distributions and stochastic optimization problems. In Marida Bertocchi, Giorgio Consigli, and Michael A. H. Dempster, editors, *Stochastic Optimization Methods in Finance and Energy*, volume 163 of *International Series in Operations Research & Management Science*, chapter 15, pages 343–387. Springer, 2011. doi: 10.1007/978-1-4419-9586-5_15.

[60] Georg Ch. Pflug and Alois Pichler. *Multistage Stochastic Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, 2014. doi: 10.1007/978-3-319-08843-3.

[61] Georg Ch. Pflug and Alois Pichler. Dynamic generation of scenario trees. *Computational Optimization and Applications*, 62(3):641–668, 2015. ISSN 1573-2894. doi: 10.1007/s10589-015-9758-0.

[62] William E. Hart, Carl D. Laird, Jean-Paul Watson, David L. Woodruff, Gabriel A. Hackebeil, Bethany L. Nicholson, and John D. Siirola. *Pyomo–optimization modeling in python*, volume 67. Springer Science & Business Media, second edition, 2017.