nature energy

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41560-017-0058-z

In the format provided by the authors and unedited.

Non-Gaussian power grid frequency fluctuations characterized by Lévy-stable laws and superstatistics

Benjamin Schäfer[1,2]*, Christian Beck[3], Kazuyuki Aihara[4], Dirk Witthaut[5,6] and Marc Timme[1,2]

[1]Chair for Network Dynamics, cfaed and Institute for Theoretical Physics, Technical University of Dresden, Dresden, Germany. [2]Network Dynamics, MPIDS, Göttingen, Germany. [3]School of Mathematical Sciences, Queen Mary University of London, London, UK. [4]Institute of Industrial Science, The University of Tokyo, Meguro-ku, Tokyo, Japan. [5]Forschungszentrum Jülich, Institute of Energy and Climate Research-Systems Analysis and Technology Evaluation, Jülich, Germany. [6]Institute for Theoretical Physics, University of Cologne, Köln, Germany. Dirk Witthaut and Marc Timme contributed equally to this work.
*e-mail: benjamin.schaefer@ds.mpg.de

SUPPLEMENTARY NOTE 1

Power grid frequency data presentation

We have a closer look at the data sets of power grid frequency measurements. Our available data includes ENTSO-E Continental European (CE) [1, 2], Great Britain (GB) [3], Mallorcan [4] and Nordic [5] grids, the 50 Hz and 60 Hz regions of Japan [6] as well as one day of the Eastern Interconnection (EI) [7]. Production data including share of wind and solar for each region was taken from [6, 8–10].

Supplementary Table I. The mean frequency is kept close to the reference frequency in all grids while standard deviations differ significantly. We list the estimates for mean, standard deviation (SD), skewness and kurtosis of distributions for different European (Continental Europe (CE), Great Britain (GB), Nordic, Mallorca), Japanese and Eastern Interconnection (EI) data sets and years including the number of data points.

Source	year/ region	# of data points	mean μ [Hz]	SD σ [Hz]	skewness β	kurtosis κ
CE (*50Hertz*)[1]	2014	31141070	49.9995	0.0202	0.047	4.04
	2015	24685140	49.9999	0.0200	-0.024	4.10
CE (*RTE*)[2]	2015	3153029	50.0003	0.0202	-0.007	3.89
Nordic (*FinGrid*) [5]	2015	31505090	50.000	0.0434	0.033	3.11
	2016	31424665	50.000	0.0456	0.046	3.10
GB (*Nationalgrid*)[3]	2014	31536000	49.9997	0.0545	0.232	3.02
	2015	31536000	49.9997	0.0544	0.258	2.91
Japan (*OCCTO*) 2016 [6]	50 Hz	91845	50.0003	0.0304	0.018	3.17
	60 Hz	91845	60.0025	0.0376	0.000	4.01
Mallorca [4]	2015	2483410	49.9999	0.0415	-0.014	4.99
EI (1d)[7]	(2014)	863995	59.9967	0.0175	0.316	2.97

Distribution measures and histograms

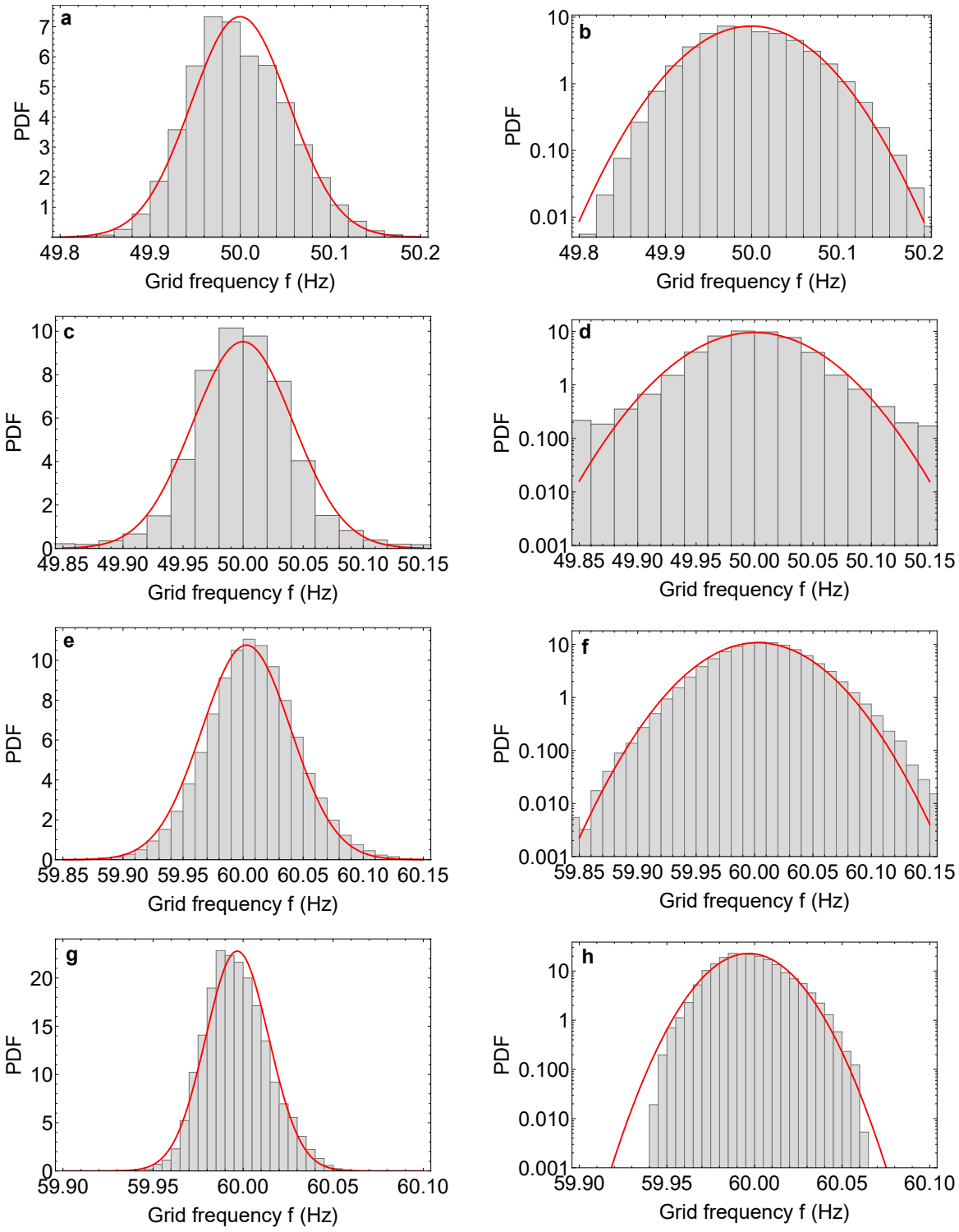
As an introduction to the data, we list mean μ, standard deviation σ, skewness β ($\beta^{\text{Gaussian}} = 0$) and kurtosis κ ($\kappa^{\text{Gaussian}} = 3$), i.e., the (normalized) first four moment of the distributions in Table I. Analyzing the figures reveals that all distributions are close to their nominal frequency of 50/60 Hz. Furthermore, all grids have either a higher kurtosis than expected from a normal distribution (Continental Europe, Nordic, Mallorca, Japan) or are skewed (Great Britain, Eastern Interconnection). Next, we visualize several data sets as histograms compared to their best-fitting normal distribution in Fig. 1.

For the Continental European (CE) grid we have two data bases, one by the German transmission system operator (TSO) *50Hertz* and the other by the French TSO *RTE*. Although the measurements were taken at different locations of the connected grid, they return close to identical statistics.

