RESEARCH ARTICLE | AUGUST 07 2023

A framework for synthetic power system dynamics

Special Collection: Nonlinear dynamics, synchronization and networks: Dedicated to Jürgen Kurths' 70th birthday

Anna Büttner ; Anton Plietzsch ; Mehrnaz Anvari ; Frank Hellmann

Chaos 33, 083120 (2023)
https://doi.org/10.1063/5.0155971

A framework for synthetic power system dynamics

Cite as: Chaos 33, 083120 (2023); doi: 10.1063/5.0155971
Submitted: 25 April 2023 · Accepted: 17 July 2023 ·
Published Online: 7 August 2023

Anna Büttner,[1,a)] Anton Plietzsch,[1,2] Mehrnaz Anvari,[1,3] and Frank Hellmann[1]

AFFILIATIONS

[1]Potsdam-Institute for Climate Impact Research, 14473 Potsdam, Germany
[2]Fraunhofer Research Institution for Energy Infrastructures and Geothermal Systems, 03046 Cottbus, Germany
[3]Fraunhofer Institute for Algorithms and Scientific Computing, 53757 Sankt Augustin, Germany

Note: This paper is part of the Focus Issue on Nonlinear dynamics, synchronization and networks: Dedicated to Juergen Kurths' 70th birthday.

[a)]**Author to whom correspondence should be addressed:** buettner@pik-potsdam.de

ABSTRACT

We present a modular framework for generating synthetic power grids that consider the heterogeneity of real power grid dynamics but remain simple and tractable. This enables the generation of large sets of synthetic grids for a wide range of applications. For the first time, our synthetic model also includes the major drivers of fluctuations on short-time scales and a set of validators that ensure the resulting system dynamics are plausible. The synthetic grids generated are robust and show good synchronization under all evaluated scenarios, as should be expected for realistic power grids. A software package that includes an efficient Julia implementation of the framework is released as a companion to the paper.

Power systems will be reconfigured as conventional generation is replaced by renewable energy sources (RES). The latter are often connected to the grid via inverters. The exact dynamical behavior and, especially, the stability of these inverter-based networks is not well understood. Thus, the availability of adequate synthetic power system models remains limited. However, it is vital to simulate future power grids to verify that this transformation does not result in undesired effects and blackouts. We introduce a framework for realistic synthetic power systems that can be used to study collective dynamical effects. We combine established methods such as realistic grid topologies and active power set points. This framework opens new avenues for predicting the stability of future power grids using advanced techniques such as graph neural networks.

I. INTRODUCTION

Synthetic power grids have become an important tool for studying the dynamics of power systems. Traditionally, most dynamical simulation studies in the engineering literature were performed using benchmark test cases, such as the "New England" IEEE 39-Bus System[1] or the IEEE Reliability Test System-1996.[2] The advantage of this approach is that models and parameters can be specified in great detail and the test cases are, therefore, highly realistic. Further, the use of standardized benchmark test cases guarantees comparability of different dynamic models and analytical methods. However, for many emerging research questions, this approach can be quite limiting and the use of automatically generated synthetic grid models might be beneficial. This is, for instance, the case when the power system in a specific region should be studied but the detailed topology and parameters of the real grid are not publicly accessible. Often there is enough data or knowledge available to generate a synthetic grid that resembles the main properties of a real grid to a reasonable degree. An example is the algorithm by Birchfield *et al.*[3,4] that generates realistic transmission network topologies from spatial load distributions based on geographic population data. The algorithm is expanded in Ref. 5 to also enable transient stability analysis of the synthetic power grids. Besides the transmission system, synthetic grids are also required for studying mid- and low-voltage grids as their exact structure is often unknown.[6] For German medium and low-voltage grids, the *DingO* model[7] is an extensive and well-documented option to generate topologies[7] and supply and demand distributions.[8] *DingO* is part of the larger research project

open eGo and is open-source software that uses freely available data.

Another important use case for synthetic power grid models is to generate large data sets of synthetic test cases that can be used to investigate the system dynamics with methods of machine learning.[9,10] A number of studies have shown that the network topology of grids has a direct influence on their dynamic stability.[11–17] However, most of these studies are based on simplistic component models for synchronous machines. Furthermore, unrealistically homogeneous parameters for the nodes and lines are used, even though it is known that heterogeneities, for example, due to the heterogeneous line lengths that directly impact the admittances, play an important role.[18] Graph-Neural-Networks have been shown to be a powerful method that could potentially extend these stability analyses to more realistic power grid models.[10,19,20] The training of such neural networks requires large data sets of realistic grids that are, for example, generated by a synthetic grid model.

Finally, synthetic grid models will be crucially important for the investigation of the dynamic effects of future power grids. Within the next decades, the power system will undergo a fundamental transformation as new transmission infrastructure is built and conventional machines are replaced by renewable energy sources (RES). A major challenge is that the exact dynamical behavior of generation units is widely unknown as renewable generation units are connected to the grid via inverters with various control schemes. In order to maintain stability in such inverter-based grids, a certain share of these controls must be grid-forming. Today, most RES are still equipped with grid-following control schemes and, hence, there is a lack of practical knowledge on the collective dynamical behavior of a large number of grid-forming generation units. It is, therefore, of great importance to do simulation studies of these systems to ensure that new technology being integrated into the grid does not lead to unexpected collective effects and blackouts.[21] Unfortunately, there is a lack of both benchmark test cases and synthetic power grid models for studying such inverter-based grids. In this work, we enable the possibility of modeling power grids with high shares of inverter-based generation units. For this, we bypass the problem that the exact dynamical models of such systems are still uncertain by using a technology-neutral model, introduced in Ref. 22, that has been shown to reproduce the behavior of a large class of different inverter controls. However, we also point out open research questions for improving the modeling of future power grids.

In this paper, we present a modular framework for generating synthetic grids that are suited for dynamic power system studies. We give an overview of all necessary steps from the generation of grid topologies, to the definition and parametrization of component models and the calculation of the steady state. The paper is accompanied by a software repository that provides an implementation of all algorithms described in this paper. Our approach is modular in the sense that users can easily adapt each step in the grid generation process to their own needs, e.g., by providing their own specific grid topologies or by using different dynamic models for the generating units in the system. We focus on extra high voltage (EHV) level transmission grids, which in the continental European transmission grid includes the 380–400 and the 220 kV voltage levels. Collective dynamical effects are traditionally studied in the highest grid layer,[23] which is why we can rely on a comprehensive foundation there. In principle, the approach presented here can be extended to all grid layers.

The framework is designed to be capable of efficiently generating large numbers of synthetic grids with very limited input data. At the same time, the component models and parameters have a comparatively high level of realism: Generator and inverter models feature voltage dynamics, the active power production, and demand are heterogeneous and the parametrization of line admittances is according to data of the German transmission grid. The framework is, therefore, well-suited for applying machine learning methods, e.g., to predict dynamical stability from the structural properties of the grid.

The main focus of the framework presented in this work is to enable the transient analysis of the collective dynamics of future power grids and especially their transient stability. The target audience of this work are interdisciplinary researchers without a strong background in power systems engineering. This framework offers them a tool to generate synthetic grids and their dynamics which finds a balance between simplicity and tractability and realism. The underlying simulation software PowerDynamics[24] that is used for this framework focuses on transient stability analysis and on elector-mechanical phenomena and is applicable for simulations in the range of seconds to several minutes.

II. SYNTHETIC POWER GRID FRAMEWORK

For this project, we have chosen a framework to structure the synthetic power grid generation process. A framework in software development is defined as a semi-complete code basis that provides a reusable structure to share among applications.[25] Users can integrate the framework into their own software and extend it to include specifically needed functionalities. The modularity and expandability of frameworks are needed for this project as researchers are interested in various properties and effects common in power grids which can be included in the framework over time. Furthermore, as more information on the structure of power grids under renewables becomes available it can easily be included in the existing software. As typical for frameworks we have developed a default structure that can be employed immediately by users. As the framework is modular each step can be interchanged as long as it adheres to the general structure. The default structure of the framework is shown in Fig. 1. In the default structure, the first step is to generate a topology or network structure for the synthetic power grid. Then, active power set points for the nodes in the network are defined. The next step is to specify the node and line models in order to populate the networks with dynamics. Then, an operation point, that fulfills certain stability criteria, is determined. In the last step, we validate the synthetic grids and ensure that the dynamic network properties are similar to those of real power grids that are carefully planned.

Most of the steps presented here have been used and validated individually in research projects before, however, they are now, for the first time, combined as a comprehensive package that is available for further research. Particularly, it is the first step toward a synthetic model of future power grids with high integration of RES. Each section contains a summary of a step in the framework as well as a critical analysis of the state-of-the-art. Secs. II A–II E, we give an outlook and show which additional work could be done to improve

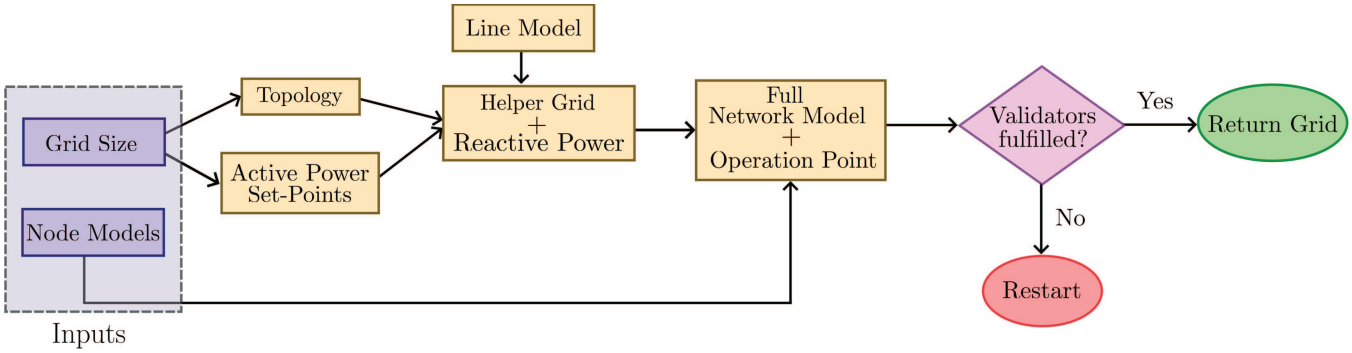

FIG. 1. The default structure of the software framework. The user only needs to input the dynamic model of the nodes and the size of the power grid. Further steps, such as the generation of the topology and the power flow in the network, are performed automatically. Before the power grid is returned by the software, its behavior is validated to fulfill the stability criteria of real power grids. This flow chart only shows the currently implemented default structure, however, as the framework is modular further options can be added over time.

the model, particularly for the representation of future power grids.

