What Kinds of Distributed Generation Technologies Defer Network Expansions? Evidence from France*

Nicolas Astier†
Ram Rajagopal‡
Frank A. Wolak§

January 3, 2021
[Latest version]

Abstract

We estimate the relationship between distributed generation investments and hourly net injections to the distribution grid across over 2,000 substations in France between 2005 and 2018. A 1 MW increase in solar PV capacity has no statistically significant impact on the highest percentiles of the annual distribution of hourly net of injections to the distribution grid. A 1 MW increase in wind capacity is predicted to reduce the 99th percentile of the annual distribution of hourly net injections to the distribution grid by 0.037 MWh. In contrast, a 1 MW investment in a distributed small hydro, non-renewable thermal, or renewable thermal generation unit predicts an almost five times larger MWh reduction in the 99th percentile of the annual distribution of hourly net injections to the distribution grid. A 1 MW investment in distributed solar PV or wind capacity predicts substantial absolute changes in both extremes of the annual distribution of hourly ramp rates of net injections to the distribution grid. For the remaining three distributed generation technologies, a 1 MW capacity increase does not predict a non-zero change in any percentile of the annual distribution of hourly ramp rates of net injections to the distribution grid. These results argue that, at least for the case of France, increases in distributed solar and wind capacity are more likely to lead to increases, rather than decreases, in distribution network investments.

Keywords: distributed generation, renewable energy, electricity grid
JEL: L94, Q42, Q48

*The views expressed in this paper only reflect the views of the authors and no external organization. The authors have received no external funding for this research.

†Precourt Institute for Energy (Bits and Watts initiative), Stanford Unversity, Stanford, CA 94305, nicolas.astier@stanford.edu

‡Department of Civil Engineering, Stanford Univesity, Stanford, CA 94305, ramr@stanford.edu

§Program on Energy and Sustainable Development and Department of Economics, Stanford University, Stanford, CA 94305-6072, wolak@zia.stanford.edu.

1 Introduction

There is considerable debate over the extent to which investments in distribution network connected (distributed) solar photovoltaic (PV) generation capacity reduces the need for future transmission and distribution network investments. Many consultant studies demonstrate significant avoided costs of future network investments associated with deploying distributed solar PV capacity. Muro and Saha (2016) survey several such studies. However, a number of scholars are more skeptical of the existence of significant avoided transmission and distribution network costs associated with investments in distributed solar capacity. For example, Davis (2018) cites two recent studies by Cohen and Callaway (2016) and Cohen et al. (2016) that find small avoided cost benefits, at most on the order of 0.2 cents per produced kWh. Borenstein (2020) takes what he calls a "macro" approach to this issue and argues that the avoided cost of grid investments associated with replacing 1 kWh of grid generated electricity with rooftop solar energy is no more than 1.2 cents/kWh and certainly well under 1 cent/kWh.

This debate is increasingly relevant to electricity supply industries where distributed generation capacity, primarily in the form of solar PV capacity, is rapidly increasing. According to the United States (US) Energy Information Administration, total annual capital investments in distribution networks by major US utilities serving about 70% of the country's electricity demand have more than doubled between 1996 and 2017 to more than 25 billion dollars in 2017.[1] If distributed solar generation reduces the need for distribution network investments by the amounts claimed in a number of consultant studies, annual distribution network investments avoided from the current annual deployment of distributed solar in the US alone could easily reach over a billion dollars.

We contribute to this debate with an empirical analysis of the relationship between distributed generation investments and hourly net injections to the distribution grid at over 2,000 substations

[1] https://www.eia.gov/todayinenergy/detail.php?id=36675

in France, where annual capital investments in distribution networks exceed 3 billion euros (Commission de Régulation de l'Energie, 2020). We use hourly net injections at these substations and the deployment of about 25 gigawatts (GW) of distributed generation capacity between 2005 and 2018 to recover estimates of how different distributed generation technologies impact the utilization of distribution network capacity. This data is used to assess the impact of a 1 MW increase in each of five distributed generation technologies – solar PV, wind, small hydro, renewable thermal, and non-renewable thermal – on the percentiles of the annual distribution of hourly net withdrawals from the distribution network.

We also estimate the impact of investments in these five distributed generation technologies on the percentiles of the annual distribution of hourly ramp rates of net injections to the distribution grid. This analysis quantifies how much different percentiles of the annual distribution of hourly differences in net injections to the distribution grid change as a result of a 1 MW increase in each of these distributed generation technologies.

We find that a 1 MW investment in distributed solar PV capacity has no statistically significant impact on the highest percentiles of the annual distribution of hourly net injections to the distribution grid. A 1 MW investment in distributed wind capacity is predicted to reduce the 99th percentile of the annual distribution of hourly net injections to the distribution grid by 0.037 MWh. In contrast, a 1 MW investment in a distributed small hydro, non-renewable thermal, or renewable thermal generation unit predicts at a least a 0.12 MWh reduction in the 99th percentile of the annual distribution of hourly net injections to the distribution grid.

A 1 MW investment in distributed solar PV or wind capacity predicts similar absolute value changes in both extremes of annual distribution of the hourly ramp rates of net injections to the distribution grid. A 1 MW increase in either wind or solar PV distributed generation capacity predicts a 0.15 MW decrease in the 1st percentile of the annual distribution of the hourly ramp

rates of net injections to the distribution network.[2] For the 99th percentile, a 1 MW increase in wind or solar PV capacity predicts a 0.14 MW increase. For each of the remaining three distributed generation technologies, a 1 MW increase in installed capacity does not predict a non-zero change in any percentile of the annual distribution of hourly ramp rates of net injections to the distribution grid.

Taken together, these results argue that at least for the case of France, increases in distributed solar and wind capacity are more likely to require increases, rather than decreases, in future network investments. Indeed, investments in these two intermittent distributed generation technologies predict very small or zero reductions in the highest percentiles of the distribution of hourly injections to distribution grids, and additional distribution network investments may be necessary to manage the significantly larger (in absolute value) extremes of the annual distribution of hourly ramp rates for net injections to the distribution grid.