Pre-processing of data sets

We had to perform some pre-processing especially with the *50Hertz* data set and the data from Mallorca: The original data set [1] contains entries set to 0, 52 or 48 Hz, while the Mallorcan data included a few very large and small values. In the case of *50Hertz*, they confirmed measurements problems leading to these small and large values. Hence, we deleted entries larger than 51Hz or smaller than 49 Hz. We associate these large deviations with blackouts or nearly blackouts, which are no longer covered by our theory. However, even after excluding these extreme values, the *50Hertz* data had some very large jumps within 1 second to a larger value by $\Delta f \sim 0.5Hz$ which we found to be most likely also artifacts. Hence, we decided to scan the data for jumps larger than $\Delta f_{max} = 0.1Hz$ and delete these if they were isolated, i.e., values before and after this value are at least different by Δf_{max}. This way the statistical measures, e.g. variance and kurtosis, of the *50Hertz* data set approached the *RTE* data set, as expected for the same synchronous region.

When computing the noise amplitude ϵ for different regions in Figure 5 in the main manuscript, we assumed that the Eastern Interconnection has a total inertia of $M = 1000$ to increase the absolute value of the estimated noise.



Supplementary Figure 1. All data sets show deviations from normal distributions. We plot the histograms for the data of Great Britain (GB), Mallorca, Japan and Eastern Interconnection (EI) together with their estimated normal distribution. **a**: GB with linear scale, **b**: GB with log-scale, **c**: Mallorca region with linear scale, **d**: Mallorca region with log-scale, **e**: Japan 60Hz region with linear scale, **f**: Japan 60Hz region with log-scale, **g**: EI with linear scale, **h**: EI with log-scale. Each histogram is either more heavy-tailed (c,d,e,f) than a Normal distribution or skewed (a,b,g,h).

Estimated distributions

Noticing deviations from normal distributions both in Table I and Fig. 1, we perform a maximum likelihood analysis, see e.g. [11], of the available data and thereby determine which standard distribution fits the data best. This is done with the assumption that one well chosen distribution would be able to fit all data sets (with differing parameters for each grid). Given a probability density function $p(x)$ and a data set $Y = \{y_1, y_2, ..., y_N\}$, we calculate the likelihood that Y is drawn from the distribution p by calculating

$$L_{p,Y} = \Pi_{i=1}^{N} p(y_i). \tag{1}$$

The maximum likelihood estimate is based on comparing at least two different distributions, e.g., $p_1(x)$ and $p_2(x)$ by computing the likelihoods for both distributions. Next, we have a look at the log-likelihood ratio

$$\log_{10}\left(\frac{L_{p_1,Y}}{L_{p_2,Y}}\right), \tag{2}$$

which is the most powerful tool to distinguish two distributions [11]. We then accept p_1 over p_2 if the ratio is larger than 1. We use the Mathematica *EstimatedDistribution* [12] routine testing several build-in distributions against the data: *HyperbolicDistribution[λ,a,b,c,d], StableDistribution[a,b,c,d], NormalDistribution[a,b], LogNormalDistribution[a,b], SkewNormalDistribution[a,b,c], LogLogisticDistribution[a,b], StudentTDistribution[a,b,c], ParetoDistribution[a,b,c,d], SechDistribution[a,b], ExponentialPowerDistribution[a,b,c], JohnsonDistribution["SU", a,b,c,d]* and *TsallisQGaussianDistribution[a,b,c]*. As an example, we note down the estimated stable distributions for some grids in Table II. Note also that especially the (generalized) hyperbolic and stable distributions use many parameters to fit the distributions.

For the heavy-tail distributions (Continental Europe, Nordic, Japan, Mallorca) normal distributions perform worse than stable distributions which get outperformed by q-Gaussian distributions in terms of likelihood (which get outperformed by a small margin by generalized hyperbolic distributions). However, skewed distributions (Great Britain and Eastern Interconnection) are best fitted by skew normal distributions but can be approximated by stable distributions as these also describe skewed distributions.

As an example, for the 50Hertz data of 2015 we use more than 24.6 million frequency measurements to compute the log-likelihood ratios of the Normal, Stable and q-Gaussian distributions:

$$\log_{10}\left(\frac{L_{\text{q-Gaussian, 50Hertz2015}}}{L_{\text{Stable, 50Hertz2015}}}\right) \approx 30360, \tag{3}$$

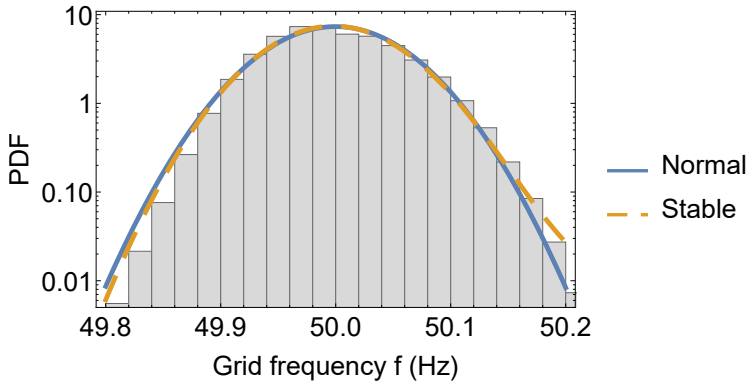
$$\log_{10}\left(\frac{L_{\text{Stable, 50Hertz2015}}}{L_{\text{Normal, 50Hertz2015}}}\right) \approx 94758, \tag{4}$$

$$\log_{10}\left(\frac{L_{\text{q-Gaussian, 50Hertz2015}}}{L_{\text{Normal, 50Hertz2015}}}\right) \approx 125118. \tag{5}$$

Based on these likelihood results, we focus on stable and q-Gaussian distributions in the main manuscript.

Supplementary Table II. Estimated stable distribution parameters: stability parameter α_S, skewness parameter β_S, location μ_S and scale parameter σ_S for different regions and years (1 day in the case of the Eastern Interconnection). All parameters are given with their standard error.

Source	Year/ Region	Stable Distribution			
		Stability para. α_S	Skewness par. β_S	Location par. μ_S	Scale par. σ_S
Continental Europe (*50Hertz*)[1]	2015	1.898±0.002	0.006±0.004	49.9999±0.00004	0.0132±0.00002
Great Britain (*Nationalgrid*)[3]	2015	1.969±0.005	0.997±0.005	50.0001±0.0002	0.0378±0.0004
Japan (*OCCTO*) 2016 [6]	50 Hz	1.988±0.002	0.237±0.120	50.0003±0.0001	0.0213±0.0001
	60 Hz	1.986±0.002	0.387±0.114	60.0025±0.0001	0.0263±0.0001
Eastern Interconnection(1d)[7]	(2014)	1.938±0.002	0.999±0.001	59.9969±0.0001	0.0121±0.00003
Nordic [5]	2015	1.987±0.003	0.999±0.001	49.9997±0.00004	0.0306±0.00002
Mallorca [4]	2015	1.818±0.003	0.074±0.005	50.0000±0.00002	0.0264±0.00001



Supplementary Figure 2. Stable distributions account for skewed distributions. Plotted is the Log plot of the histogram of the Great Britain 2015 data and its best normal as well as stable distribution fit. The skewed stable distribution is a better description both for low and high frequencies than the normal distribution.

SUPPLEMENTARY NOTE 2

Additional Fokker-Planck results

We extend the Fokker-Planck results obtained in the main text by calculating the standard deviation and noise amplitude assuming Gaussian noise and adding treatment for primary control with deadzones. Deadzones arise naturally in power grid control [13] where it is only possible to determine the frequency to a finite precision. Hence, one could argue that the non-Gaussian nature of the observed distribution could be explained by Gaussian noise combined with the nonlinear control.