For the analysis of the resulting grids, we also provide stochastic models that characterize fluctuation processes that are typical at the timescale of interest.

A. Grid topology

The default topologies in our framework are generated using the random growth algorithm introduced in Ref. 26. We choose this model as it is conceptually straightforward to generate a large number of interesting and plausible topologies and as it has little computational complexity, which is convenient for generating large ensembles of synthetic test cases. However, it is at the conceptual end of the synthetic grid spectrum. If the interest is to study dynamics on more realistic topologies, other models should be employed. Two examples of algorithms that generate such realistic network topologies are introduced in Refs. 3 and 7. The advantage of these models is their high degree of realism and their ability to depict real-world electricity grids instead of only reproducing statistical properties. However, both algorithms are only able to produce a limited number of topologies, which hinders their usage in certain machine learning projects that aim at studying the dynamics stability such as in Refs. 10 and 19.

The random growth algorithm[26] generates synthetic networks that resemble real-world EHV power grids with respect to the exponentially decaying degree distribution and the mean degree. The algorithm includes first an initialization phase, where a spatially embedded minimum spanning tree is generated, and then a growth phase. The growth phase makes use of a heuristic target function for the trade-off between the total line length, which determines the costs, and the smallest number of edges that would need to be removed to disconnect the grid into two parts, which influences the redundancy.

The default parameters of the growth algorithm have been set to $[N_0, p, q, r, s] = [1, 1/5, 3/10, 1/3, 1/10]$, as employed in Ref. 16, where N_0 is the initial number of nodes in the minimum spanning tree, p, q are the probabilities for generating a new redundant line, r is the exponent for the trade-off between redundancy and cost, and s is the probability of splitting an existing line.

Since distribution grids typically exhibit rather different network structures (mostly radial and ring topologies[7]), these parameters have to be adapted when the growth algorithm should be used for modeling lower voltage levels.

For the default step, we assume that there is no correlation between the grid topology and the positioning of generation units in future grids. We, thus, assume that the transmission system topology will remain very similar to today, even if the position of generation units will be correlated to the renewable energy potentials and the location of the generation, thus, changes. This may not be entirely realistic and future studies should consider that the grid will be expanded and adapted to the new supply sources. However, such changes are expensive and time-consuming[27] and, thus, likely to be limited. To properly incorporate these aspects, a synthetic geographical model, potentially incorporating economic optimization, such as in Ref. 28, is needed.

B. Active power distribution

In order to correctly represent the dynamics of the power grid, a realistic distribution of power in the grid is required. For this purpose, the *ELMOD-DE*[29] data set, an open-source spatially distributed, nodal dispatch model for the German transmission system is consulted. This data set has been chosen as it contains real data on demand and generation and has been accumulated from reliable sources such as the German Transmission Operators and the European Network of Transmission System Operators for Electricity (ENTSO-E). The *ELMOD* data set represents the current load and capacity distribution, which means that RES are still in the minority. The analysis shown here is suitable for the distribution of active power in synthetic grids, which should represent the status quo as most buses are either generation-heavy or load-heavy. Following Ref. 30, which also analyses the data set, we examine the net power ΔP at each node given in the data set.

The *ELMOD* data set includes a time series for the total demand P_{tot} in all of Germany. The demand is distributed to the individual

nodes by introducing the nodal load share ls_m, which specifies the proportion of the consumption of a node m from the total demand P_{tot}. It is distinguished between two different types of load scenarios, off-peak and on-peak. Egerer[29] define on-peak and off-peak as the highest and lowest load level meaning the maximum and minimum of P_{tot}, respectively. The data set gives the load shares ls_m for both scenarios the off-peak and on-peak. Following the calculations of Ref. 30, we will always work with the off-peak scenario. However, simulating the on-peak scenario can easily be achieved by using the on-peak load shares and repeating the following calculations. The consumption at a node $P_{con,m}$ is then given by

$$P_{con,m} = P_{tot} \cdot ls_m. \tag{1}$$

The *ELMOD* data set includes the installed capacity for each generation unit c_m^k, which is the maximum power output the unit k connected to node m can produce. As multiple power plants can be connected to a node m, the nodal capacity C_m is given by the sum of all capacities at the node $C_m = \sum_k c_m^k$. Typically, the full capacity of a generation unit is not available. In addition to the approach by Taher *et al.*,[30] we also include the availability factors a^{tech} for each technology during the off-peak scenario. The nodal availability A_m is then given by

$$A_m = \sum_k c_m^k \cdot a^{tech}. \tag{2}$$

The total available power is defined as $A_{tot} = \sum_m A_m$. As there are no data about how much power each node generates at a given time point we follow the approach given in Ref. 30 and reduce the nodal availability A_m by the factor $x = \frac{P_{tot}}{A_{tot}}$, such that generation and consumption are balanced. The nodal generation $P_{gen,m}$ is, thus, given by $P_{gen,m} = A_m \cdot x$. Finally, we can define the net nodal power ΔP_m as

$$\Delta P_m = P_{gen,m} - P_{con,m}. \tag{3}$$

Figure 2 shows the distribution of the net nodal powers ΔP as a histogram. It can be seen that the distribution is bimodal and asymmetric and that the power generation is heavy-tailed. The heavy tail in the power distribution can be explained by the structure of today's power grid where the power is mostly produced by a small number of large generators. In the *ELMOD* data set, 301 nodes are classified as net consumers, while only 137 are net generators. For a future RES-heavy scenario, the capacities and availabilities should be replaced with a model for the deployment of wind and solar renewable resources.

Following Ref. 30, the active power P of each node is sampled from a bimodal distribution, given by

$$p(P) = \frac{1}{2\sigma\sqrt{2\pi}} \left(\exp \frac{(P - P_0)^2}{2\sigma^2} + \exp \frac{(P + P_0)^2}{2\sigma^2} \right), \tag{4}$$

in this work, we will use $P_0 = \overline{\Delta P_{380}} \approx 131\,\text{MW}$.

The topologies used here mimic the extra high voltage 380 kV transmission grids. All following calculations are performed in a Per-Unit-System (p.u.), meaning that an appropriate base power P_{base} and base voltage V_{base} have to be chosen. As this work only examines the highest voltage layer of the grid, the base voltage is simply chosen as $V_{base} = 380$ kV. To define the base power for the 380 kV level, we extract all nodes that are connected to 380 kV lines

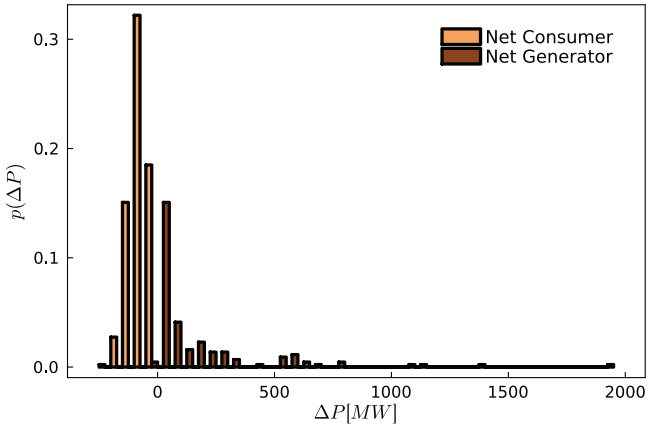

FIG. 2. Histograms of the net nodal generation and consumption in the *ELMOD-DE*[29] data set during the off-peak scenario. The distribution is bimodal and asymmetric. The power generation shows a heavy tail with a small share of net producers that generate more than 750 MW.

and calculate the mean $\overline{\Delta P_{380}} \approx 131\,\text{MW}$. Based on the available data, we choose $P_{base} = 100$ MW as the base power for the synthetic power grids.

For this work, we will adopt the bimodal model which was introduced in Ref. 30. How this distribution will change due to the increasing share of RES but also changing consumption remains an open research question. A promising possibility is to base the distribution of active power supply on the renewable potentials of geographical areas. For this purpose, established software packages, such as atlite,[31] could be consulted. For the consumption side, new sectors with additional loads will be connected to the electric grid, for example, electric cars or hydrogen production.

Furthermore, it should also be taken into account that the set points for the power change in the grid over time due to the evolution of the demand over the day and year. Typically, these set points are updated every 15 min based on a cost optimization procedure. It would be valuable to study a grid and its dynamics under different load scenarios. Moreover, the demand is not constant between two dispatch times, but fluctuates, for example, studied in Ref. 32. In Sec. III B, we will apply the model for realistic demand fluctuations, which has been derived in Ref. 32, to our power grids. Future work could also consider that the generation is typically distributed via an optimal power flow calculation to find the optimal dispatch. Additionally, it could be considered that the demand is typically not randomly distributed in space but centered around population centers or industry-rich regions. This clustering could result in a higher loading of lines that connect production- and consumption-heavy centers, as we can, for example, see in the German power grid.[33]

C. Power grid model

On the most abstract level, we will mathematically describe power grids as systems of differential-algebraic equations (DAEs). The algebraic constraints are most commonly introduced via the load models, but can also appear when several generation units are

present at a bus. Explicit DAEs are defined as

$$\dot{x} = f(x, y), \tag{5}$$

$$0 = g(x, y), \tag{6}$$

where Eqs. (5) and (6) represent the differential and algebraic equations, respectively. The vector x holds the differential variables, whose derivatives appear in the DAE, while the vector y gives the algebraic variables, whose derivatives do not appear.

The specific models for the nodes and lines as well as for the networks are introduced in Secs. II C 1 and II C 2.

1. Node models

Our synthetic grids will consist of grid-forming components, for example, power plants and novel types of inverters that contribute to grid stability and components without grid-forming capabilities, such as loads or grid-following inverters, that have to rely on an already stable grid. For this work, we have decided to use elementary nodal models to depict components with and without grid-forming abilities that are able to cover a large range of dynamical actors.