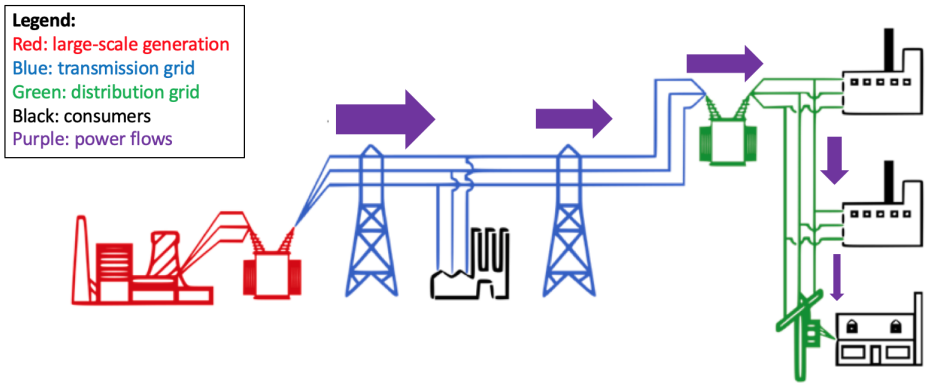
The remainder of the paper is organized as follows. Section 2 describes the historical operation of electricity supply industries and how investments in distributed generation capacity are changing this operating paradigm. This background explains why investments in distributed generation capacity can reduce the need for future investments transmission and distribution network capacity. Section 3 describes our approach to quantify the impact of investments in different distributed generation technologies on the need for future investments in transmission and distribution network capacity, and how our work differs from previous work on this topic. Section 4 describes the data sources used for our analysis. Section 5 details our empirical strategy. Section 6 presents our main results. Section 7 discusses both the limitations and policy implications of our results. Finally, Section 8 concludes.

[2] We compute hourly ramp rates as the difference between two subsequent hourly net load levels (in MWh). Because we measure changes in load levels over the course of 1 hour, we use MW as the unit for hourly ramps.

2 Historical Industry Operation and Distributed Generation

Historically, electricity supply industries consisted of generating electricity at large-scale facilities that took advantage of economies of scale in production, and then transmitting it at a high voltage to local distribution grids.[3] This electricity was then transformed to a lower voltage and transferred to final consumers through a local distribution grid. Figure 1 shows a large-scale generation unit, the high voltage transmission grid, a transformer station which steps down the voltage, and the distribution grid moves the electricity at a safe voltage to final consumers. As illustrated by the arrows in Figure 1, historically electricity flowed in one direction, from large-scale generation units to final consumers.

Figure 1: Historical structure and operation of the power grid (illustrative, adapted from US Department of Energy)



Starting in the early 2000s, electricity supply industries in many jurisdictions saw significant investments in more environmentally friendly small-scale generation units, located near final consumers. In most cases, wind and solar PV were the primary technologies deployed. Investments in small hydro, renewable thermal and non-renewable thermal technologies also occurred. Because

[3]Production and transmission at high voltage limits the magnitude of the energy losses between the point of production and the point where the energy is injected to the distribution grid.

of their small size these generation units could be connected to the distribution grid. Figure 2 presents a sample electricity supply industry with distribution network connected generation, what is typically referred to as distributed generation.

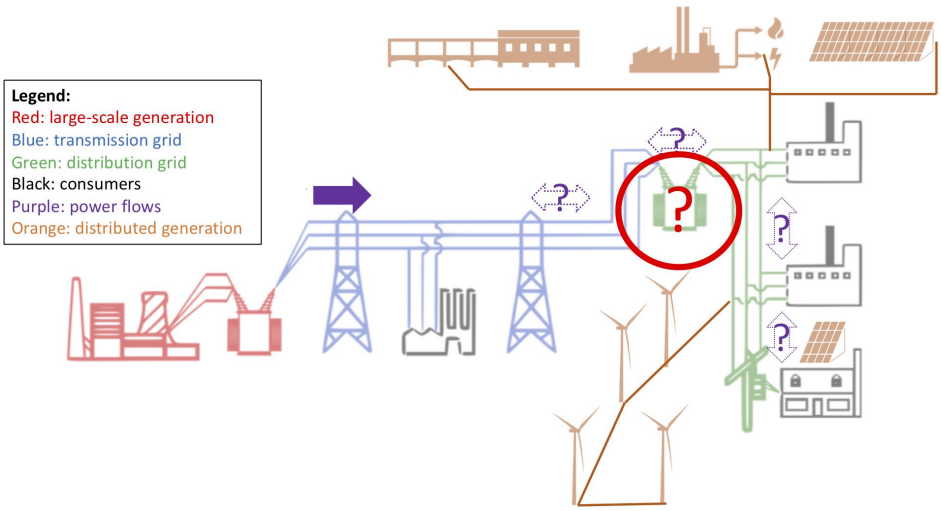
Because they are connected closer to final consumers, distribution generation units may reduce the magnitude of grid power flows. In addition, with sufficient distributed generation capacity, electricity may even no longer always flow from large-scale generation units to final consumers. Distributed generation capacity could thus reduce both the need for transmission capacity to move energy from large-scale generation units to final consumers and distribution network capacity to move the high-voltage energy from the transformer in the red circle Figure 2 to final consumers.

For the case of non-controllable distributed generation, such as wind and solar PV capacity, when these resources produce energy relative to when system peaks occur determines the extent to which investments in these technologies can actually reduce the need for future transmission and distribution network investments. In contrast, controllable distributed generation units, such as renewable and non-renewable thermal generation units, can be operated to reduce the need for future transmission and distribution network investments. Specifically, if these units operate during periods when the final demand for electricity is at or near its annual peak, they can reduce the need for future transmission and distribution network expansions.

Transitioning from the simplified model of an electricity supply industry in Figures 1 and 2 to the case of France, Figure 3 illustrates the overall structure of the country's electricity grid,[4] and highlights examples of distributed generation units. Distributed generation units connect to the grid at consumers' premises (panel (e)) or to an upstream substation that was initially designed to supply end-consumers (panels (f) and (g)). The interface between the transmission and distribution

[4]Panels (a), (b), (e), (f) and (g) are screenshots taken from: https://www.arcgis.com/home/webmap/viewer.html?webmap=02d413bcbe184384ba424fc40f9b8ce8. Panels (c) and (d) are screenshots taken from: https://www.enedis.fr/cartographie-des-reseaux-denedis

Figure 2: Schematic impact of distributed generation on the operations of the power grid (illustrative, adapted from US Department of Energy). The red circle highlights the role of distribution substations as the interface between the transmission and distribution grids.