Solving Fokker-Planck equations

In the main text we made use of a general time-independent solution of a Fokker-Planck equation. Given a general stochastic system whose dynamics is defined by

$$\frac{\mathrm{d}}{\mathrm{d}t}X = a\left(X\right) + \sqrt{b\left(X\right)}\xi, \tag{6}$$

where ξ is white Gaussian noise based on a Wiener process, its stationary Fokker-Planck equation is given by

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[a\left(x\right)p\left(x\right)\right] - \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left[b\left(x\right)p\left(x\right)\right] = 0, \tag{7}$$

which is solved to give [14]

$$p\left(x\right) = \frac{c}{b(x)}\exp\left[2\int_0^x a(s)/b(s)\mathrm{d}s\right], \tag{8}$$

with normalization constant c. In the case of $a\left(x\right) = -a_0 x$ and $b\left(x\right) = b_0$, the final distribution is a Gaussian distribution, as is the case for the power grid. We assumed that the probability and its derivative vanish at infinity to derive (8) [14]:

$$\lim_{|x|\to\infty} p\left(x,t\right) = 0, \tag{9}$$

$$\lim_{|x|\to\infty} \frac{\partial}{\partial x}p\left(x,t\right) = 0. \tag{10}$$

Assuming that the power grid is dominated by Gaussian noise, we formulate the Fokker-Planck equation in the main manuscript for the bulk angular velocity $\bar{\omega}$ as

$$\frac{\partial p}{\partial t} = \gamma\frac{\partial}{\partial\bar{\omega}}\left(\bar{\omega}p\right) + \frac{1}{2}\sum_{i=1}^{N}\frac{\epsilon_i^2}{M^2}\cdot\frac{\partial^2 p}{\partial\bar{\omega}^2}, \tag{11}$$

which is solved by the probability density function

$$p\left(\bar{\omega}\right) = \sqrt{\frac{\gamma M^2}{\pi\sum_{i=1}^{N}\epsilon_i^2}}\exp\left[-\bar{\omega}^2\frac{\gamma M^2}{\sum_{i=1}^{N}\epsilon_i^2}\right], \tag{12}$$

with damping γ, the number of nodes N, noise amplitude ϵ and summed inertia $M = \sum_i M_i$.

Standard deviation predictions

The standard deviation of the bulk angular velocity $\bar{\omega}$, assuming that all nodes have the same noise amplitude $\epsilon_i \approx \epsilon$ and unit inertia $M_i = 1\ \forall i$, is given by

$$\sigma = \sqrt{\frac{\epsilon^2}{2N\gamma}}. \tag{13}$$

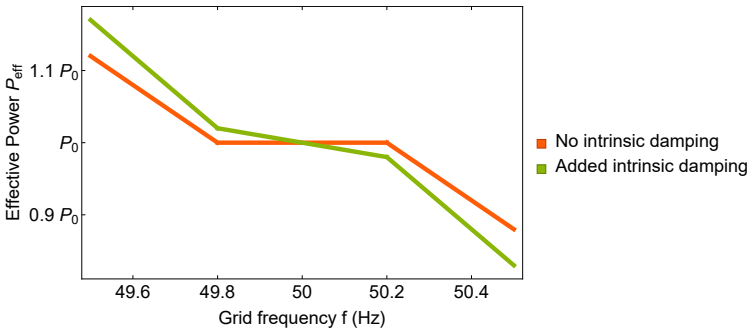
This standard deviation is dependent on the number of nodes N in the grid, i.e., synchronous regions with less production, i.e., with fewer nodes will have a broader distribution and hence higher risk of large fluctuations.

Deadzones

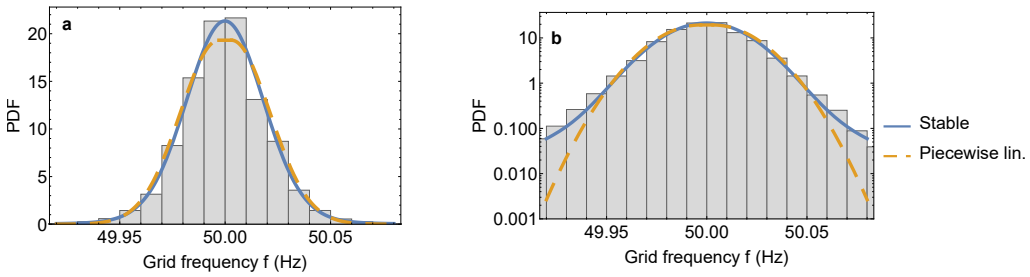
The power grid frequency (and angular velocity) cannot be determined to arbitrary precision, giving rise to *deadzones* of control, i.e., for a small interval $\omega = 0 \pm \Delta\omega_D$ there is no (primary) control activated at the swinging machines. Could those deadzones explain the non-Gaussian distributions in the frequency assuming Gaussian noise but nonlinear control? These deadzones are typically of the order $\Delta\omega_D \approx 2\pi\,(10...200\,mHz)$ [13, 15], see Fig. 3 for an illustration where we split the damping γ into intrinsic damping γ_D that arises from damper windings etc. and the (primary) control damping γ_C which is only active outside of the deadzone. Given our solution of the Fokker-Planck equation, we calculate the probability density function for piecewise linear control to be

$$p(\omega) = \frac{\sqrt{\gamma_C}}{2\Delta\omega_D + \sqrt{\pi}} \begin{cases} \exp\left[-\gamma_C\left(\Delta\omega_D + \omega\right)^2\right] & \omega < -\Delta\omega_D, \\ 1 & else, \\ \exp\left[-\gamma_C\left(\Delta\omega_D - \omega\right)^2\right] & \omega > \Delta\omega_D, \end{cases} \tag{14}$$

where we used $\gamma_D = 0$. Performing a maximum likelihood analysis to estimate the deadzone ω_D, we reach the conclusion that the most likely value for the deadzone parameter is $\omega_D \approx 0$, i.e., we do not need a deadzone to model our real power grid frequency data. Furthermore, stable distributions still outperform Gaussian noise in terms of likelihood even when deadzones are included, see Fig. 4.



Supplementary Figure 3. Deadzones of control lead to piecewise linear power(frequency) functions. We plot the effective power $P_{eff} = P_0 + \text{control}$ as a function of frequency using piecewise linear functions modeling deadzones of primary control $\gamma_C > 0$, with intrinsic damping $\gamma_D > 0$ (green) and without (orange).



Supplementary Figure 4. Stable distributions outperform Gaussian noise with deadzones in terms of likelihood. We compare stable distributions (blue) with a piecewise linear probability density function modeling deadzones of primary control (dashed orange) and the real *50Hertz* 2015 data set (histogram). **a**: We plot the histogram of the real data together with the PDFs of the best-fitting stable distribution and a piecewise linear. **b**: We repeat the plot with a log-scale of the PDF. Parameters are $\Delta\omega_D = 3$ mHz and $N\gamma_C/\epsilon^2 = 1500$. We note that increasing the damping γ_C increases the peak at 50 Hz but also results in flatter tails. Hence, we improve one property at the expense of the other as the piecewise linear distribution tends to have a peak lower than the data (and the stable distribution) and flatter tails.

SUPPLEMENTARY NOTE 3

Additional generalized Fokker-Planck results

The main text focused on results of the (ordinary) Fokker-Planck equation which only covers Gaussian noise. We add to this by presenting the generalized Fokker-Planck equation following [16], treating Lévy-stable and generalized hyperbolic distributions explicitly, providing a time-dependent solution and showing the extension to arbitrary composite distributions.