In this work, PQ-buses[34] are used to represent the components without grid-forming behavior. The PQ-bus locally fixes the active and reactive power of node m,

$$0 = (P_{set,m} + \mathrm{i}Q_{set,m}) - v_m \cdot i_m^*, \tag{7}$$

where $P_{set,m}$ and $Q_{set,m}$ are the active and reactive power set points of the node, and v_m and i_m are the complex voltage and current of node m, which completely describe the physics of a balanced three-phase AC system.[22] The model can depict either loads or sub-networks of consumers and RES that are connected to the grid via grid-following inverters. The PQ-bus (7) is a constraint equation as given in Eq. (6) and forces us to use the DAE description of the power grids.

To represent grid-forming components, we use the normal-form, a technology-neutral model for grid-forming actors, that has been introduced in Ref. 22. It has been shown that various models of grid-forming components, such as droop-controlled inverters[35] and synchronous machine models,[36] can be expressed by the normal form. The normal form has been validated by numerical simulations and lab measurements of a grid-forming inverter so far, and work to identify normal form parameters for a wide range of grid-forming actors is ongoing. A normal form at node m with a single internal variable, the frequency ω_m, is given by

$$\begin{aligned} v_m &= v_m v_m^*, \\ \dot{\omega}_m &= A^{\omega,m} + B^{\omega,m}\delta\omega_m + C^{\omega,m}\delta v_m + G^{\omega,m}\delta P_m + H^{\omega,m}\delta Q_m, \\ \frac{\dot{v}_m}{v_m} &= A^{v,m} + B^{v,m}\delta\omega_m + C^{v,m}\delta v_m + G^{v,m}\delta P_m + H^{v,m}\delta Q_k, \end{aligned} \tag{8}$$

where v_m is the complex voltage. δP_m and δQ_m represent the difference between the active and reactive power to the set points. δv_m is the difference in the squared voltage magnitude v_m to the squared voltage set point. The other coefficients are the modeling parameters that capture all the differences between the various models the normal form can represent. The parameters $A^{\omega,m}$ and $A^{v,m}$ are zero when the system is, as in our case, defined in the co-rotating reference frame. In the normal form, all structural differences between models are absorbed in the parametrization.

The free parameters for the normal form can be gathered by approximating other models; moreover, it is also possible to derive them from experimental data, which has also been performed in Ref. 22 for a specific type of inverter in a lab. For the example provided in this work, we will use a normal form approximation of a droop-controlled inverter[35] whose parameters can be derived analytically.

The exact ability of the normal form to cover all needed dynamics is a subject of current research. Future work will include measurements on different types of inverters and deriving the parameters of the normal form from the data. This is a crucial step to study the dynamics and stability of realistic future power grids, which will consist of a variety of interacting grid-forming inverters.

In addition, we use a slack bus[34] for the load flow calculation. The slack bus locally fixes the voltage v_m of node m,

$$0 = v_{set,m} - v_m, \tag{9}$$

where $v_{set,m}$ is the set point voltage. The voltage magnitude $|v_{set,m}|$ of the slack is typically set to 1 p.u. and its voltage angle is $\phi_m = 0°$. The slack bus is not included in the resulting dynamic synthetic power grid. It is only an ancillary component that is used as the reference for all other buses in the system while solving the load flow problem, as described in Sec. II D. The active and reactive power of the slack bus are free to change to compensate for the power imbalance in the network. Therefore, it is assumed that the slack bus has a large amount of energy stored, which can be released quickly. The slack bus is typically considered to be a large power plant or battery, a connection point to a higher grid layer, or another part of the power system which is not modeled explicitly.

2. Line model

For this work, the Pi-Model, see, for example, Ref. 37, is used. In the Pi-Model, the impedance $Z_{km} = \frac{1}{Y_{km}}$ is placed in the center of the line. The capacitance between the line and the ground is also taken into account by introducing the shunt admittance $Y_{sh,km}$ which is placed, in parallel, at both ends of the line. The current on the lines connecting node k and m is then given by[37]

$$i_{km} = Y_{km}(v_k - v_m) + Y_{sh,km}v_k, \tag{10}$$

$$i_{mk} = Y_{km}(v_m - v_k) + Y_{sh,km}v_m, \tag{11}$$

where Y_{km} is the admittance of a line connecting node k and m and $Y_{sh,km}$ is the shunt admittance. v_k and v_m are the complex nodal voltages. Combining the nodal and line models, we obtain the full network model. The current injected at node k is given by

$$i_k = \sum_m i_{km}, \tag{12}$$

and the power flow in the network is defined as

$$S_k = v_k i_k^* = P_k + \mathrm{i}Q_k, \tag{13}$$

where S_k is the apparent power at node k and P_k and Q_k are the real and reactive power injected at k, respectively.[34]

TABLE I. Standard overhead line parameters according to Ref. 38 for the typical number of cables and wires.

Voltage level	R (Ω/km)	X (Ω/km)	C_{sh} (nF/km)
380 kV	0.025	0.25	13.7

The impedance and shunts are calculated according to the *dena* model of standard 380 kV overhead power lines[38] given in Table I. The model includes the resistance and reactance per unit length R and X, respectively. Furthermore, the model includes the shunt capacitance per unit length C_sh, which is used to calculate the shunt admittances.

As X is specified for the nominal frequency of 50 Hz, which is why we use a static line model here. The total admittances are calculated according to

$$k_c = \frac{c}{c_t}, \tag{14}$$

$$k_w = \frac{w}{w_t}, \tag{15}$$

$$Y_{km} = \frac{k_c k_w}{(R + jX) \cdot l_{km}}, \tag{16}$$

$$Y_{sh,km} = \frac{-(j\omega C_{sh}) k_c k_w}{2} \cdot l_{km}, \tag{17}$$

where l_{km} is the line length in kilometers. For consistency, we fix the grid frequency ω, in the shunt admittance $Y_{sh,km}$, to the nominal frequency. The coefficients k_c and k_w define the ratio between the typical number of cables c_t and wires w_t and the actual numbers of cables c and wires w in the line.[39] The typical numbers of cables and wires are 3 and 4, respectively, for transmission lines in the 380 kV level in Germany.[39] In the default version of the algorithm, we assume that all transmission lines have the typical number of cables and wires. In Sec. II E 4, we introduce an additional step in the algorithm where probabilistic power flow scenarios are considered. The line capacities are increased, by adding new cables to existing lines, if a load scenario leads to an overload.

To calculate the line properties, the lengths of the transmission lines are needed. As the model of Ref. 26 generates an embedded topology, but does not provide a spatial scale, we need an additional step to determine the spatial scale. This is done by requiring that the line lengths of the synthetic grids resemble the line lengths of real EHV grids.

The line lengths l_{mk} in kilometers are obtained by converting the Euclidean distances d_{mk} of the lines, which are generated by the random growth model.[26] The conversion factor c_l is given by the mean length $\langle l \rangle$ of overhead lines in the extra high voltage (EHV) level, that concerns voltages equal or greater than 220 kV, divided by the mean euclidean distance $\langle d \rangle$,

$$c_l = \frac{\langle l \rangle}{\langle d \rangle}, \tag{18}$$

TABLE II. Comparison of transmission line lengths between different models. The values for the synthetic grid were calculated by generating 10 000 different topologies. The mean line length is given by $\langle l \rangle$, the standard deviation of the line length is σ_l, and the minimal line length by $l_{\min}$.

	$\langle l \rangle$ (km)	σ_l (km)	$l_{\min}$ (km)
SciGRID[39]	37.13	36.59	0.06
ELMOD-DE[29]	40.98	35.54	0.42
Synthetic grids	37.13	34.6	0.06

$$l_{mk} = c_l \, d_{mk}. \tag{19}$$

Additionally, we used the shortest line $l_{\min}$ in the EHV level as a threshold. The admittances of lines that are shorter than $l_{\min}$ are set to the threshold impedance of the shortest line.

The mean line length was determined from the *SciGRID* data set,[39] which consists of openly available geographic data of the German power grid. At the time of the creation of the data set, the coverage of the EHV level in Germany was around 95%,[39] which thus offers an excellent basis for such a study.

The *ELMOD* data set[29] also offers a network topology that is based on network plans by the transmission system operators (TSOs) and OpenStreetMap data. Since the data in *SciGRID* are better documented and the study deals much more intensively with the network topology, we base our transmission line lengths on *SciGRID*. Still, for completeness, we will also analyze the data from *ELMOD*. A comparison between *SciGRID*, *ELMOD*, and our synthetic grids, which are based on *SciGRID*, is given in Table II.

In Table II, it can be seen that the mean line length, as well as the standard deviation of the line length of *SciGRID* and *ELMOD*, matches well. Furthermore, it can be seen that our synthetic grid line length shows a standard deviation that matches the *SciGRID* as well as the *ELMOD*. The most significant difference between the two data sets is the minimum line length $l_{\min}$, which is about 400 m in *ELMOD* and about 60 m in *SciGRID*. For the reasons that were stated above, we have adopted $l_{\min}$ from *SciGRID*.

Future work would also include not only analyzing the mean and standard deviation of the length but also matching the distributions of line lengths (see Appendix A). This goes beyond the random growth algorithm[26] which is currently used and would require an algorithm that considers line lengths, node locations, and a spatial embedding. A preliminary study[40] on extending the algorithm which uses different node positioning rules has been performed but it does not deal with recovering the correct line length distribution.

D. Operation point and reactive power

Finding a stable operation point for synthetic power grids is challenging as power systems are generally non-linear and multistable. The AC load flow has no guarantee for convergence. Even if it converges only the synchronous fixed points whose voltage magnitudes are all close to 1 p.u. are physically meaningful for power grids. Securing the voltages to be close to their nominal values is a difficult task to accomplish. Typically, the reactive powers of the nodes are adjusted to control the voltage magnitudes in the power

grid.[34] However, for many synthetic grid models, there is no prior information about the reactive power flow.

Reactive power planning is considered to be one of the most intricate problems in power grid planning.[41] The review article[41] gives an excellent overview of the objectives and constraints that are considered in reactive power planning. Instead of implementing one of the complex established models presented in Ref. 41, we use a straightforward method to solve the reactive power flow. We employ the voltage stability objective, which is also a standard objective according to Ref. 41, and assume that it has to be met perfectly. This requirement uniquely determines the reactive powers at the nodes.