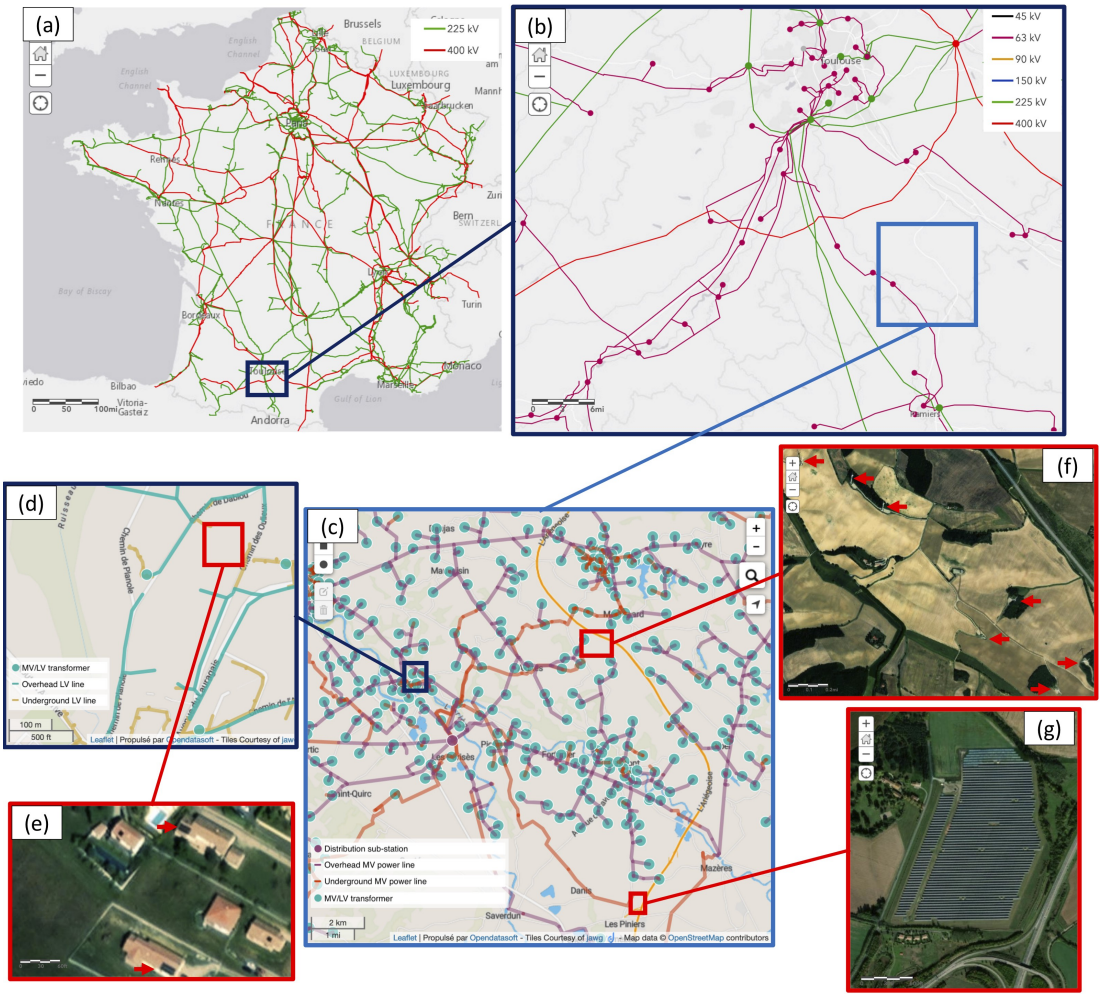
grids consists in assets called distribution substations. In France, these substations typically host 63/20 or 90/20kV transformers.[5] Although the exact voltage levels used in the electricity grid differ across countries, these substations would correspond to the edge of the sub-transmission network in the United States (US Department of Energy, 2015).

Power flows measured at a given substation consist in the aggregation of all the consumption and generation of users connected downstream the substation. We thus refer to these power flows as *net* load levels, with the convention that positive (resp. negative) values correspond to hours where local consumption exceeds local generation (resp. local generation exceeds local consumption).

Over the past two decades, France has experienced a tremendous increase in the installed capacities of different distributed generation technologies. The vast majority of distributed generation installations belong to one of the following categories: wind, solar, small hydro, and thermal units

[5] Very dense urban areas are however supplied by 225/20kV transformers.

Figure 3: (a) French transmission grid (225 and 400kV power lines) as of 1 September 2016 ; (b) Zoom on the high voltage (63kV) sub-transmission grid for a given area ; (c) Zoom on the area around a single distribution substation (purple dot), with the corresponding medium voltage (20kV) power lines and MV/LV transformers ; (d) Zoom on a given neighborhood with its MV/LV transformers and LV power lines. Panels (e)-(f) show examples of distributed generation units connected to the substation. Panel (e) highlights two houses with residential PV systems (about 5-6kW each). Panel (f) is a 14.3MW wind unit (arrows point to the 7 wind generators). Panel (g) is a 10.5MW utility-scale PV unit (sources: ArcGIS, RTE, Enedis).

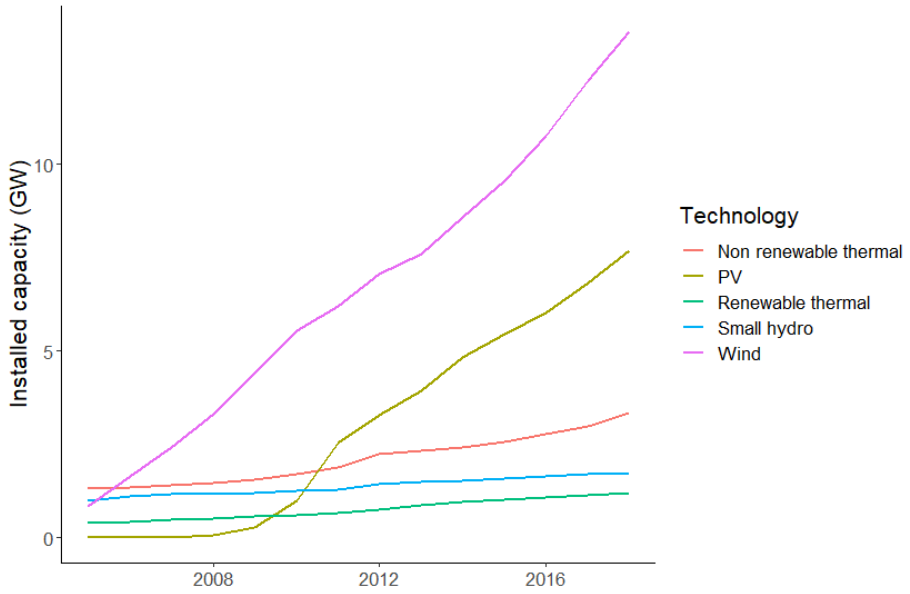


using either renewable (e.g. wood, waste) or non-renewable (e.g. gas, diesel) fuels. Figure 4 shows how the installed distributed generation capacities we observe have evolved between 2005 and 2018.[6]

[6] As discussed in Section 4, our final dataset includes the vast majority of distributed generation units located in

Although all technologies exhibit a significant upward trend, solar and wind have by far experienced the most impressive growth, from under 1 GW to over 7 GW and 13 GW, respectively. As of 2018, there was approximately 28 GW of distributed generation in France, which is non-negligible for a power system whose historical peak demand is about 100 GW.