Generalized Fokker-Planck equation

Let us assume that the evolution of state $x(t)$ is characterized by the stochastic differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = a(x,t)\,x(t) + b(x,t)\,\xi, \tag{15}$$

with intrinsic dynamic $a(x,t)$, noise amplitude b and arbitrarily distributed noise ξ. We assume the noise to be uncorrelated, i.e.

$$\langle \xi(t)\,\xi(s)\rangle = \delta(t-s), \tag{16}$$

with Dirac delta function $\delta(...)$ and to have mean zero:

$$\langle \xi(t)\rangle = 0. \tag{17}$$

Then the generalized Fokker-Planck equation is given by

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}\left(a(x,t)\,p(x,t)\right) + \mathcal{F}^{-1}\left[\varphi_k(t)\ln S_k^{\mathrm{in}}\right], \tag{18}$$

where

$$\mathcal{F}^{-1}[u_k] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{ikx}u_k \mathrm{d}k \tag{19}$$

is the inverse Fourier transform, φ_k is the characteristic function of the probability density function $p(x,t)$ given as $\varphi_k(t) = \left\langle \mathrm{e}^{-ikx(t)}\right\rangle$ and S_k^{in} is the characteristic function of the (input) noise at time $t=1$ $S_k^{\mathrm{in}} = \left\langle \mathrm{e}^{-ik\eta(1)}\right\rangle$ with noise generating process η. A partial differential equation like (18) requires an initial and boundary conditions to be solved. We assume the initial condition to be a delta peak at zero

$$p(x, t=0) = \delta(0). \tag{20}$$

For the boundary conditions we assume the probability density function and its spatial derivative to vanish at infinity:

$$\lim_{|x|\to\infty} p(x,t) = 0, \tag{21}$$

$$\lim_{|x|\to\infty} \frac{\partial}{\partial x}p(x,t) = 0. \tag{22}$$

The intrinsic dynamic $a(x,t)$ is often obtained from a potential $U(x)$ as

$$a(x) = -\frac{\mathrm{d}U}{\mathrm{d}x}(x). \tag{23}$$

For the special case of a quadratic potential $U(x) = \gamma x^2/2$ and additive noise with $b=1$, the stochastic equation of motion simplifies to

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = -\gamma x(t) + \xi. \tag{24}$$

For this dynamic of x, the generalized Fokker-Planck equation (18) is solved by the methods of characteristics [17] given as

$$p\left(x,t\right)=\mathcal{F}^{-1}\left[\exp\left(-\frac{1}{\gamma}\int_{0}^{k}\frac{1}{z}\ln\left(\frac{S_{z\cdot e^{-\gamma t}}^{\text{in}}}{S_{z}^{\text{in}}}\right)\mathrm{d}z\right)\right], \tag{25}$$

where S_k^{in} is the characteristic function of the (arbitrarily distributed but uncorrelated) noise generating process [16].

Using the characteristic function of the (output) distribution S_k^{out} and the characteristic function of the input noise distribution S_k^{in}, we note the following important relations of the stationary distributions:

$$S_{k}^{\text{out}}=\exp\left[\frac{1}{\gamma}\int_{0}^{k}\frac{\ln\left(S_{z}^{\text{in}}\right)}{z}\mathrm{d}z\right], \tag{26}$$

$$S_{k}^{\text{in}}=\exp\left[\gamma\cdot k\cdot\frac{\partial}{\partial k}\ln S_{k}^{\text{out}}\right]. \tag{27}$$

Levy-stable noise

Assuming our system is described by a Langevin equation of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}x\left(t\right)=a\left(x,t\right)x\left(t\right)+b\left(x,t\right)\xi, \tag{28}$$

and is subject to Lévy-stable distributed noise with characteristic function $S_k\left(\alpha_S,\beta_S,\sigma_S\right)$:

$$S_{k}^{\text{in}}\left(\alpha_{S},\beta_{S},\sigma_{S}\right)=\exp\Big[-\sigma_{S}^{\alpha_{S}}\left|k\right|^{\alpha_{S}}\cdot\left(1-\operatorname{sgn}\left(k\right)i\beta_{S}\cdot\tan\frac{\pi\alpha_{S}}{2}\right)\Big], \tag{29}$$

with stability parameter α_S, skewness parameter β_S and scale parameter σ_S (setting the location parameter $\mu_S=0$). Then, the generalized Fokker-Planck equation [16] is given by

$$\begin{aligned}\frac{\partial p}{\partial t}\left(x,t\right)=&-\frac{\partial}{\partial x}\left[a\left(x,t\right)p\left(x,t\right)\right]\\&+\sigma_{S}^{\alpha_{S}}\frac{\partial^{\alpha_{S}}}{\partial\left|x\right|^{\alpha_{S}}}\left[b\left(x,t\right)^{\alpha_{S}}p\left(x,t\right)\right]\\&+\sigma_{S}^{\alpha}\beta_{S}\tan\left(\frac{\pi\alpha_{S}}{2}\right)\frac{\partial^{\alpha_{S}-1}}{\partial\left|x\right|^{\alpha_{S}-1}}\left[b\left(x,t\right)^{\alpha_{S}}p\left(x,t\right)\right].\end{aligned} \tag{30}$$

The fractional derivative of a function $h\left(x\right)$ is defined as

$$\frac{\partial^{\alpha_{S}}}{\partial\left|x\right|^{\alpha_{S}}}h\left(x\right)=-\mathcal{F}^{-1}\left[\left|k\right|^{\alpha}h_{k}\right]. \tag{31}$$

Time-dependent solution

We formulate the time-dependent solution given the quadratic potential $U\left(x\right)=\gamma x^{2}/2$ and assuming Lévy-stable noise input. We compute

$$\frac{S_{z\cdot e^{-\gamma t}}^{\text{in}}}{S_{z}^{\text{in}}}=\exp\left[-\sigma_{S}^{\alpha_{S}}\left|z\right|^{\alpha_{S}}\left(e^{-\gamma\alpha_{S}t}-1\right)\right] \tag{32}$$

and can now compute the final characteristic function

$$S_z^{\text{out}}(t) = \exp\left[-\frac{1}{\gamma}\int_0^k \frac{1}{z}\ln\left[\frac{S_{z\cdot e^{-\gamma t}}^{\text{in}}}{S_z^{\text{in}}}\right]\mathrm{d}z\right] \tag{33}$$

$$= \exp\left[\frac{\sigma_S^{\alpha_S}}{\gamma}\left(e^{-\gamma\alpha_S t}-1\right)\int_0^k \frac{|z|^{\alpha_S}}{z}\mathrm{d}z\right] \tag{34}$$

$$= \exp\left[-\frac{(1-e^{-\gamma\alpha_S t})}{\gamma\alpha_S}\sigma_S^{\alpha_S}|z|^{\alpha_S}\right] \tag{35}$$

$$= S_z^{\text{in}}\left(\alpha_S, 0, \sigma_S\cdot\left(\frac{(1-e^{-\gamma\alpha_S t})}{\gamma\alpha_S}\right)^{1/\alpha_S}\right). \tag{36}$$

So we get a time-dependent scaling parameter

$$\sigma_S(t) = \sigma_S\cdot\left(\frac{(1-e^{-\gamma\alpha_S t})}{\gamma\alpha_S}\right)^{1/\alpha_S}, \tag{37}$$

which is zero at the initial condition, consistent with the ansatz in [16]: $\sigma_S(0) = 0$, i.e., the probability density function is a delta function $p(t=0,\bar{\omega}) = \delta(\bar{\omega})$. For $t\to\infty$, the scaling parameter approaches the value stated in the main manuscript as follows:

$$\sigma_S(t) \to \sigma_S\cdot\left(\frac{1}{\gamma\alpha_S}\right)^{1/\alpha_S}, \text{as } t\to\infty. \tag{38}$$

In this calculation we assumed zero mean ($\mu_S = 0$) and neglected skewness since it does not change.

Generalized hyperbolic distribution

The generalized hyperbolic distribution (describing among others, generalized inverse Gamma and StudentT distributions) returned one of the highest likelihoods for the Continental European data and hence is worth special attention. Assuming that the final distribution is given as a generalized hyperbolic distribution, we have

$$S_k^{\text{out}} = \frac{\left(\sqrt{\alpha_H H^2\delta}\right)^{\lambda}\left(\sqrt{k^2+\alpha_H H^2\delta}\right)^{-\lambda}K_\lambda\left(\sqrt{k^2+\alpha_H H^2\delta}\right)}{K_\lambda\left(\sqrt{\alpha_H H^2\delta}\right)}, \tag{39}$$

where K_λ is the modified Bessel function of the second kind, λ and α_H are shape parameters, δ the scale parameter [18] and we set the location parameter μ and skewness parameter β to 0, in accordance to our estimate. Setting the above as our final distribution, we get an initial noise input distribution as

$$S_k^{\text{in}} = \exp\left(-\frac{k^2\alpha_H\delta\cdot K_{\lambda+1}\left(\sqrt{k^2+\alpha_H H^2\delta}\right)}{\sqrt{k^2+\alpha_H H^2}\cdot K_\lambda\left(\sqrt{k^2+\alpha_H H^2\delta}\right)}\right), \tag{40}$$

which unfortunately is not a well known standard distribution. However, we relate the variances of the input and output distributions assuming $\epsilon_i = \epsilon$ as

$$\left(\frac{\sigma_{F_K}^2}{\sigma_{S_k}^2}\right)_{\text{Hyperbolic}} = \frac{\epsilon^2}{2\gamma N}, \tag{41}$$

which demonstrates how increasing the damping γ or the number of nodes N decreases fluctuations. On the other hand, increasing the noise amplitude ϵ increases the final distribution width. Furthermore, this is exactly the relation of two variances we get when assuming Gaussian noise, see (13) and setting $\sigma_{S_k} = 1$.