We generate an ancillary power grid with the same topology and line models as the full power grid. The ancillary power grid consists of PV buses where all nodes are constrained to have voltages magnitudes of $V_m = 1p.u.$ and the same active power that they generate in the actual power grid. One of the nodes is randomly turned into the slack bus (9) of the system that accounts for any power imbalances, for example, due to line losses. The reactive powers of the ancillary grid are found by using the power flow calculation of `PowerModel.jl`[42] and a root-finding algorithm to find a steady state. The operation point of the ancillary grid is used as the initial guess for the operation point search of the actual grid.

As the synthetic grids generated in this work have less than 10 000 nodes, our approach still leads to feasible power flow solutions. Once the grids become bigger, a more in-depth reactive power flow planning algorithm, such as in Ref. 43, will be needed to find feasible operation points.

E. Validators

Real-life power grids are planned carefully to lead to stable operations. Synthetic processes can never fully capture this planning stage. Instead, we use a rejection sampling approach. Synthetic power grids whose dynamics do not satisfy the stability properties of real-life power grids are rejected. In this section, we introduce a set of validators that review the stability of the synthetic power grids in their operation point. To assess our default settings, we generated a set of synthetic networks with different sizes and studied the number of rejections. We generated power grids ranging from 100 to 1300 nodes with a step size of 25 nodes. For each grid size, we generate 100 power grids and can report that no grid was rejected.

1. Voltage magnitude

First, we verify that the nodal voltage magnitudes fulfill the standard of the EN 50160 report.[44] The report specifies that the average 10 min root mean square voltage has to stay within the bounds of $\pm 10\%$ for 95% of the week. We assure this by validating that all nodal voltage magnitudes are $V \approx 1p.u.$ in the operation point. If the set points of the system and the parametrization have been chosen properly, the voltage condition should be fulfilled. Even if the reactive power is chosen to ensure a stable power flow with good voltage magnitudes, incorrectly specified control dynamics or machine parameters, can still lead to a violation of the voltage conditions in the operating point. Thus, even in this case, the verification of the voltage condition is still essential in order to catch such unrealistic parametrizations.

2. Line loading stability margin

In a stable operation of the power grid, no line is overloaded. There are different thresholds for the allowed loading of a transmission line. In this work, we focus on the threshold which is determined by the stability margin and depends on the physically possible limit of the line $P_{\max}$.

The power flow transferred over a line connecting node m and k, neglecting the reactive power flow and line losses, is given by

$$P_{mk} = \frac{v_m v_k}{X_{mk}} \sin(\theta_{mk}), \tag{20}$$

where v_m and v_k are the nodal voltage magnitudes, X_{mk} is the line reactance, and θ_{mk} is the difference in the voltage angles of node m and k. The transferred power becomes maximal when $\theta_{mk} = \frac{\pi}{2}$. Thus, the physically possible limit of the line is $P_{\max} = \frac{V_m V_k}{X_{mk}}$. To assure a stable power system, transmission lines are operated well below this limit and the so-called stability margin sm is introduced.[45] The transferred power of a line P_{rated} must, therefore, be below a threshold given by $P_{rated} \leq P_{\max}(1 - sm)$. In this study, we choose $sm = 0.3$ as suggested in Ref. 45. If any line loading in our power grid violates this threshold, we reject the power grid.

3. Small signal stability analysis

Since the grids we consider in this work are described by DAEs, we cannot simply study the eigenvalues of the Jacobian in the equilibrium to determine the linear stability of the system. Instead, we perform a small signal stability analysis for DAEs according to Ref. 46.

In this approach, the eigenvalues of the so-called reduced Jacobian, or state matrix J_{red} are examined. The reduced Jacobian is set up by decomposing the full Jacobian matrix J into the following blocks containing the partial derivatives:

$$J = \begin{bmatrix} \partial_x f & \partial_y f \\ \partial_x g & \partial_y g \end{bmatrix}, \tag{21}$$

where for a function h and a variable z, the matrix of partial derivatives of h with respect to z is given by $\partial_z h$. Again, the differential equations and the algebraic equations are given by f and g, respectively.

Following Ref. 46, the reduced Jacobian is defined as

$$J_{red} = \partial_x f - D, \tag{22}$$

$$D = \partial_y f \left(\partial_y g\right)^{-1} \partial_x g, \tag{23}$$

where D is the degradation matrix. The eigenvalues of J_{red} can be examined as usual again, meaning that power grids whose eigenvalues of J_{red} have positive real parts are classified as linearly unstable. Power grids whose operation point is linearly unstable would not exist in reality and, therefore, have to be rejected before any further investigations are performed.

4. Probabilistic capacity expansion

So far we have only assured that the synthetic power grids are stable under a single power set point that was drawn from the probability distribution (4) or any other source. However, in real power

grids, the set points are updated regularly, e.g., in Germany, a new demand plan is implanted every 15 min. Therefore, it is important to also verify the stability of the grid under different set points. In principle, all validators can be applied to an ensemble of set points. In this work, we only focus on the capacity of lines, as this is the most directly affected by the demand, and assure that there is always enough line capacity to cover the expected load cases.

We sample completely new set points from the bimodal distribution (4) but double the mean power P_0 in order to study the system under more stress. A more realistic analysis of high-stress power flow scenarios would require an extensive investigation of the expected set points and is, therefore, beyond the scope of this paper.

For each new scenario, we calculate the load flow in the grid and then analyze the line loading as given in Sec. II E 2. If a line is overloaded, we add three additional cables to the line to increase its admittance as in Eq. (16). This approach is repeated for N different scenarios. So far no new cables were added for all performed simulations. This is to be expected since, in the *SciGRID*[39] data set, more than 90% of the EHV transmission lines have the typical number of cables. It is nevertheless important to validate the grid under different load scenarios to assure its stability. Furthermore, this capacity evaluation could become important once more realistic load scenarios are evaluated, which in the future could include the weather-dependent time series generated by atlite.[31]

While these validators cover the most basic functioning of the grid, further conditions can also be considered. A natural extension would be to ensure N-1 security,[47] meaning that any component may fail and all other components stay within their operational bounds.

III. NODAL FLUCTUATIONS

Due to the increasing share of variable RES, i.e., wind and solar energy, power grids are exposed to new sources of fluctuations. RES are fluctuating at different time scales[48,49] and, particularly, have intermittent fluctuations at short time scales.[50] Along with supply-side fluctuations, recent studies of high-resolution recorded electricity consumption demonstrate intermittent fluctuations on the demand-side[32,51,52] as well. To generate synthetic power grids that imitate the dynamics of real power systems at such short time scales, fluctuations have to be considered both on the supply and demand side.

Here, we introduce the stochastic processes that generate fluctuating wind and solar power, as well as demand time series. These models have been derived to ensure that these synthetic time series have the same short time-scale stochastic characteristics as empirically observed in real data. Therefore, one can confidently use the synthetic time series for further research in power grids and consider the response of power systems to these fluctuations. The effects on the grid frequency are illustrated in Sec. IV.

In addition to the supply and demand fluctuations, the fluctuating time series of the measured grid frequency can also be analyzed directly. The authors of Ref. 53 have introduced a data-driven model to generate realistic synthetic frequency trajectories that reproduce critical statistical properties such as heavy tails in the probability distributions. The authors have shown that market activities induce frequency fluctuations. Specifically, the probability to observe large deviations from the nominal frequency is increased by market activities. Future work should, therefore, take markets into account as drivers of fluctuations.

A. Supply fluctuations

The intermittent nature of wind speed and solar irradiance, along with their turbulent-like behavior, which transfers to wind and solar power and, consequently, to power grids has been widely discussed.[48,50,54,55] As demonstrated in these studies, wind and solar power are non-Gaussian time series and have heavy-tailed probability distribution functions (PDFs). Extreme fluctuations, such as a 90% reduction in power in just a few seconds, occur often in RES. These fluctuations can present additional challenges for maintaining the stability of power systems.

Here, we employ a non-Markovian Langevin-type stochastic process,[56] as well as a jump-diffusion model[57] to generate respectively wind and solar power with similar short time-scale characteristics as the empirical data sets. The Langevin-type model used here is

$$\dot{P}_{wind}(t) = P_{wind}(t)\left(\Gamma - \frac{P_{wind}(t)}{P_0}\right) + \sqrt{\kappa P_{wind}^2(t)}\,n(t) \tag{24}$$

where Γ and P_0 are constant parameters and κ is a parameter with which one can tune the intensity of the noise n. The exact values of the parameters used in our simulations are given in Sec. IV. The noise n is obtained from the following Langevin equation:

$$\dot{n}(t) = -\gamma n(t) + \zeta(t), \tag{25}$$

where ζ is a Gaussian noise with $\langle \zeta(t) \rangle = 0$ and $\langle \zeta(t)\zeta(t') \rangle = \delta(t - t')$. The jump-diffusion model emulating short time-scale fluctuations in solar power is

$$dP_{solar}(t) = D^{(1)}(P_{solar}, t)dt + \sqrt{D^{(2)}(P_{solar}, t)}dw(t) + \eta dJ(t), \tag{26}$$

where $D^{(1)}$ and $D^{(2)}$ are, respectively, the drift and diffusion coefficients. In Eq. (26), dw is the Wiener process and dJ is the Poisson process with jump size η, which is assumed to be a normally distributed random number, i.e., $\eta \sim N(0, \sigma_\eta)$. The Poisson process comprises also a jump rate, which we call λ. The advantage of the jump-diffusion model is that it is a non-parametric model. All model parameters, the drift and diffusion coefficients, and the jump rate, can be calculated from empirical data sets. For the estimation of the parameters, we refer the interested reader to Ref. 57.

B. Demand fluctuations

Standard load profiles used to balance energy in the grid in advance have a time resolution of 15 min. Shorter time scales are balanced by control mechanisms rather than by trading. To study the dynamics at short time scales, the load profiles are, thus, of limited use. Instead, we consider empirical measurements of loads that have a high enough resolution to reveal short-term fluctuations such as in Refs. 32, 51, and 58.