Figure 4: Total installed capacities of distributed generation by year (in GW) and technology observed in our final dataset (see Section 4).



Distributed generation units benefit from different types of public support mechanisms, whose nature varies by technology and has changed over time.[7] Although the total installed capacity of each technology is largely influenced by targets set by the government, the location and characteristics of distributed generation units are the result of decentralized decisions. Most financial support schemes are not location-specific and grid connection charges are only mildly differentiated at the level of administrative regions, which on average aggregates about two hundred distribution systems. As a result, the growth in distributed generation capacities at the substation level is not

mainland France.

[7]For more details, see for example https://www.cre.fr/Transition-energetique-et-innovation-technologique/soutien-a-la-production/Dispositifs-de-soutien-aux-EnR.

the result of a centralized optimization. In particular, the observed location and characteristics of distributed generation units may differ significantly from the outcome one would obtain from an idealized planning exercise trying to minimize their impact on the electricity system.

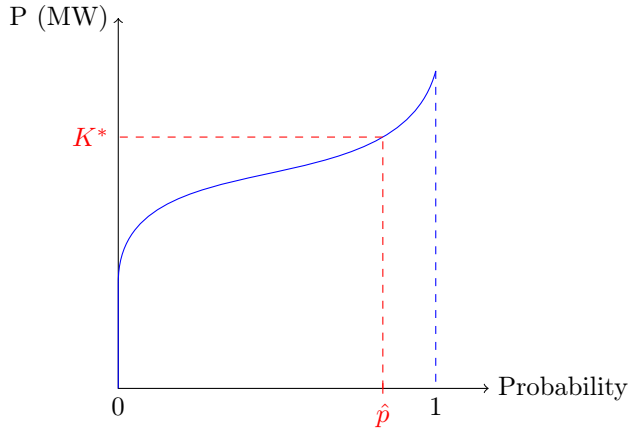
3 Measuring the Impact of Distributed Generation Capacity

The most direct strategy to answer our main research question is to estimate the impact that an additional MW of each distributed generation technology has on the power flows at distribution substations, which correspond to the red circle on Figure 2.

We first study how the percentiles of the annual distribution of the hourly net load levels at each substation change in response to investments in different distributed generation technologies over our sample period. The percentiles of the annual distribution of the hourly net load levels indeed map to what is known in the power systems literature as the load duration curve. Up to a change in the direction of the x-axis, the load duration curve corresponds to the inverse cumulative distribution function of hourly net load levels. This curve allows grid planners to assess the probability (measured in number of hours per year) that net load may exceed a given level. Planning rules typically use one or several probability thresholds to decide on the size of a given network component (Persoz et al., 1984). Figure 5 illustrates this procedure. Given the load duration curve faced by a substation (in blue), a grid planner can infer how much capacity K^* is needed to ensure a given reliability level $\hat{p}$.

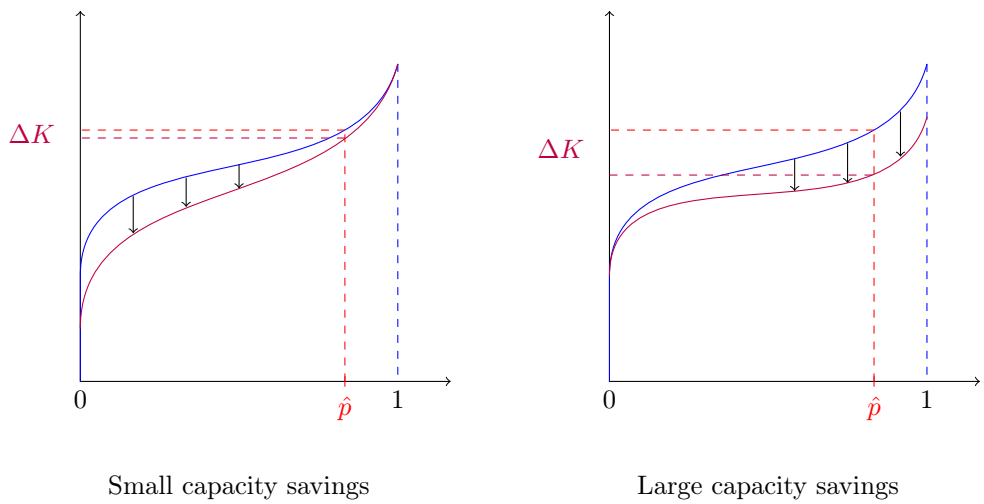
Whether or not a given distributed generation unit decreases significantly the grid capacity needed to ensure a given level of reliability of supply will largely depend on the extent to which this unit generates electricity during the peak hours for that substation. Figure 6 illustrates this intuition. If the output from the distribution generation unit is not coincident with local peak

Figure 5: Illustrative use of a load duration curve for grid planning purposes



demand hours, it is unlikely to defer grid expansions (left panel). By contrast, its impact on needed network capacity will be much larger if the unit generates electricity during peak hours (right panel).