Composite distributions

Using generalized Fokker-Planck equations, we also treat composite distributions in the case that the power fluctuations are generated by multiple stochastic processes. Suppose that we are aware that our input noise ξ is not generated from a single but from a composite process Z, i.e., $\xi \sim Z$ with the composite noise as a sum of processes $Z = \sum_{i=1}^{M} X_i$, where any weighting factors in the sum are absorbed in the individual distributions. A sum of different independent processes is also known as a convolution and is easily handled in terms of the characteristic equations of the distributions

$$S_Z^{\text{in}} = \prod_{i=1}^{M} S_{X_i}, \tag{42}$$

see e.g. [11].

As an example, we consider the sum of a normal distribution X and an α-stable distribution Y

$$Z = X + Y, \tag{43}$$

with

$$X \sim N\left(\mu_N, \sigma_N\right), \tag{44}$$

$$Y \sim S\left(\alpha_S, \beta_S, \mu_S, \sigma_S\right), \tag{45}$$

$$S_X^{\text{in}}\left(k\right) = \exp\left[i\mu_N k - \frac{\sigma_N^2 k^2}{2}\right], \tag{46}$$

$$S_Y^{\text{in}}\left(k\right) = \exp\left[i\mu_S k - \left|\sigma_S k\right|^{\alpha_S}\left(1 - i\beta_S \text{sign}\left(k\right)\tan\left(\frac{\pi\alpha_S}{2}\right)\right)\right]. \tag{47}$$

The composite distribution Z has the characteristic function

$$S_Z^{\text{in}}\left(k\right) = S_X^{\text{in}}\left(k\right)\cdot S_Y^{\text{in}}\left(k\right) = \exp\left[ik\left(\mu_N + \mu_S\right) - \frac{\sigma_N^2 k^2}{2} - \left|\sigma_S k\right|^{\alpha_S}\left(1 - i\beta_S \text{sign}\left(k\right)\tan\left(\frac{\pi\alpha_S}{2}\right)\right)\right]. \tag{48}$$

Note that stable distributions are closed under convolution if they have the same stability parameter α_S. In our example this would only be true for $\alpha_S = 2$ which resembles two normal distributions. Otherwise, the distribution of Z is neither a normal nor a stable distribution. We simplify (48) by setting the skewness parameter to zero $\beta_S = 0$ and assuming both distributions have 0 mean $\mu_N = \mu_S = 0$ as follows:

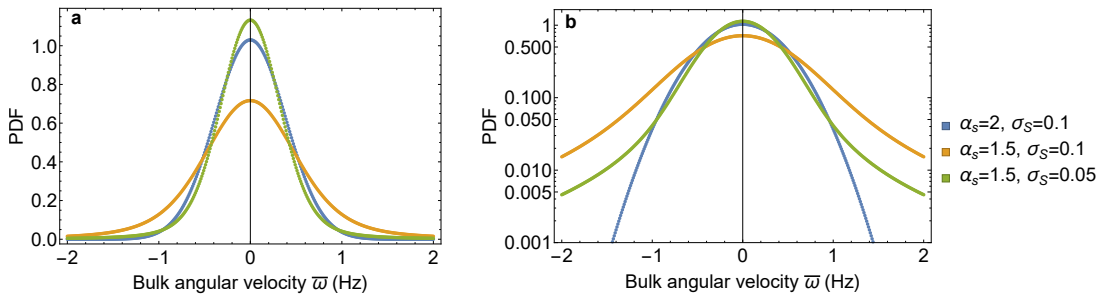
$$S_{Z,\,\beta=0}^{\text{in}}\left(k\right) = \exp\left[-\frac{\sigma_N^2 k^2}{2} - \left|\sigma_S k\right|^{\alpha_S}\right]. \tag{49}$$

We now apply the solution of the generalized Fokker-Planck equation (25) to get the characteristic equation of the output distribution as

$$S^{\text{out}}\left(k\right) = \exp\left[\frac{1}{\gamma}\int_0^k \frac{\ln\left(S_z\right)}{z}\text{d}z\right] \tag{50}$$

$$= \exp\left[-\frac{\sigma_N^2 k^2}{4\gamma} - \frac{\left|\sigma_S k\right|^{\alpha_S}}{\gamma\alpha_S}\right]. \tag{51}$$

The probability density function does not have a closed form but we plot some examples of such a composite distribution in Fig. 5.

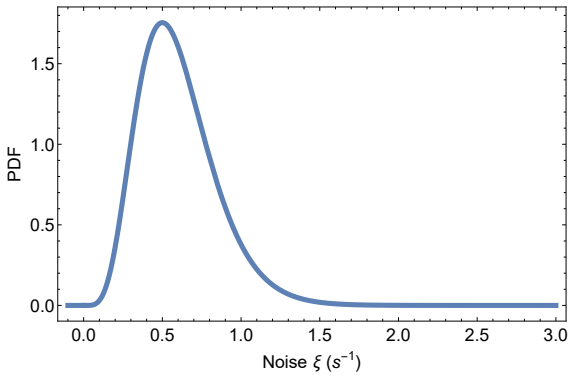


Supplementary Figure 5. The generalized Fokker-Planck equation also handles composite distributions. **a**: We depict the final PDF of a power grid with input noise consisting of a (0-mean) normal distribution $N(0, \sigma_N)$ and a (0-mean, 0-skewness) stable distribution $S(\alpha_S, 0, 0, \sigma_S)$. **b**: We repeat this with a log-scale of the PDF. We fix the standard deviation of the normal distribution at $\sigma_N = 0.1$ but test different values for stability parameter α_S and the scale parameter σ_S of the stable distribution. Especially in the Log-Plot we clearly observe the power-law tails of the composite distribution, noting that stable distributions are normal distributions for $\alpha_S = 2$.

SUPPLEMENTARY NOTE 4

Noise drawn from a Gamma distribution

As was shown in [19, 20], noise in power grids can be well approximated by Weibull, Normal, Log-Normal but also Beta or Gamma distributions which have support $x \in [0, 1)$ and $x \in [0, \infty)$, respectively. Here, we demonstrate how the presented framework for treating arbitrary noise distributions applies to noise drawn from a Gamma with shape parameter $\alpha = 6$ and scale parameter $\beta = 0.1$, as shown in Fig. 6.



Supplementary Figure 6. Probability density function (PDF) of a Gamma noise distribution with shape parameter $\alpha = 6$ and scale parameter $\beta = 0.1$.