Here, we apply the superstatistics model introduced in Ref. 32 to generate the short time-scale fluctuations of the demand side. Following the superstatistical approach, the demand fluctuations are

obtained by taking the L2 norm of several Gaussian distributions plus a constant offset μ_{MB},

$$P^{fluc}(t) = \sqrt{(z_1(t))^2 + (z_2(t)^2) + \cdots + (z_J(t))^2} + \mu_{MB}, \quad (27)$$

where we use $J = 3$ as discussed in Ref. 32, and $z_i(t)$ is obtained from the following Langevin equation:

$$dz_i(t) = \gamma z_i(t)dt + \epsilon dw_i, \quad (28)$$

where dw_i is the Wiener process with a mean 0 and standard deviation $\sigma = \epsilon/\sqrt{2\gamma}$. We employ the same parameter values μ_{MB}, γ, and ϵ as reported in Ref. 32.

It should be noted that the stochastic time series we have introduced here is based on empirical measurements of power grid actors that are typically not directly connected to the highest level of the power grid. As not all producers and consumers connected to a bus are perfectly correlated, the fluctuations would be attenuated in reality. Unfortunately, few or no measurements of the actual correlations of fluctuations exist, which is why we need to leave this point to future work.

IV. SIMULATION EXAMPLES

In this section, we generate a fully electrified synthetic power grid, whose structure is shown in Fig. 3, and study its behavior in response to the three different fluctuation processes that have been introduced in Sec. III. The synthetic grid that we consider here consists of 100 nodes with an equal share of grid-following and grid-forming inverters. We expect that future power grids will have a high share of variable renewable energies and, therefore, we consider multi-node fluctuations in this example. We assume that the grid-forming inverters are equipped with sufficiently large storage units. Hence, the RES fluctuations are only fed into the grid via the grid-following inverters.

The fluctuations $P_{fluc,i}(t)$ are added to the set points $P_{set,i}$ of the nodes. This results in the following equation for the active power P_i at node i:

$$P_i(t) = P_{set,i} + P_{fluc,i}(t). \quad (29)$$

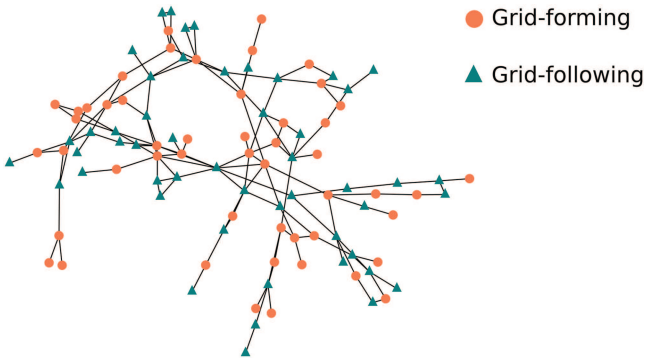

FIG. 3. Network structure of a synthetic power grid. Triangular and circular nodes depict grid-following and grid-forming inverters, respectively.

TABLE III. Performance measures for completely correlated fluctuations.

	$\|\|\mathcal{L}\|\|_{sync}$	$\|\|\mathcal{L}\|\|_{dev}$	$\Delta f_{\max}$(Hz)	$RoCoF_{\max}$(Hz/s)
Wind	0.027	1.237	0.787	0.734
Demand	0.081	2.044	1.446	7.545
Solar	1.619	0.698	0.454	15.295

For the different processes, we will analyze the two edge cases, completely correlated fluctuations, meaning that all nodes have the same fluctuating time series $P_{fluc}(t)$, and second, completely uncorrelated fluctuations where all nodes have different fluctuating time series.

In order to compare the results, we will study four performance measures, the synchronization norm $||\mathcal{L}||_{sync}$,[59] the L_2 norm of the average deviation from the nominal grid frequency $||\mathcal{L}||_{dev}$,[60] the maximal absolute frequency deviation $\Delta f_{\max}$, and the maximal absolute Rate of Change of Frequency $RoCoF_{\max}$,

$$||\mathcal{L}||_{sync} = \sqrt{\frac{1}{T}\int_0^T \frac{1}{N}\sum_{m=1}^{N}\left(\omega_m(t) - \frac{1}{N}\sum_{k=1}^{N}\omega_k(t)\right)^2 dt}, \quad (30)$$

$$||\mathcal{L}||_{dev} = \sqrt{\frac{1}{T}\int_0^T \frac{1}{N}\sum_{m=1}^{N}(\omega_m(t) - \omega_0)^2\, dt}, \quad (31)$$

$$\Delta f_{\max} = \frac{1}{2\pi}\max_t(|\omega(t) - \omega_0|), \quad (32)$$

$$RoCoF_{\max} = \frac{1}{2\pi}\max_t|\dot{\omega}(t)|, \quad (33)$$

where ω_0 is the nominal angular grid frequency. The indices m, k run over all N grid-forming inverters as the grid-following inverters have no internal frequency dynamics (7).

The synchronization norm (30) measures the synchronicity in the power grid. A large synchronization norm expresses a lack of synchronization. The synchronization norm, however, neglects any fluctuation of the so-called bulk,[61] the joint response of the entire power grid, of synchronous frequencies. Therefore, the authors of Ref. 60 introduce the deviation norm $||\mathcal{L}||_{dev}$, which measures the contribution of the bulk to the fluctuations. In Ref. 60, it has been shown that the bulk is the dominant contributor in response to single-node renewable energy fluctuations.

The results are summarized in Tables III and IV. In the cases of wind and demand fluctuations, it can be seen that the deviation norm $||\mathcal{L}||_{dev}$ is larger than the synchronization norm. This indicates that the bulk fluctuations are the main contributors for both

TABLE IV. Performance measures for completely uncorrelated fluctuations.

	$\|\|\mathcal{L}\|\|_{sync}$	$\|\|\mathcal{L}\|\|_{dev}$	$\Delta f_{\max}$(Hz)	$RoCoF_{\max}$(Hz/s)
Wind	0.027	0.181	0.118	0.096
Demand	0.074	0.302	0.176	0.274
Solar	1.344	0.101	0.218	9.713

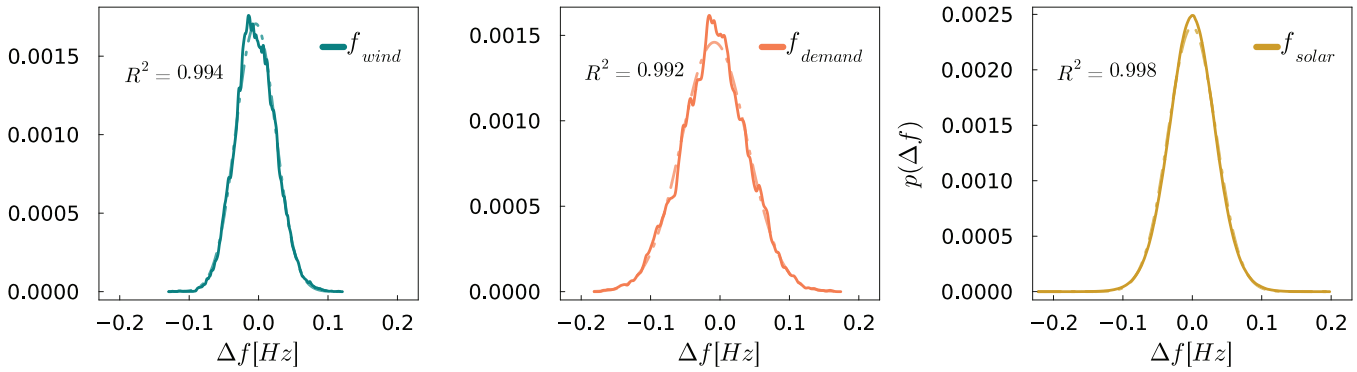

FIG. 4. Probability density function of the frequency response to the uncorrelated fluctuation processes. Bold lines represent the PDFs determined via kernel density estimation. The dashed lines show the best-fitted normal distribution. Each subfigure includes the coefficient of determination R^2 of the fitted normal distribution.

edge cases, the correlated or uncorrelated fluctuations. In the case of solar power, we can see that the synchronization norm is larger than the deviation norm, which expresses a lower degree of synchronization potentially induced by the jump-like nature of the fluctuations. For all processes, we can see that the deviation norm is smaller for the uncorrelated case than for the correlated case, which is to be expected.

Furthermore, it can be seen that the maximal absolute frequency deviation $\Delta f_{\max}$ and the maximal absolute Rate of Change of Frequency $RoCoF_{\max}$ are always smaller in the uncorrelated cases than in the correlated cases as expected. As mentioned in Sec. III B, the actual fluctuations should be attenuated as typically multiple producers and consumers are connected to a single node whose fluctuations are not perfectly correlated. Thus, the result that we present here should be considered a worst-case estimate. This explains why the frequency response for the uncorrelated fluctuations is relatively severe and even surpasses 0.1 Hz.

For all fluctuation processes considered in this work, we find that the voltage magnitudes of the nodes stay close to the set point of 1 p.u., which is to be expected as we simulate active power fluctuations which couple to the frequency.[34]

Figure 4 shows the probability density function for the frequency deviations in the uncorrelated cases determined via kernel density estimation and the corresponding fitted normal distribution. Furthermore, we have calculated the coefficient of determination R^2 for all fits and can see that $R^2 > 0.99$ for all three processes. Although, all fluctuation processes for the demand and the supply are non-Gaussian the response of the frequency is again Gaussian. This is in contrast to the results found for the frequency distribution in Refs. 62 and 53, which show non-Gaussian statistics. However,

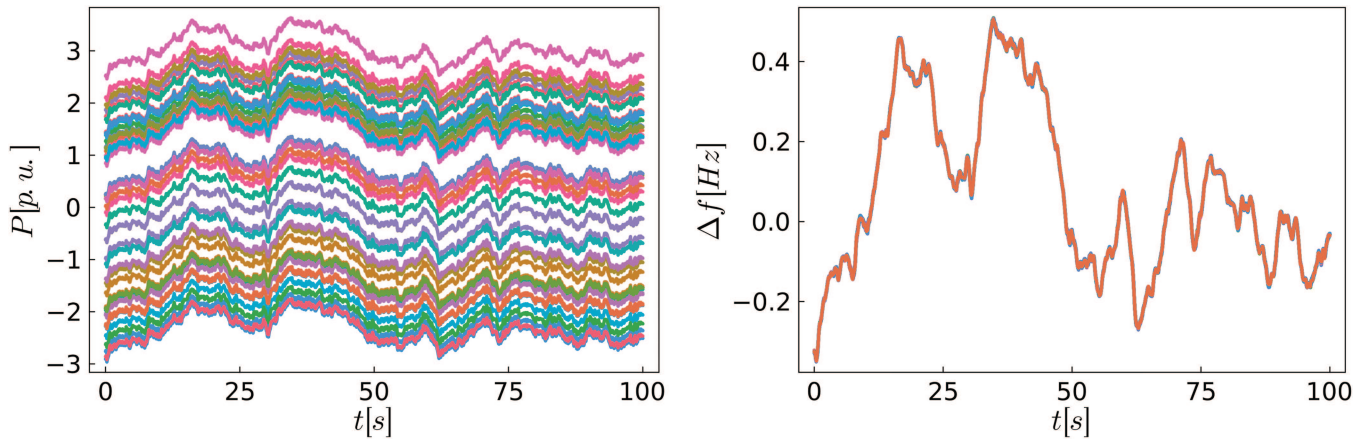

FIG. 5. Results for completely correlated demand fluctuations at the nodes. The figure on the left shows the active powers of the grid-following inverters. The frequency response of the grid-forming inverters is shown in the figure on the right side. The parameters $[\gamma, \epsilon, \mu_{MB}] = [0.016, 33.81, 0.03]$, as in Ref. 32, were used to generate the demand fluctuations.