Figure 6: Illustration of how distributed generation may shift the load duration curve of a substation, decreasing significantly or not the network capacity needed to meet a given reliability standard



Small capacity savings

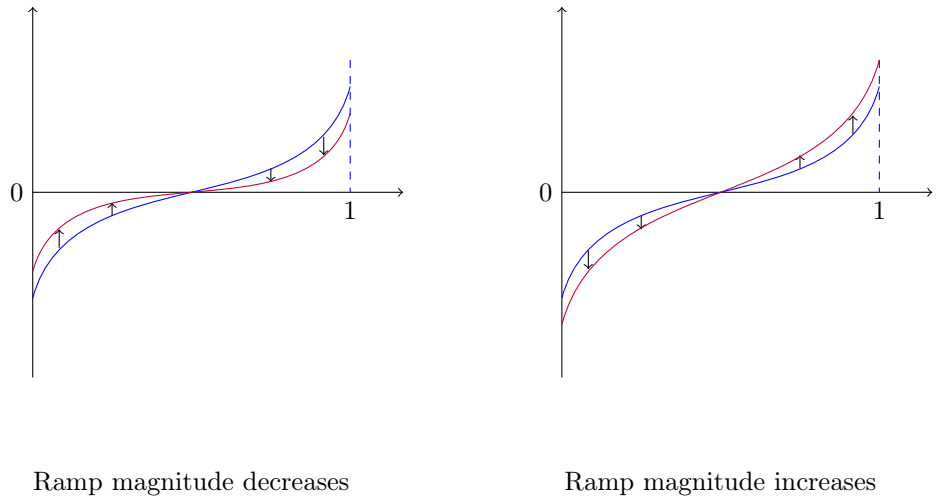
Large capacity savings

In France, peak electricity demand is reached during the winter, due to a large electric heating load. Because co-generation units produce both heat and electricity, they seem likely to generate electricity under such circumstances. By contrast, rooftop solar units have a lower output during

the winter and stop producing completely after sunset. As a result, they seem less likely to generate electricity when most needed. In what follows, we will quantify empirically this intuition.

In recent years, power system engineers have also paid increased attention to very high variations in net load levels over short periods of time, known as "ramps". The most well-known example is perhaps the so-called "duck curve" in California,[8] where the rapid decrease in PV output at sunset makes it necessary to ramp up more than 10,000 MW of controllable generation capacity in three hours. At the distribution grid level, it is likely that very large variations in the net load levels will put more stress on electrical components. This may in turn accelerate equipment aging and make operational constraints (e.g. phase balancing, voltage bounds) more likely to bind, leading to increase investments in local distribution networks.

Figure 7: Illustration of how distributed generation may shift the ramp duration curve of a substation, decreasing or increasing the magnitude of hourly ramps



We thus also look at the impact of distributed generation on the distribution hourly ramps, defined as the difference between two consecutive hourly net load levels. More precisely, we investigate how different distributed generation technologies are changing the shape of the ramp duration

[8] https://www.caiso.com/documents/flexibleresourceshelprenewables_fastfacts.pdf

curve supplied by substations. For a given substation in a given year, the hourly ramp in hour h is defined as the difference between the load level in hour $h+1$ and the load level in hour h. Hourly ramps can then be sorted in increasing order to build a ramp duration curve. By construction, the integral of the ramp duration curve is close to zero.[9] As a result, the ramp duration curve starts at negative values, which correspond to hours during which net load is decreasing at the highest rates, and ends at positive values, which correspond to hours during which net load is increasing at the highest rates. The flatter the ramp duration curve, the less severe are the observed ramps. In particular, the two extremities of the ramp duration curve materialize the most extreme variations in net load levels.

Figure 7 illustrates two contrasted ways in which a distributed generation unit may impact the ramp duration curve faced by a given substation. In both panels, the blue curve represents the pre-existing ramp duration curve and the purple curve the ramp duration curve after the addition of a distributed generation unit. Broadly speaking, two situations may be envisioned: distributed generation may either reduce the severity of ramps, rotating the ramp duration curve clockwise (left panel) ; or it may exacerbate their magnitude, rotating the ramp duration curve in the other direction (right panel). Which situation arises in practice is a question of empirical nature, whose answer may depend on the particular technology of the distributed generation unit. In what follows, we will thus investigate it empirically.

A large literature focuses on the on-going transition of the electricity generation mix. Numerous studies take a system-wide perspective and simulate the behavior of a given electricity system under a variety of prospective scenarios. Their main objective has been to explore the consequences of having large amounts of intermittent renewable capacities in the generation mix. These studies for example found that the marginal social value of wind and PV decreases with installed capacities

[9] If $R(h) \equiv L(h+1) - L(h)$ is the hourly ramp for hour h (where $L(h')$ is the load level in hour h'), then $\sum_{h=1}^{8759} R(h) = L(8760) - L(1)$. This difference is negligible relative to the total load $\sum_{h=1}^{8760} L(h)$ supplied by the substation over the course of a full year.

(Bistline, 2017), making non-intermittent resources valuable assets of the generation mix (Sepulveda et al., 2018). Recent work argues that the sharp decline in the capital costs of wind and PV could nonetheless enable an affordable near-complete decarbonization of the power sector in the medium run (Phadke et al., 2020). However, these studies only account for (simplified) transmission grid constraints (if any), and ignore completely sub-transmission and distribution grids. Other studies do explore spatial heterogeneity and estimate the social value of intermittent renewable technologies across numerous countries and/or electricity systems (Callaway et al., 2018; Gillingham and Ovaere, 2020). However, sub-transmission and distribution grids are again either ignored or accounted for using evidence from a handful of distribution systems. Because a large fraction of wind and PV generation comes from small units connected to distribution grids, this simplification may overlook important operational challenges. Indeed, as the amount of distributed resources connected to distribution grids increases, distribution grid operators may have to revisit how they operate their network. Such claims have however mainly relied on qualitative arguments so far (Burger et al., 2019). The power systems literature does provide detailed case studies looking at the operating constraints faced by distribution grids with increasing amounts of distributed resources. For example, Navidi et al. (2019) propose a two-layer architecture to coordinate the operation of distributed resources using credible communication protocols. Empirical validations however rely on simulations using standardized distribution grids. Cohen et al. (2016) look in more details at the impact of distributed PV on a handful of distribution feeders, using a mix of real-life and simulated data. In both cases, assessing the external validity of the obtained results proves challenging.

Our work differs from these studies by utilizing 14 years of hourly net load levels for over 2,000 distribution systems, as well as annual investments in solar PV, wind, small hydro, renewable thermal, and non-renewable thermal generation capacity in each of these distribution networks over the same time period, to disentangle the impact of investments in these five distributed generation

technologies on the percentiles of the annual distribution of net load levels and the percentiles of the annual distribution of hourly ramp rates. Our empirical approach has the advantage of using the actual hourly utilization of the distribution network to determine whether distributed generation investments are likely to allow reductions in future investments in distribution network capacity.