The generalized Fokker-Planck equation of the bulk angular velocity $\bar{\omega}$ was solved to give the stationary solution

$$p\left(\bar{\omega}\right) = \mathcal{F}^{-1}\left[\exp\left(-\frac{1}{\gamma}\int_0^k \frac{1}{z}\ln\left(\frac{1}{S_z^{\mathrm{in}}}\right)\mathrm{d}z\right)\right], \tag{52}$$

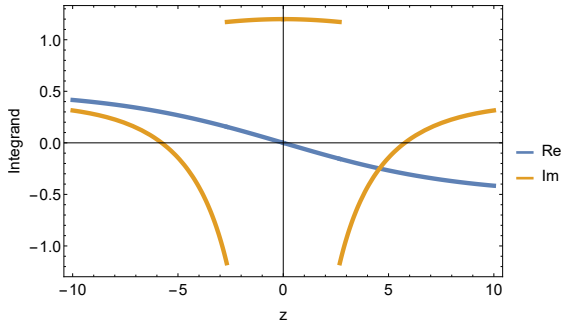
where γ is a damping constant, $\mathcal{F}^{-1}$ is the inverse Fourier transform and S_z^{in} is the characteristic function of the noise. This equation can be applied to arbitrary distributions, as long as the integral over the integrand

$$\mathrm{Integrand}(z) = \frac{1}{z}\ln\left(\frac{1}{S_z^{\mathrm{in}}}\right) \tag{53}$$

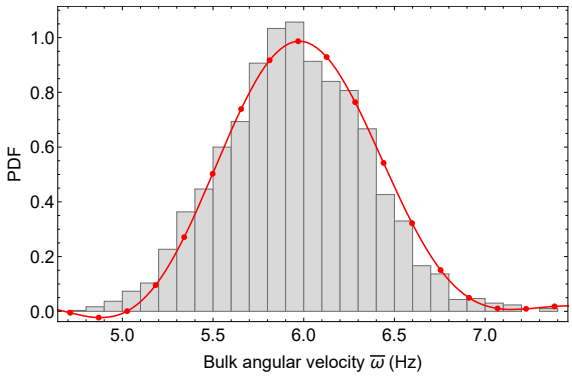
exists.

To treat the Gamma distribution mentioned above, we plot this integrand in Fig. 7. Unfortunately, including the complex logarithm induces jumps in the imaginary part of the integrand and thereby required us to split the integral into parts which can be continuously integrated. As a side note, the Beta distribution [19, 20], even shows divergence of the integrand making it more difficult to treat.

After integrating the integrand (53), the resulting expression is highly oscillatory and can not be treated easily by automatic inverse Fourier transforms, as required in Eq. (52). Hence, we performed a discrete Fourier transform of a finite interval $z \in [-10, 10]$ to finally obtain the distribution of the bulk angular velocity $\bar{\omega}$ as predicted by equation (52) (red line) versus the histogram obtained from stochastic simulations in Supplementary Figure 8.



Supplementary Figure 7. Integrand of the Gamma distribution, which needs to be integrated for any values of z.



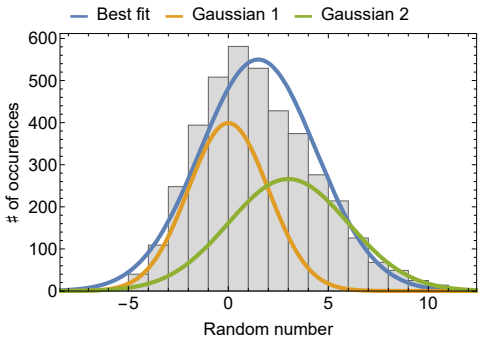
Supplementary Figure 8. Probability density function (PDF) of the bulk angular velocity $\bar{\omega}$ when subject to noise distributed according to a Gamma distribution with shape parameter $\alpha = 6$ and scale parameter $\beta = 0.1$. The red line is based on the prediction of Eq. (52) and the histogram is obtained from stochastic simulations of the power grid being subject to noise from a Gamma distribution.

SUPPLEMENTARY NOTE 5

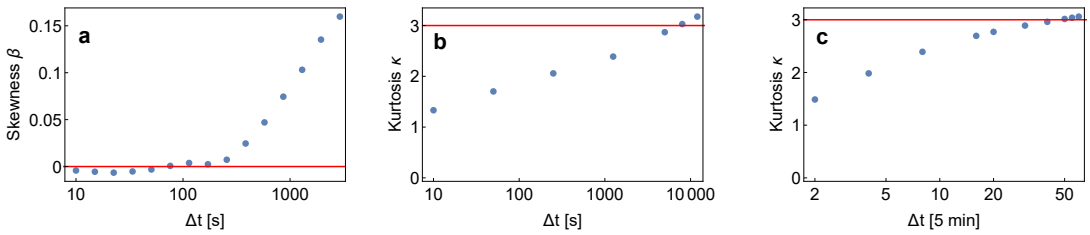
Superstatistics

In the main text we introduced the idea of superstatistics as presented in [21–23] where the total distribution is seen as several aggregated Gaussian distributions with changing damping γ and noise amplitude ϵ. We present an illustration of the concept in Fig. 9, where we combine two Gaussian distributions to form a distribution that is both skewed and heavy-tailed. Note also that skewed distributions could arise from the grid responding differently to frequencies above the reference frequency value than to ones below it.

To extract the long time scale we need to determine the kurtosis given a certain interval length Δt: For heavy-tailed distributions this Δt is the time interval for which the averaged kurtosis of individual intervals of length Δt is equal to $\kappa = 3$. Similarly, for skewed distributions, such as the Great Britain grid, we determine the long time scale by looking for the longest interval such that the skewness averaged over all intervals equals 0, see Fig. 10.



Supplementary Figure 9. Few Gaussian distributions give rise to skewed and heavy-tailed distributions. We display two Gaussian distributions (green and orange) from which we take 2000 samples each to form a histogram. This histogram is no longer well-described by another Gaussian distribution, as can be seen when comparing histogram and most likely Normal plot (blue). Instead the resulting data is skewed with skewness $\beta \approx 0.45$ and has a kurtosis of $\kappa \approx 3.26$, compared to $\kappa^{\text{Gaussian}} = 3$.



Supplementary Figure 10. We determine the long time scale T. **a**: For skewed distributions (Great Britain (GB), Eastern Interconnection (EI)) the long time scale is the time for which the averaged skewness has a minimum. **b-c**: For heavy-tailed distributions (b: Continental Europe (CE) and c: Japan 60Hz) the long time scale is the time for which the averaged kurtosis is approximately $\kappa(\Delta t = T) \approx 3$.

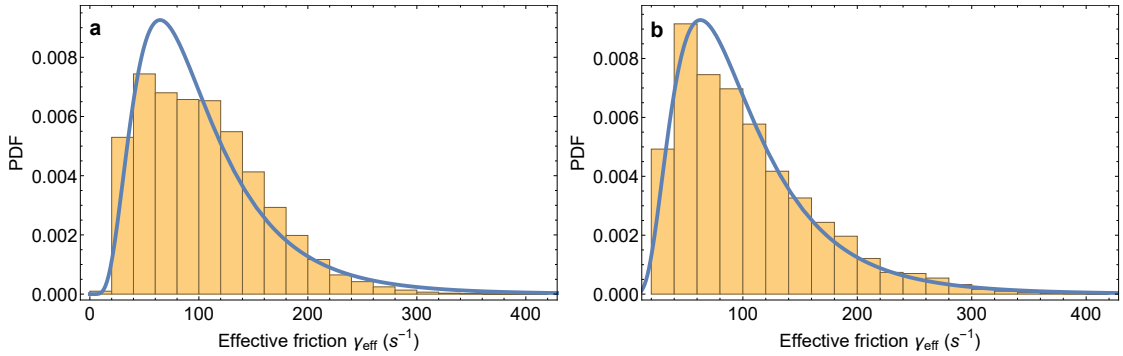
Computing effective friction

As soon as we computed the long time scale T for a given data set, we extract the distribution of the effective friction γ_{eff} which is changing over time as

$$\gamma_{\text{eff}}(t_0) = \frac{1}{\langle x^2 \rangle_{t_0,T} - \langle x \rangle^2_{t_0,T}}, \tag{54}$$

where $\langle ... \rangle_{t_0,\Delta t} = \int_{t_0}^{t_0+\Delta t} ... \mathrm{d}t$. Following [21–23] we expect γ_{eff} to follow a log-normal or alternatively a χ^2 or inverse χ^2 distribution, see e.g. [24, 25] for coverage of log-normal distributions.

In the main text the Japanese data follows a log-normal distribution very well, while we observe larger deviations from the predicted log-normal distribution for the *50Hertz* data set, see Fig. 11.