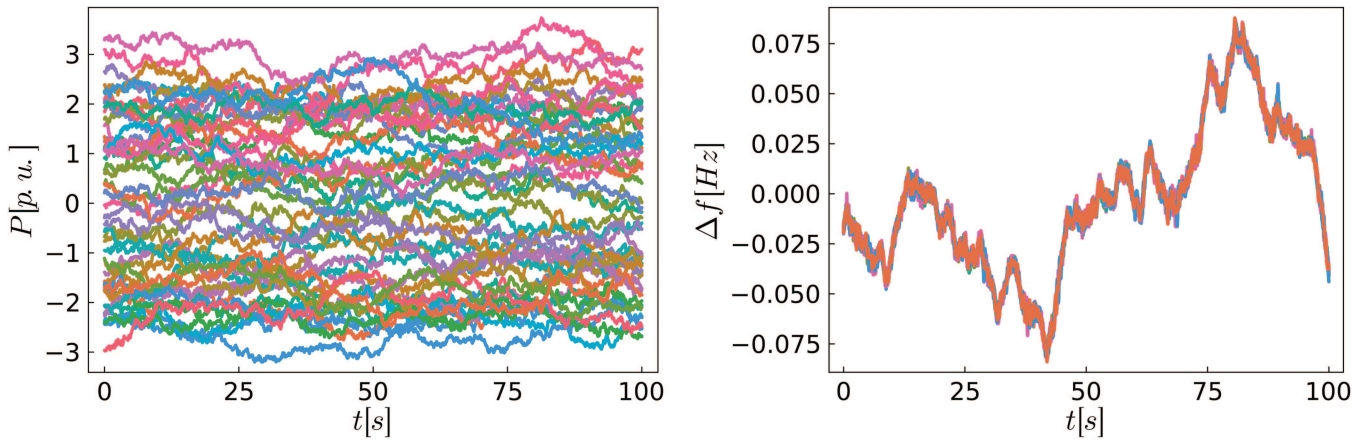

FIG. 6. Results for completely uncorrelated demand fluctuations at the nodes. The figure on the left shows the active powers of the grid-following inverters. The frequency response of the grid-forming inverters is shown in the figure on the right side. The parameters $[\gamma, \epsilon, \mu_{MB}] = [0.016, 33.81, 0.03]$, as in Ref. 32, were used to generate the demand fluctuations.

our model does not include market activities that have been shown to increase the probability of large frequency deviations.[53] Furthermore, we do not expect the synthetic case of our renewable-inverter scenario to reflect the type of frequency dynamics that occur in today's grid where the dynamics are mostly driven by synchronous generation and a lower share of RES. This result further underlines the importance of studying fluctuation processes in power grids and including market activities in future studies.

In the following, we will go into more detail about the results of the demand fluctuations. The results for the wind and solar fluctuations can be found in Appendix B.

Figures 5 and 6 show the results for the correlated and uncorrelated demand fluctuations, respectively. In this example, we use the coefficients for the stochastic process, introduced in Ref. 32, which have been extracted from the NOVAREF data set[63] that consists of high-resolution demand profiles. In a transmission grid, the number of consumers is significantly higher than in the data sets analyzed in Ref. 32.

These results show that our fluctuation model cannot be considered fully realistic on the system level. While the individual time series of demand, solar, and wind are close to those observed in reality, the model does not include realistic correlations between separate injections. We assume perfect correlation at the nodal level, leading to unrealistically large jumps in energy at individual nodes. In addition, full or no correlation across nodes is assumed where the latter leads to incorrect Gaussian behavior for the whole system statistics.[64] The question of the correlation of short-time scale fluctuations at the length scales relevant to the grid is still open. To our knowledge, no direct measurements exist, and thus data-driven models are not available.

Yet, these extreme examples demonstrate that we are able to generate robust and stable synthetic grids. Having highly robust synthetic grids also opens the door to future research that studies grids that are under severe stress, possibly from compound events, meaning that multiple stressors occur at once. Extreme scenarios that can destabilize grids include the loss of multiple lines, as grids are built N-1 secure, special weather conditions that can cause storage to be locally depleted, causing grid-forming inverters to have to compromise on their grid-forming capabilities and to inject fluctuations as well.

V. CONCLUSIONS

In this work, a framework to generate synthetic power grid models for studying collective dynamical effects has been introduced. For the first time, the following established methods are combined to obtain synthetic power grids: realistic grid topologies,[26] active power set points[29,30] and short-term fluctuations, node[22] and line models. Finally, we introduce validators that ensure our power grid and its operation point fulfill established stability criteria[45,46] and reject the sample otherwise. Each element in the framework can be substituted as long as it adheres to the general structure, thus making the approach modular. For the default elements, we have chosen methods that have already been used and validated in various research projects. We have reviewed these established approaches and draw attention to possible improvements in Secs. II A–II E, in particular, in order to investigate electricity grids with a high share of renewable energy. We have identified two elements that need improvement, the generation of network topologies and the distribution of active power supply.

The topologies created with the random growth model[26] cannot reflect the distribution of transmission line lengths in the empirical *SciGRID* data set.[39] The model has been designed to resemble network properties, such as the degree distribution, of real EHV power grids. However, the positioning of the nodes is uniformly random, which does not reflect the growth of real power grids. Grid growth is driven by population and demand growth processes that are far from uniform. We assume that it is possible to correct the length

distribution by introducing an additional step in the algorithm that considers the geographical location of the nodes. Furthermore, we have assumed that the transmission system topology will remain very similar to today. Future studies should consider how the energy transition influences the topology, as, for example, RES are connected to the grid differently than large power plants and the grid evolves to adapt to the new locations.

The major issue in the distribution of active power supply for our synthetic model is that the *ELMOD-DE*[29] specifies scenarios that reflect the current power supply. As we are interested in studying future dynamics as well, a new method for generating active power distributions is needed. Atlite[31] is a software tool that generates weather-dependent power generation potentials and time series for renewable energy technologies. These potentials and time series are promising and could be used to update the active power supply in our model. Further, as the time series depend on the weather, they could also be used to study the synthetic grid under multiple supply scenarios.

Besides the generation of the synthetic grid dynamics in stable operation points, we also include the major drivers of fluctuations at short time scales. We have implemented the three major drivers of short-term fluctuations in future power grids, solar, wind, and demand. As an example, we study a fully synthetic power grid under these fluctuations. We have decided to add the fluctuations only to the components without grid-forming capabilities as grid-forming components will usually be equipped with sufficient storage. We find that the synthetic grid shows good synchronicity under all three fluctuation scenarios. We saw that there is a relevant contribution to the joint response of synchronous frequencies.

It remains a challenge to find a balance between the simplicity and tractability of the model and realism. We have outlined a wide range of points at which realism can be increased. In the current state, the complete model is already well suited to be used in further research projects. This includes developing methods to study compound and extreme events that particularly stress the system. More immediately, it will allow us to advance the study of dynamic power grid stability using graph neural networks.[10,19,20] It enables for the first time to generate a large and robust set of heterogeneous DAE models that will challenge the GNN models and allow us to take one step closer to predicting the dynamic stability of real power grids.

ACKNOWLEDGMENTS

A. Büttner acknowledges support from the German Academic Scholarship Foundation. This research is partially funded within the framework of the "Technology-oriented systems analysis" funding area of the BMWi's 7th Energy Research Program "Innovation for the energy transition" (FKZ: 03EI1016B) and the DFG-project (ExSyCo-Grid, 410409736), as well as the DFG-project (CoComusy, Grant No. KU 837/39-1/RA 516/13-1).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Anna Büttner: Conceptualization (lead); Investigation (lead); Software (lead); Writing – original draft (lead); Writing – review & editing (equal). **Anton Plietzsch:** Conceptualization (supporting); Investigation (supporting); Software (supporting); Writing – original draft (supporting); Writing – review & editing (equal). **Mehrnaz Anvari:** Conceptualization (supporting); Investigation (supporting); Writing – original draft (supporting); Writing – review & editing (equal). **Frank Hellmann:** Conceptualization (supporting); Investigation (supporting); Software (supporting); Writing – original draft (supporting); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in GitHub repository at https://github.com/PIK-ICoNe/SyntheticPowerGrid_Paper_Companion, Ref. 65. The implementation of the synthetic power grid framework can be found in a separate GitHub repository at https://github.com/PIK-ICoNe/SyntheticPowerGrids.jl, Ref. 66.

APPENDIX A: LINE LENGTH DISTRIBUTION

See Fig. 7 for histograms of the line lengths in the SciGRID data set[39] and of our synthetic model.