4 Data

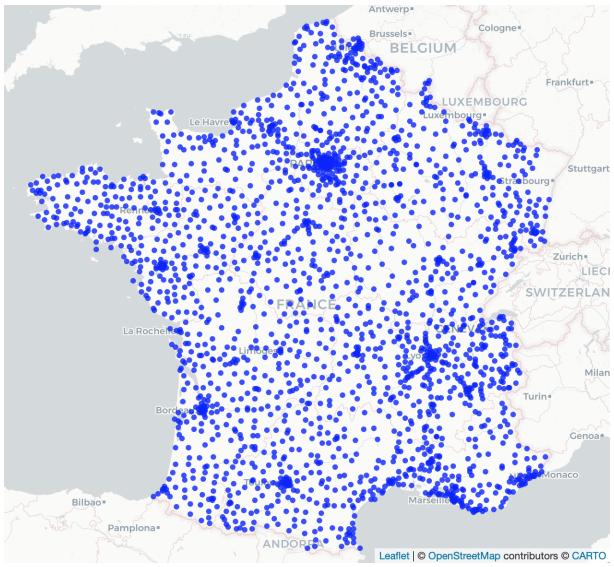
This section describes the two datasets we use in our empirical work. The first is hourly net load levels at over 2,000 distribution substations in France for 2005 through 2018. The second dataset is the location, commissioning date, and nameplate capacity for investments in five types of distributed generation technologies over our sample period.

4.1 Substation hourly net load levels

We observe hourly net load levels for 2,216 substations.[10] 2,112 are observed over the 14-year period. Out of the remaining 114 substations, 90 correspond to substations commissioned during the period. We discuss how we account for entry/exit at the end of Appendix A. These substations are operated by the main distribution system operator in France, which oversees 95% of the distribution grid. The substations we observe thus cover the vast majority of mainland France electricity demand from low and medium voltage customers (Figure 8). As explained above, substation hourly load levels are net of distributed generation. In other words, for each substation and each hour between 1 January 2005 and 31 December 2018, we observe the difference between total consumption and total generation during that hour from all users connected downstream the substation.

[10] We are very grateful to the French National Regulatory Authority, the Commission de régulation de l'énergie (CRE), for granting us access to this data under a non-disclosure agreement. CRE need not share the views and opinions expressed in this paper, which are the responsibility of the authors alone.

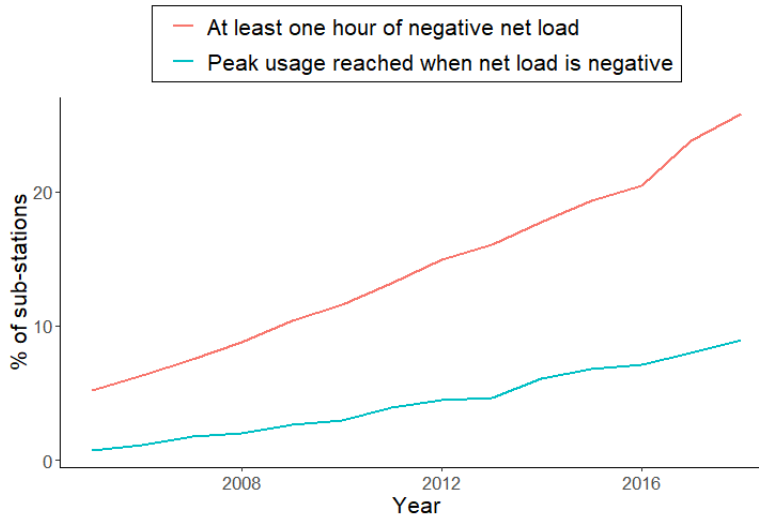
Figure 8: Location of the observed distribution substations



From this raw dataset of 250+ million observations, we can compute summary statistics for the distribution of hourly net load levels at a given substation in a given year. Appendix C provides detailed information on a number of classic summary statistics, such as the mean or standard deviation of hourly net load levels. The main take-away from this analysis is that the most dramatic changes that occurred between 2005 and 2018 relate to reverse power flows, that is to hours during which local generation exceeds local consumption. During such hours, power flows in the opposite direction as the one displayed on Figure 1 at the distribution substation.

Figure 9 shows that the fraction of substations that have experienced at least one hour of reverse power flows has increased from 6% in 2005 to more than 25% in 2018. In other words, over a quarter of substations now have to deal with hours during which electricity is flowing from the distribution to the transmission grid. In addition, the fraction of substations for which peak usage (in absolute value) was reached during an hour with reverse power flows has increased from less than 1% in 2005 to almost 9% in 2018. This new use pattern of the distribution grids suggests that, in the mid-run, the lowest quantiles of the distribution of hourly net load levels will require curtailing/storing excess

Figure 9: Evolution of the percentage of substations that (i) experienced reverse power flows in a given year; (ii) reached their peak usage (in absolute value) when net load was negative, that is during an hour where they were moving electricity from the distribution to the transmission grid.



generation, expanding the grid, or both. When discussing our results, we will thus also dedicate attention to the lowest hourly net load levels supplied by substations.

As described in Section 5, our empirical strategy uses a given substation in a given year as the unit of observation, which gives us 30,000+ units. For each substation and each year, we observe a distribution of hourly net load levels. Each distribution can be used to build both a load duration curve and a ramp duration curve. In other words, we observe a load/ramp duration curve for each substation and each year. We keep track of these curves by extracting the 1st, 10th, 25th, 50th, 75th, 90th and 99th percentiles of the distribution of hourly net load levels (resp. hourly ramps) for each substation in each year.[11] Our final data on observed power flows thus consist in fourteen panels (7 percentiles for both the load and ramp duration curves), where the temporal dimension is years and the spatial dimension consists in distribution substations.

[11] Given the monotonous nature of load/ramp duration curves, extracting different percentiles would yield similar results to the ones we report.

4.2 Distributed Generation Capacity

Information on distributed generation units is retrieved from the public inventory of French power plants.[12] This inventory provides detailed information on the universe of power plants in France, based on data from transmission and distributor system operators. As of 31 December 2018, 44,000+ observations were listed in this dataset, out of which 42,000+ correspond to installations located in mainland France and connected to a distribution grid. These sites range from a few kW to 50 MW. With a negligible number of exceptions,[13] distributed generation units belong to one of the five following categories: wind, solar,[14] small hydro, and thermal units using either renewable (e.g. wood, waste) or non-renewable (e.g. gas, diesel) fuel. We also observe the installed capacities[15] and commissioning dates[16] of distributed generation units.

Two types of observations are listed in the inventory. First, most observations (28,000+) correspond to distributed generation units listed individually. We observe the location of these units down to the county or sub-county level. Importantly, for the vast majority of these observations (26,000+), we also perfectly observe the identifier of the upstream substation to which they

[12] Publicly available at: www.data.gouv.fr/en/datasets/registre-national-des-installations-de-production-delectricite-et-de-stockage-au-31-decembre-2018/. The dataset was downloaded on 28 August 2020.