Supplementary Figure 11. The variance of Gaussian noise follows approximately a log-normal distribution. Plotted are the histograms of the effective friction γ_{eff} value based on the *50Hertz* 2015 trajectory compared to the best-fitting log-normal distribution. **a**: Full year of 2015. **b**: First million (of 30 mill.) data points only, i.e. approximately 10 days. We notice that the data using a shorter trajectory give the better fit to a log-normal distribution. Most likely this is due to several time scales entering the frequency trajectory. Correlations exist on the half an hour, hour, day, weeks and more time scales. Hence, modeling the varying noise with one long time scale is limited and using the full trajectory reveals this problem.

How to derive q-Gaussians

Here we derive how a χ^2 distribution of the effective friction γ_{eff} leads to a q-Gaussian distribution of the frequency, following [21] and provide numerical evidence how a log-normal γ_{eff} distribution also leads to a q-Gaussian.

Consider the Langevin equation

$$\dot{x} = -\gamma x + \epsilon\xi, \tag{55}$$

with noise amplitude ϵ and damping γ. Now define the effective friction $\gamma_{\mathrm{eff}} = \frac{\gamma}{\epsilon^2}$ and assume that it is following a χ^2 distribution:

$$p\left(\gamma_{\mathrm{eff}}\right) = \frac{1}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{n}{2\gamma_0}\right)^{n/2} \gamma_{\mathrm{eff}}^{n/2-1} \exp\left(-\frac{n\gamma_{\mathrm{eff}}}{2\gamma_0}\right), \tag{56}$$

with degree n, mean γ_0 and Gamma function Γ. Next, assume that the changes of γ_{eff} are much slower than the relaxation time scale defined by $\tau = 1/\gamma$ during which the system settles down for one fixed γ_{eff}. Then, the conditional probability to find the system in state x at fixed γ_{eff} is

$$p\left(x|\gamma_{\mathrm{eff}}\right) = \sqrt{\frac{\gamma_{\mathrm{eff}}}{2\pi}} \exp\left(-\frac{1}{2}\gamma_{\mathrm{eff}}x^2\right), \tag{57}$$

and the marginal probability distribution (probability to observe x independent of the value of γ_{eff}) is

$$p\left(x\right) := \int p\left(x|\gamma_{\mathrm{eff}}\right) p\left(\gamma_{\mathrm{eff}}\right) \mathrm{d}\gamma_{\mathrm{eff}} \tag{58}$$

$$= \frac{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{\gamma_0}{\pi n}\right)^{1/2} \frac{1}{\left(1+\frac{\gamma_0}{n}x^2\right)^{n/2+1/2}}, \tag{59}$$

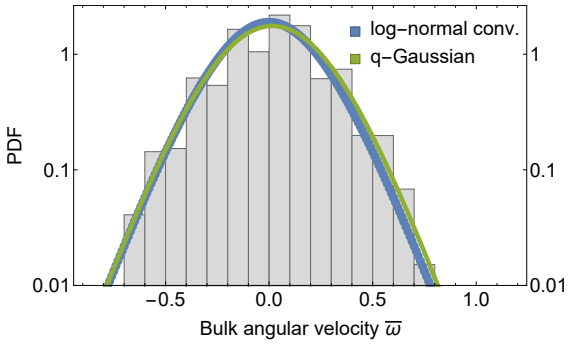
which is a q-Gaussian and can be re-written as

$$p(x) \sim \frac{1}{\left(1+\frac{1}{2}\tilde{\gamma}\,(q-1)\,x^2\right)^{1/(q-1)}}, \tag{60}$$

with $q = 1 + 2/(n+1)$ and $\tilde{\gamma} = 2/(3-q)\gamma_0$.

Let us review the consequences for our data-driven approach. When we record a χ^2 distribution for the effective friction γ_{eff}, it strongly supports the q-Gaussian modeling of the bulk angular velocity $\bar{\omega}$.

However, we did not fit our distribution of γ_{eff} with a χ^2 distribution but used a log-normal distribution instead. Unfortunately, we cannot derive the q-Gaussian distribution analytically based on a log-normal distribution but we can compare the predicted PDF or $p(\bar{\omega})$ when convoluting a log-normal distribution with the histogram and the estimated q-Gaussian. Supplementary Figure 12 displays this comparison for the 60Hz Japan data which works very well: The q-Gaussian based on the original data and the PDF based on the convolution of the log-normal distribution are close to identical, see also [23] for more discussion on the role of log-normal distributions in superstatistics.



Supplementary Figure 12. The log-normal distribution of γ_{eff} predicts nearly the q-Gaussian estimate for the 60Hz Japanese grid. We plot the histogram data of the Japanese 60 Hz region, together with its q-Gaussian estimate. In addition, we compute the expected PDF given that the Gaussian distributions change based on a γ_{eff} following a log-normal distribution. All quantities match very well.

SUPPLEMENTARY NOTE 6

Monte-Carlo simulations

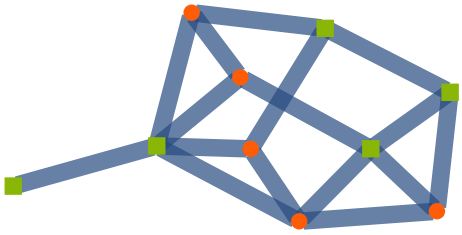
We validate our predicted probability density functions (PDF) based on the Fokker-Planck equation

$$p\left(\bar{\omega}\right)=\sqrt{\frac{\gamma M^{2}}{\pi\sum_{i=1}^{N}\epsilon_{i}^{2}}}\exp\left[-\bar{\omega}^{2}\frac{\gamma M^{2}}{\sum_{i=1}^{N}\epsilon_{i}^{2}}\right] \tag{61}$$

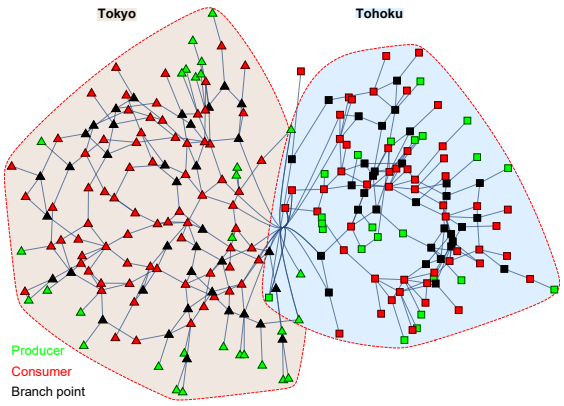
and the generalized Fokker-Planck equation

$$p\left(x,t\right)=\mathcal{F}^{-1}\left[\exp\left(-\frac{1}{\gamma}\int_{0}^{k}\frac{1}{z}\ln\left(\frac{S_{z\cdot e^{-\gamma t}}^{\mathrm{in}}}{S_{z}^{\mathrm{in}}}\right)\mathrm{d}z\right)\right], \tag{62}$$

by simulating the swing equations on realistic power grid networks which are subject to either Gaussian or Lévy-stable noise. As topologies we chose the elementary two node system (one producer connected to one consumer), a 10 node, see Fig. 13 and the Tokyo-Tohoku power grid topology, see Fig. 14.



Supplementary Figure 13. We test the (generalized) Fokker-Planck predictions with simulations on a toy power grid. Here we show the 10 node system with producers (green) and consumers (red). We set $\gamma=0.1$, $P^{-}=-1/s^{2}$ for consumers and $P^{+}=1/s^{2}$ for producers with homogeneous coupling of $K=4/s^{2}$.



Supplementary Figure 14. We test the (generalized) Fokker-Planck predictions with simulations on a realistic power grid. Here we show the joined Tokyo (Triangle) and Tohoku (squares) grid with producers (green), branches (black) and consumers (red) [26]. We set $\gamma=0.1$ and $P^{-}=-1/s^{2}$ for consumers and $P^{+}\approx 2.38/s^{2}$ for producers, with admittance matrix and positions of producers/consumers in the grid given by [26]. For our simulations, we increase the coupling by a factor of $k=15$.