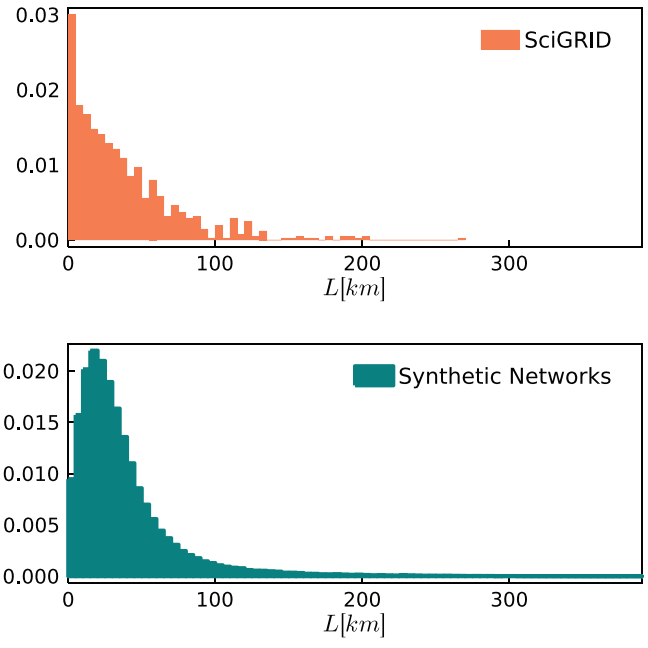

FIG. 7. Histograms of the line lengths in the *SciGRID* data set[39] and of our synthetic model. Both distributions show heavy tails. The data for the *SciGRID* lines indicate a scale-free distribution but the quantity of data is too small to make accurate statements. Further investigations are necessary.

APPENDIX B: RES FLUCTUATION EXAMPLES

The results for the wind (Figs. 8 and 9) and solar fluctuations (Figs. 10 and 11).

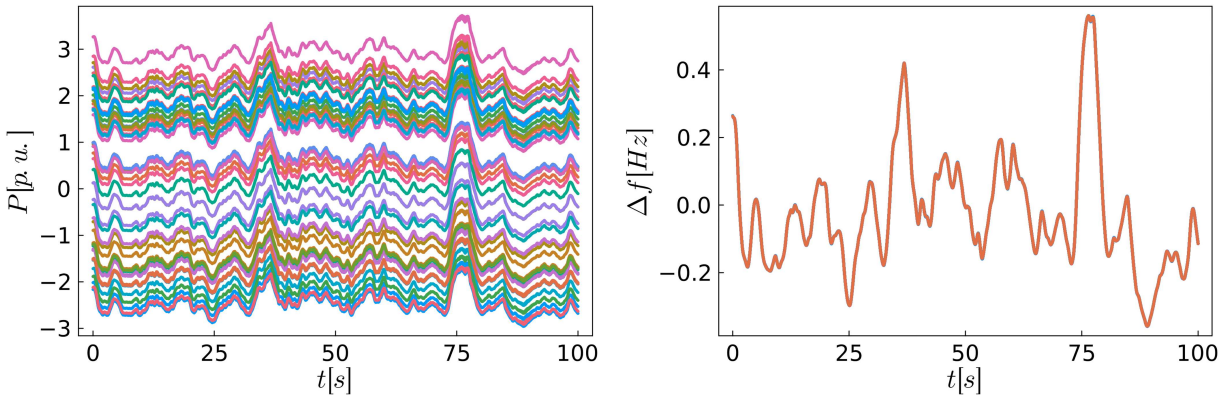

FIG. 8. Results for completely correlated wind power fluctuations. The parameters $[D, \gamma, g, \epsilon] = [0.1, 1.0, 0.5, 1.0]$, as in Ref. 56, were used to generate the wind power fluctuations.

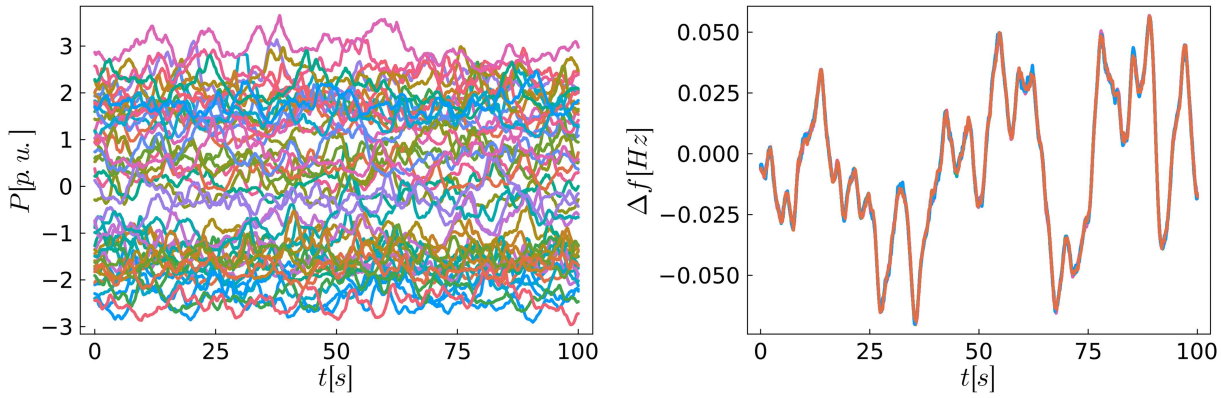

FIG. 9. Results for completely uncorrelated wind power fluctuations. The parameters $[D, \gamma, g, \epsilon] = [0.1, 1.0, 0.5, 1.0]$, as in Ref. 56, were used to generate the wind power fluctuations.

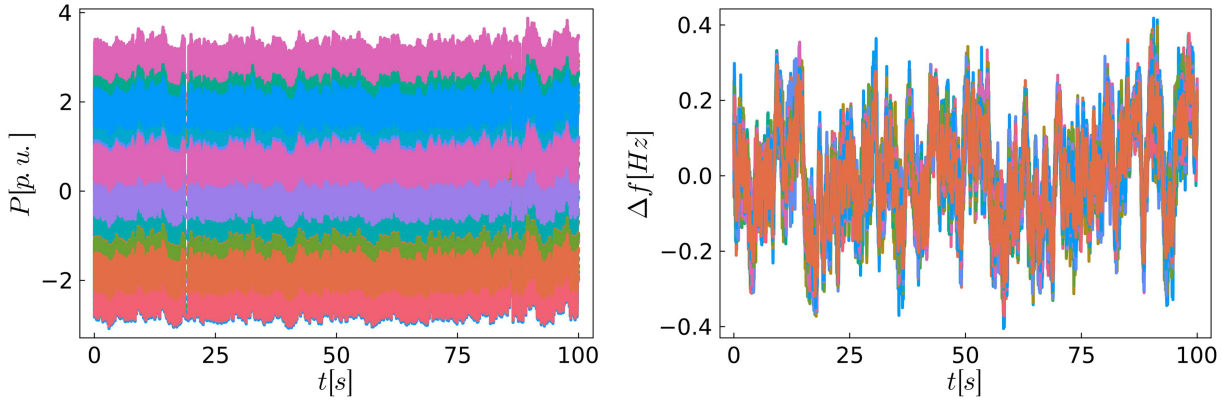

FIG. 10. Results for completely correlated solar power fluctuations. The parameters $[D^{(2)}, \lambda, \sigma_\eta] = [0.001, 0.01, 0.02]$, as in Ref. 57, were used to generate the solar power fluctuations.

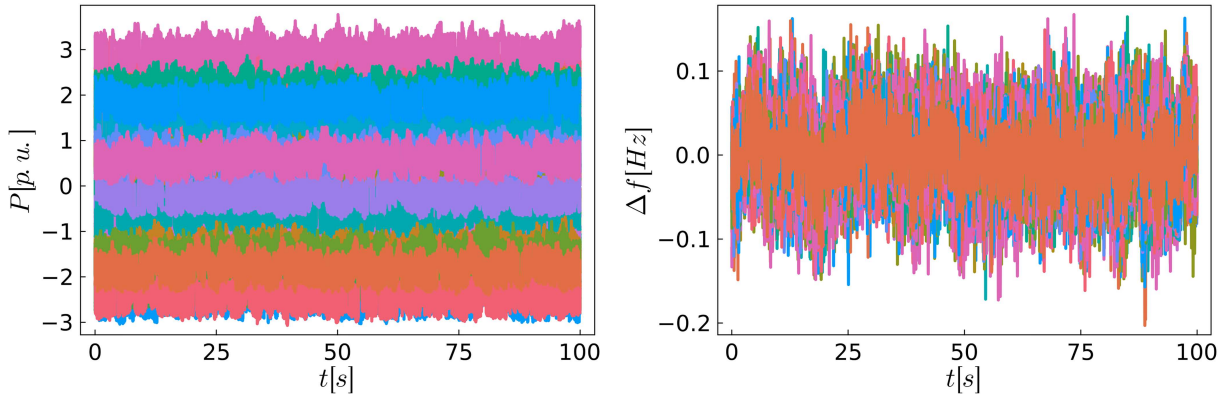

FIG. 11. Results for completely uncorrelated solar power fluctuations. The parameters $[D^{(2)}, \lambda, \sigma_\eta] = [0.001, 0.01, 0.02]$, as in Ref. 57, were used to generate the solar power fluctuations.