[13] These exceptions are (i) 1 geothermal unit located in a county most likely supplied by a substation we do not observe ; (ii) 2 pilot units harnessing ocean energy ; (iii) 3 battery storage units that were commissioned only very recently ; and (iv) 51 units labeled as "other technology" due to missing information or mistakes. We are able to infer the technology of 33 out these latter 51 units based on the fuel used, their name or an internet search of their characteristics (name, location, etc.).

[14] Only 6 units out of tens of thousands are labeled as thermodynamic solar, the rest of units consisting in photovoltaic panels. The paper hence uses interchangeably the terms "solar" and "PV".

[15] The inventory makes a distinction between the installed capacity of a unit and its contracted connection capacity with the grid operator. In practice, a single capacity metric is available for 38,000+ observations, suggesting that system operators often used these concepts interchangeably when entering data into the inventory. For observations with both installed and connection capacities information, both figures are similar (either equal or with an absolute difference lower than 10% of installed capacity) for almost 4,000 units, in which case we use the reported installed capacity. 95 observations have neither installed nor connection capacities information, but do provide another capacity metric which we use as a proxy. Finally, 141 installations have reported installed and connection capacities that differ by more than 10%. For these units, we compute the implied capacity factors from reported energy production (when available) and choose the capacity that imply the most credible (extrapolating to similar units when energy production is not available).

[16] The inventory also makes a distinction between the date at which a unit is commissioned and the date at which its connection to the grid is effective. Again, for the vast majority of observations (39,000+), either both dates are identical or a single date is reported. For the remaining observations, the later date is taken into account since any discrepancy between both commissioning and connection dates is most likely due to a ramping up period during which the generation unit does not produce at full capacity.

connect.[17] This information allows us to match accurately these installations to the distribution substations for which we observe net load levels. Second, the remaining 14,000+ observations listed in the inventory do not correspond to individual units. Indeed, in order to preserve individual owners' privacy, the majority of smaller units (<36kW) are aggregated by groups of at least 10 installations (see Appendix A). We do not directly observe the upstream substation to which such aggregated observations connect. Although they represent a third of the observations listed in the public inventory, they add up to a relatively small total capacity due to their modest unitary size compared to larger-scale installations. Table 1 summarizes, for each distributed generation technology, the total capacity as of 31 December 2018 of installations for which we perfectly observe the upstream substation, as well as of installations for which we do not observe it and that may be either aggregated or listed individually.

Table 1: Installed capacities of distributed generation (MW) as of 31 December 2018 in mainland France by technology and availability of upstream substation information. The last column computes, for each technology, the percentage of total installed capacities for which upstream substation information is available.

Technology	Units with known substation (MW)	Units with unknown substation (MW)		Fraction known
		listed individually	aggregated	
Wind	13,012	987	14	92.9%
PV	5,787	247	1,739	74.5%
Small hydro	1,906	83	5	95.6%
Renewable thermal	1,158	81	9	92.8%
Non renewable thermal	3,328	218	26	93.2%

We implement an assignment procedure to infer the substation to which distributed generation installations whose upstream substation is unknown are most likely to connect. To do so, we leverage our knowledge of both the GPS coordinates of the substations and the location of generation units down to the (sub)county level. Indeed, (sub)counties represent a sub-division of mainland France into over 45,000 spatial units, which is an order of magnitude higher than the number of distribution

[17] Because the inventory has been published every year starting in 2017 and individually listed units generally have a unique identifier, we were able to retrieve substation information for about 300 additional observations from the 2017 and 2019 inventories.

substations.[18] This very fine spatial granularity allows us to form reasonable guesses about which substation is most likely to supply electricity to a given spatial unit. Appendix A provides more details on our assignment procedure. Appendix B presents sensitivity analyses. We find that our results are robust to alternative specifications of the assignment procedure, including ignoring altogether installations whose upstream substation is unknown.

Table 2: First columns: summary statistics of substation level installed capacities (in MW) by technology. The unit of observation is a given substation in a given year (N = 30,091). Last column: total capacity by technology as of 2018 in our final dataset, both in absolute value and as a percentage of the total capacity listed in the public inventory of power plants.

Statistic	Mean (MW)	St. Dev. (MW)	Min (MW)	Pctl(25) (MW)	Pctl(75) (MW)	Max (MW)	Total 2018 (% inventory)
Wind	3.114	11.499	0	0	0	189	13,567 (96.8%)
PV	1.392	3.829	0	0.01	1.2	101	7,695 (99.0%)
Small hydro	0.641	2.642	0	0	0	63	1,717 (86.1%)
Renewable thermal	0.354	1.709	0	0	0	35	1,198 (96.0%)
Non renewable thermal	0.974	2.743	0	0	0	45	3,334 (93.3%)

In the end, our dataset on distributed generation records, for every distribution substation and each year, the installed capacities (as of 31 December of each year) of distributed generation connected to this substation, broken down by technology. Table 2 provides descriptive summary statistics for each technology, where the unit of observation is a given substation in a given year. Not surprisingly, because the installed capacities of distributed generation were small in 2005 (see Figure 4), a large number of observations are at zero. Nonetheless, we observe a significant amount of variation which we will be able to leverage. The last column in Table 2 shows the total installed capacity by technology in our final dataset as of 31 December 2018, both in absolute value and as a fraction of the total capacity listed in the public inventory of power plants (see Table 1). Quite remarkably, we are able to match the vast majority of distribution generation units currently installed in mainland France to the substations for which we observe hourly net load levels.[19]

[18] Our assignment procedure relies on a sub-division of mainland France into 45,508 spatial units, with a mean surface of 11.9 km^2 (4.6 miles squared).