Swing equation simulations Assuming symmetrical coupling $K_{ij}=K_{ji}$, balanced power $\sum_{i}P_{i}=0$ and homogeneous damping to inertia ratio γ we have the following equation of motion

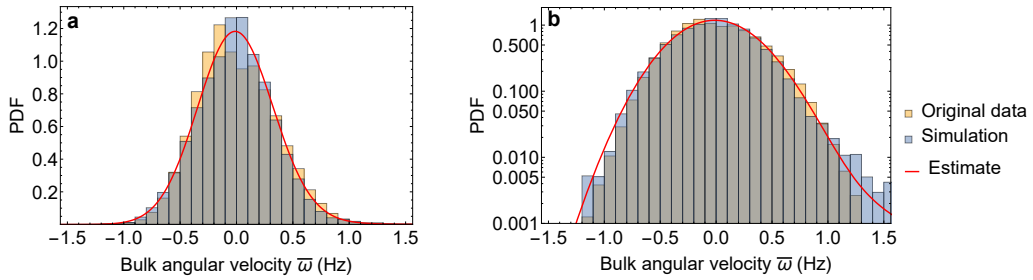
$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{\omega} = -\gamma\bar{\omega} + \bar{\epsilon}\xi, \tag{63}$$

with mean noise amplitude $\bar{\epsilon}$ and noise ξ. To solve this equation, we discretize time into intervals of length Δt and compute

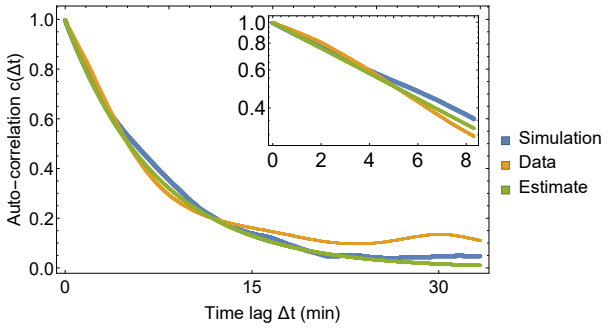
$$\begin{aligned} \Delta\bar{\omega} &= -\gamma \cdot \bar{\omega}\,(t) \cdot \Delta t + X \cdot \sqrt{\Delta t}, \\ \bar{\omega}\,(t + \Delta t) &= \bar{\omega}\,(t) + \Delta\bar{\omega}, \\ t &= t + \Delta t, \end{aligned} \tag{64}$$

with X as our random variable drawn from a previously defined distribution (normal or stable).

Comparison of data and simulations In order to reproduce the data of the ENTSO-E Great Britain grid, we estimate the stability parameter of the measurements as $\alpha_{S,GB} \approx 1.97$, the scale parameter as $\sigma_{S,GB} = 0.2420$ and calculate the damping based on the autocorrelation to be $\gamma_{GB} = 0.00215/s$. The simulations reproduces very well the initial estimate and the original data in the histograms, see Fig. 15 and in the autocorrelation function, see Fig. 16.



Supplementary Figure 15. Simulations reproduce the heavy-tailed ensembles. **a**: We compare the histograms of the original frequency measurements (transformed to angular velocity ω) with simulations and the estimated stable distribution (based on the original data) using the Great Britain 2015 data. **b**: We repeat the plot with a log-scale of the PDF.



Supplementary Figure 16. Simulations reproduce the autocorrelation. We compare the autocorrelation function of the original frequency measurements (transformed to angular velocity ω) with simulations and the estimated exponential decay using the Great Britain 2015 data.

SUPPLEMENTARY REFERENCES

[1] 50Hertz Transmission GmbH. ENTSO-E Netzfrequenz. http://www.50hertz.com/de/Maerkte/Regelenergie/Regelenergie-Downloadbereich (2010-2016).

[2] Réseau de Transport d'Electricité (RTE). Network frequency (2014-2016). URL https://clients.rte-france.com/lang/an/visiteurs/vie/vie_frequence.jsp.

[3] National Grid. Frequency data. http://www2.nationalgrid.com/Enhanced-Frequency-Response.aspx (2014-2016).

[4] Tchuisseu, E. B., Gomila, D., Brunner, D. & Colet, P. Effects of dynamic-demand-control appliances on the power grid frequency. *Preprint at https://arxiv.org/abs/1704.01638* (2017).

[5] Fingrid. Nordic power system frequency measurement data (2015-2016). URL http://www.fingrid.fi/en/powersystem/Power%20system%20management/Maintaining%20of%20balance%20between%20electricity%20consumption%20and%20production/Frequency%20measurement%20data/Pages/default.aspx.

[6] Organization for Cross-regional Coordination of Transmission Operators, Japan (OCCTO). Japanese grid frequency (2016). URL http://occtonet.occto.or.jp/public/dfw/RP11/OCCTO/SD/LOGIN_login#.

[7] Power Information Technology Lab, University of Tennessee, Knoxville and Oak Ridge National Laboratory. FNET/GridEye (2014). URL http://powerit.utk.edu/fnet.html. 1 day data set "20140905".

[8] U.S. Department of Energy. Eia-411: Coordinated bulk power supply and demand program report (2016). URL https://www.eia.gov/electricity/data/eia411/.

[9] ENTSO-E. Monthly production for a specific year for 2015. https://www.entsoe.eu/db-query/production/monthly-production-for-a-specific-year (2016).

[10] ENTSO-E. Statistical factsheet 2015. https://www.entsoe.eu/Documents/Publications/Statistics/Factsheet/entsoe_sfs2015_web.pdf (2016).

[11] Bohm, G. & Zech, G. *Introduction to statistics and data analysis for physicists* (DESY, 2010).

[12] Wolfram Research Inc. Mathematica. Champaign, Illinois (2014).

[13] Machowski, J., Bialek, J. & Bumby, J. *Power System Dynamics: Stability and Control* (John Wiley & Sons, 2011).

[14] Gardiner, C. W. *Handbook of stochastic methods* (Springer Berlin, 1985), 3 edn.

[15] Bollen, M. H. & Gu, I. *Signal processing of power quality disturbances*, vol. 30 (John Wiley & Sons, 2006).

[16] Denisov, S., Horsthemke, W. & Hänggi, P. Generalized Fokker-Planck equation: Derivation and exact solutions. *The European Physical Journal B* **68**, 567–575 (2009).

[17] Polyanin, A. D., Zaitsev, V. F. & Moussiaux, A. *Handbook of First-Order Partial Differential Equations* (CRC Press, 2001).

[18] Fischer, M. Generalized hyperbolic distributions. In *International Encyclopedia of Statistical Science*, 589–590 (Springer, 2011).

[19] Carpaneto, E. & Chicco, G. Probabilistic characterisation of the aggregated residential load patterns. *IET Generation, Transmission & Distribution* **2**, 373–382 (2008).

[20] Soubdhan, T., Emilion, R. & Calif, R. Classification of daily solar radiation distributions using a mixture of dirichlet distributions. *Solar Energy* **83**, 1056–1063 (2009).

[21] Beck, C. Dynamical foundations of nonextensive statistical mechanics. *Physical Review Letters* **87**, 180601 (2001).

[22] Beck, C. & Cohen, E. G. D. Superstatistics. *Physica A* **322**, 267–275 (2003).

[23] Beck, C., Cohen, E. G. D. & Swinney, H. L. From time series to superstatistics. *Physical Review E* **72**, 056133 (2005).

[24] Clark, P. K. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: Journal of the Econometric Society* 135–155 (1973).

[25] Castaing, B., Gagne, Y. & Hopfinger, E. Velocity probability density functions of high reynolds number turbulence. *Physica D* **46**, 177–200 (1990).

[26] Nagata, M. *et al.* Node-wise robustness against fluctuations of power consumption in power grids. *The European Physical Journal Special Topics* **223**, 2549–2559 (2014).