REFERENCES

[1] T. Athay, R. Podmore, and S. Virmani, IEEE Trans. Power Appar. Syst. **169**, 573–584 (1979).
[2] C. Grigg, P. Wong, P. Albrecht, R. Allan, M. Bhavaraju, R. Billinton, Q. Chen, C. Fong, S. Haddad, S. Kuruganty *et al.*, IEEE Trans. Power Syst. **14**, 1010–1020 (1999).
[3] A. B. Birchfield, K. M. Gegner, T. Xu, K. S. Shetye, and T. J. Overbye, IEEE Trans. Power Syst. **32**(2), 1502–1510 (2017).
[4] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, IEEE Trans. Power Syst. **32**(4), 3258–3265 (2017).
[5] T. Xu, A. B. Birchfield, K. S. Shetye, and T. J. Overbye, "Creation of synthetic electric grid models for transient stability studies," in *The 10th Bulk Power Systems Dynamics and Control Symposium (IREP 2017)* (International Institute of Research and Education in Power System Dynamics, 2017).
[6] B. Höflich, P. Richard, J. Völker, C. Rehtanz, M. Greve, B. Gwisdorf, J. Kays, T. Noll, J. Schwippe, A. Seack *et al.*, "dena-Verteilnetzstudie Ausbau-Und Innovationsbedarf der Stromverteilnetze in Deutschland bis 2030" (Deutsche Energie-Agentur, Berlin, 2012).
[7] J. Amme, G. Pleßmann, J. Bühler, L. Hülk, E. Kötter, and P. Schwaegerl, J. Phys.: Conf. Ser. **977**, 012007 (2018).
[8] L. Hülk, L. Wienholt, I. Cußmann, U. P. Müller, C. Matke, and E. Kötter, Int. J. Sustainable Energy Plann. Manage. **13**, 79 (2017).
[9] Y. Che and C. Cheng, Chaos **31**, 053129 (2021).
[10] C. Nauck, M. Lindner, K. Schürholt, H. Zhang, P. Schultz, J. Kurths, I. Isenhardt, and F. Hellmann, New J. Phys. **24**, 043041 (2022).
[11] P. J. Menck, J. Heitzig, J. Kurths, and H. J. Schellnhuber, Nat. Commun. **5**, 3969 (2014).
[12] P. Schultz, J. Heitzig, and J. Kurths, New J. Phys. **16**, 125001 (2014).
[13] H. Kim, S. H. Lee, and P. Holme, New J. Phys. **17**, 113005 (2015).
[14] H. Kim, S. H. Lee, and P. Holme, Phys. Rev. E **93**, 062318 (2016).
[15] F. Hellmann, P. Schultz, C. Grabow, J. Heitzig, and J. Kurths, Sci. Rep. **6**, 29654 (2016).
[16] J. Nitzbon, P. Schultz, J. Heitzig, J. Kurths, and F. Hellmann, New J. Phys. **19**, 033029 (2017).
[17] H. Kim, M. J. Lee, S. H. Lee, and S. W. Son, Chaos **29**, 103132 (2019).
[18] M. F. Wolff, P. G. Lind, and P. Maass, Chaos **28**, 103120 (2018).
[19] C. Nauck, M. Lindner, K. Schürholt, and F. Hellmann, "Towards dynamic stability assessment of power grid topologies using graph neural networks," arXiv:2206.06369 [physics] (2023), see http://arxiv.org/abs/2206.06369.
[20] C. Nauck, M. Lindner, K. Schürholt, and F. Hellmann, "Towards dynamic stability analysis of sustainable power grids using graph neural networks," arXiv:2212.11130 [cs,eess] (2022), see https://www.climatechange.ai/papers/neurips2022/16.
[21] P. Christensen, G. K. Andersen, M. Seidel, S. Bolik, S. Engelken, T. Knueppel, A. Krontiris, K. Wuerflinger, T. Bülo, J. Jahn *et al.*, "High penetration of power electronic interfaced power sources and the potential contribution of grid forming converters," Technical Report (European Network of Transmission System Operators for Electricity, 2020).
[22] R. Kogler, A. Plietzsch, P. Schultz, and F. Hellmann, *PRX Energy* (American Physical Society, 2022).
[23] D. Witthaut, F. Hellmann, J. Kurths, S. Kettemann, H. Meyer-Ortmanns, and M. Timme, Rev. Mod. Phys. **94**, 015005 (2022).
[24] A. Plietzsch, R. Kogler, S. Auer, J. Merino, G. de Muro A, J. Liße, C. Vogel, and F. Hellmann, SoftwareX **17**, 100861 (2022), ISSN 2352-7110, see https://www.sciencedirect.com/science/article/pii/S2352711021001345.
[25] P. Tahchiev, F. Leme, V. Massol, and G. Gregory, *JUnit in Action* (Manning Publications, 2011).
[26] P. Schultz, J. Heitzig, and J. Kurths, Eur. Phys. J.: Spec. Top. **223**, 2593 (2014).
[27] ENTSO-E, see https://consultations.entsoe.eu/system-development/tyndp2020/consult_view/ for "Ten-Year Network Development Plan 2020—Main Report" (2020).
[28] T. Brown, D. Schlachtberger, A. Kies, S. Schramm, and M. Greiner, Energy **160**, 720–739 (2018), ISSN 0360-5442, see https://www.sciencedirect.com/science/article/pii/S036054421831288X.
[29] J. Egerer, "Open source electricity model for Germany (ELMOD-DE)," Technical report DIW Data Documentation, 2016, see http://hdl.handle.net/10419/129782.
[30] H. Taher, S. Olmi, and E. Schöll, Phys. Rev. E **100**, 062306 (2019).
[31] F. Hofmann, J. Hampp, F. Neumann, T. Brown, and J. Hörsch, J. Open Source Software **6**, 3294 (2021).
[32] M. Anvari, E. Proedrou, B. Schäfer, C. Beck, H. Kantz, and M. Timme, Nat. Commun. **13**, 1–12 (2022).
[33] A. Singh, P. Eser, N. Chokani, and R. Abhari, Energies **8**, 14168–14181 (2015).
[34] J. Machowski and J. W. Bialek, *Power System Dynamics—Stability and Control* (Wiley & Sons, 2008).
[35] J. Schiffer, R. Ortega, A. Astolfi, J. Raisch, and T. Sezi, Automatica **50**(10), 2457–2469 (2014).
[36] K. Schmietendorf, J. Peinke, R. Friedrich, and O. Kamps, Eur. Phys. J. Spec. Top. **223**, 2577–2592 (2014).
[37] G. Andersson, "Power system analysis," Lecture 227-0526-00 Script, ETH Zürich, 2012.
[38] See https://www.dena.de/newsroom/publikationsdetailansicht/pub/dena-verteilnetzstudie-ausbau-und-innovationsbedarf-der-stromverteilnetze-in-deutschland-bis-2030/ for "dena 2012 dena-Verteilnetzstudie: Ausbau- und Innovationsbedarf der Stromverteilnetze in Deutschland bis 2030."
[39] W. Medjroubi, U. P. Müller, M. Scharf, C. Matke, and D. Kleinhans, Energy Rep. **3**, 14 (2017).

[40]P. Schultz, F. Hellmann, J. Heitzig, and J. Kurths, "A network of networks approach to interconnected power grids," arXiv:1701.06968 (2017).
[41]W. Zhang, F. Li, and L. M. Tolbert, IEEE Trans. Power Syst. **22**, 2177–2186 (2007).
[42]C. Coffrin, R. Bent, K. Sundar, Y. Ng, and M. Lubin, "Powermodels.jl: An open-source framework for exploring power flow formulations," in *2018 Power Systems Computation Conference (PSCC)* (IEEE, 2018), pp. 1–8.
[43]A. B. Birchfield, T. Xu, and T. J. Overbye, IEEE Trans. Power Syst. **33**, 6667–6674 (2018).
[44]European Committee for Electrotechnical Standardization, "Voltage characteristics of electricity supplied by public electricity networks," standard DIN EN 50160:2020-11 (British Standards Institution (BSI), 2020).
[45]T. Gonen, *Electrical Power Transmission System Engineering: Analysis and Design* (CRC Press, 2014), ISBN 9781482232226.
[46]F. Milano, *Power System Modelling and Scripting*, Power Systems (Springer, 2010).
[47]A. Shokri Gazafroudi, F. Neumann, and T. Brown, Int. J. Electr. Power Energy Syst. **137**, 107702 (2022), ISSN 0142-0615, see https://www.sciencedirect.com/science/article/pii/S0142061521009297.
[48]J. Apt, J. Power Sources **169**, 369–374 (2007).
[49]A. Woyte, R. Belmans, and J. Nijs, Sol. Energy **81**, 195–206 (2007).
[50]M. Anvari, G. Lohmann, M. Wächter, P. Milan, E. Lorenz, D. Heinemann, M. R. R. Tabar, and J. Peinke, New J. Phys. **18**, 063027 (2016).
[51]A. Wright and S. Firth, Appl. Energy **84**, 389–403 (2007).
[52]A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro, and A. M. Tonello, "GREEND: An energy consumption dataset of households in Italy and Austria," in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)* (IEEE, 2014), pp. 511–516.
[53]L. R. Gorjão, M. Anvari, H. Kantz, C. Beck, D. Witthaut, M. Timme, and B. Schäfer, IEEE Access **8**, 43082–43097 (2020).
[54]P. Milan, M. Wächter, and J. Peinke, Phys. Rev. Lett. **110**, 138701 (2013).
[55]A. E. Curtright and J. Apt, Prog. Photovoltaics: Res. Appl. **16**, 241–247 (2008).
[56]K. Schmietendorf, J. Peinke, and O. Kamps, Eur. Phys. J. B **90**, 1–6 (2017).
[57]M. Anvari, B. Werther, G. Lohmann, M. Wächter, J. Peinke, and H. P. Beck, Sol. Energy **157**, 735–743 (2017).
[58]A. Marszal-Pomianowska, P. Heiselberg, and O. K. Larsen, Energy **103**, 487–501 (2016), ISSN 0360-5442.
[59]M. Andreasson, E. Tegling, H. Sandberg, and K. H. Johansson, "Coherence in synchronizing power networks with distributed integral control," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)* (IEEE, 2017), pp. 6327–6333.
[60]A. Plietzsch, S. Auer, J. Kurths, and F. Hellmann, Chaos **32**, 113114 (2022), ISSN 1054-1500, see https://pubs.aip.org/aip/cha/article-pdf/doi/10.1063/5.0122898/16499190/113114_1_online.pdf.
[61]X. Zhang, S. Hallerberg, M. Matthiae, D. Witthaut, and M. Timme, Sci. Adv. **5**, eaav1027 (2019), see https://www.science.org/doi/abs/10.1126/sciadv.aav1027.
[62]B. Schäfer, C. Beck, K. Aihara, D. Witthaut, and M. Timme, Nat. Energy **3**(2), 119–126 (2018).
[63]M. Lange and M. Zobel, "NOVAREF, Erstellung neuer Referenzlastprofile zur Auslegung, Dimensionierung und Wirtschaftlichkeitsberechnung von Hausenergieversorgungssystemen," Technical Report (NEXT ENERGY, 2016).
[64]H. Haehne, K. Schmietendorf, S. Tamrakar, J. Peinke, and S. Kettemann, Phys. Rev. E **99**, 050301 (2019).
[65]A. Büttner and A. Plietzsch (2023). "Synthetic power grid paper companion," GitHub. https://github.com/PIK-ICoNe/SyntheticPowerGrid_Paper_Companion
[66]A. Büttner, A. Plietzsch, M. Bornemann, and F. Hellmann, "Synthetic power grids package," GitHub. https://github.com/PIK-ICoNe/SyntheticPowerGrids.jl