[19] The remaining capacities seem most likely to be connected to substations that we do not observe. For example,

5 Empirical Strategy

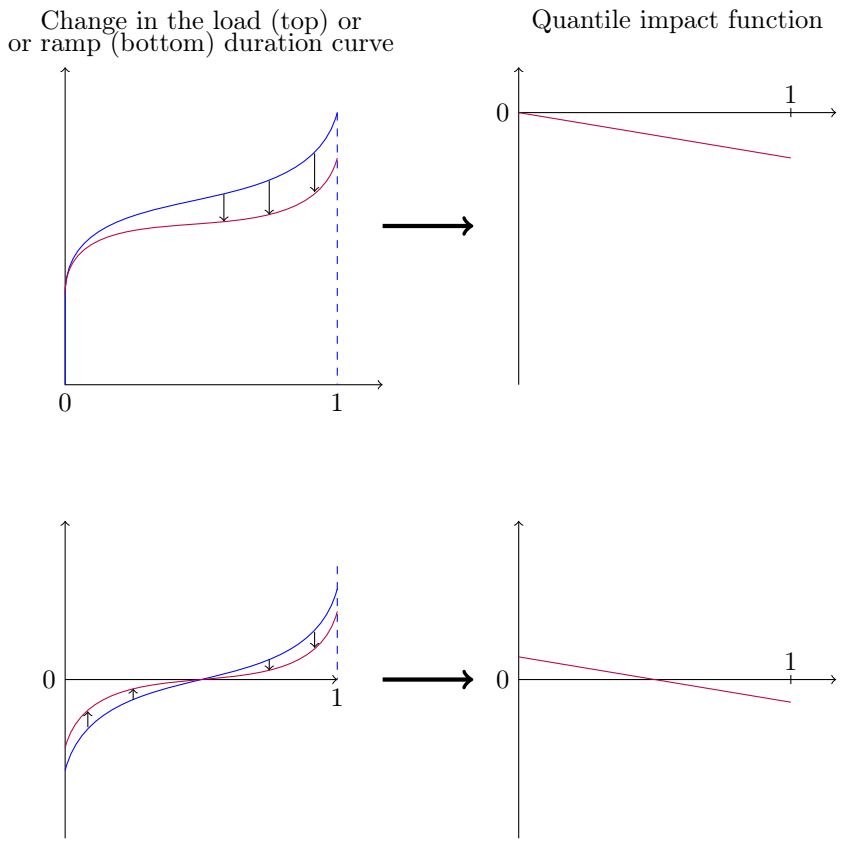
5.1 Quantile impact functions

In order to assess whether or not different distributed generation technologies are likely to significantly defer network expansions, we look at how each technology has shifted, on average, the load and ramp duration curves supplied by distribution substations. In other words, we estimate, for each distributed generation technology, the average impact that adding 1 MW of capacity has had on both the load and ramp duration curves supplied by substations.

We characterize the changes in the load/ramp duration curve induced by a given technology through a quantile impact function. Figure 10 illustrates the intuition behind such a function. For a given technology, the quantile impact function maps each quantile index (from 0 to 1) to the average effect (in MWh/MW) that adding 1 MW of this technology has had on the corresponding quantile of the load/ramp duration curve. A downward sloping quantile impact function indicates that the corresponding technology tends to shrink the distribution of interest, which is a desirable feature for both duration curves. For the ramp duration curve, it indeed suggests that the magnitude of the highest ramps (in absolute value) decreases with the addition of distributed generation of the corresponding technology. For the load duration curve, it indicates that the output from the technology of interest tends to be higher during substation peak hours than during off-peak hours. In addition, technologies that have a higher impact on the top quantiles of the load duration curve are more likely to drive network capacity down (see Figure 6). Conversely, as more and more substations experience reverse power flows (see Figure 9), technologies that have a milder impact on the bottom quantiles, and thus that contribute less to amplify the magnitude of reverse power flows, are less likely to require network expansions.

a number of small hydro units are located on the Northern part of the Rhine river, which is one of the few areas where we lack information on substation load (see Northeastern region on Figure 8).

Figure 10: Illustration of the intuition behind quantile impact function for the load duration curve (top) and the ramp duration curve (bottom)



5.2 Estimation

We use a seemingly unrelated regressions framework with a two-way fixed effect model to estimate quantile impact functions. In other words, for both the load and ramp duration curves, we run the following regression for the main percentiles $q \in \{0.01, 0.1, 0.25, 0.5, 0.75, 0.9, 0.99\}$:

$$Y_{q,s,y} = \sum_t \beta_{q,t} K_{t,s,y} + \delta_s + \delta_y + \epsilon_{s,y} \tag{1}$$

where $Y_{q,s,y}$ denotes the q-th quantile of either the load or ramp duration curve for substation s in year y, $K_{t,s,y}$ the installed capacity of distributed generation technology t connected to substation

s as of year y, and δ_s and δ_y are respectively substation and year fixed effects. We thus estimate fourteen linear regressions (7 percentiles for the load duration curve and 7 percentiles for the ramp duration curve) using ordinary least squares.

Fixing a given technology t and a given duration curve of interest (either the load or the ramp duration curve), the 7-tuple $(\hat{\beta}_{0.01,t}, \hat{\beta}_{0.1,t}, \hat{\beta}_{0.25,t}, \hat{\beta}_{0,5,t}, \hat{\beta}_{0.75,t}, \hat{\beta}_{0.9,t}, \hat{\beta}_{0.99,t})$ then corresponds to the estimated quantile impact function for that technology and duration curve. Indeed, the coefficient $\hat{\beta}_{q,t}$ captures the average impact (in MWh or MW) that adding 1 MW of technology t is having on the q-th quantile of the duration curve of interest.

Because distributed generation output is non-negative and capped by installed generation capacity, we expect that $\hat{\beta}_{q,t} \in [-1, 0]$ for the load duration curve. In other words, $-\hat{\beta}_{q,t}$ captures the fraction of the nameplate capacity from technology t that is on average generated during hours that correspond the q-th quantile of the distribution of hourly net load levels. For example, $\hat{\beta}_{0.5,PV} = -0.2$ means that adding 1 MW of distributed PV generation decreases on average the median ($q = 0.5$) hourly net load level supplied by distribution substations by 0.2 MWh.

For the ramp duration curve, we expect that $\hat{\beta}_{q,t} \in [-1, 1]$. Indeed, the additional variation in hourly net load levels between two subsequent hours that may be attributable to a distributed generation unit can at most be its installed capacity. Coefficients have otherwise a similar interpretation as for the load duration curve. For example, $\hat{\beta}_{0.75,Wind} = 0.1$ means that adding 1 MW of distributed wind generation increases on average the third quartile of the distribution of hourly ramps supplied by distribution substations by 0.1 MW. Importantly, because ramps are signed, the lowest hourly ramps are negative. As a result, negative coefficients for the lowest quantiles actually tend to *increase* the magnitude of ramps in absolute value.

6 Main results

6.1 Impact of distributed generation on substation load duration curve

We first estimate the impact of the different distributed generation technologies on the load duration curve supplied by substations. As discussed in Section 5, the shape of the obtained quantile impact function for each technology provides relevant information on whether or not this technology may contribute to defer or avoid network expansions. Table 3 shows the obtained results, which are represented graphically on Figure 11. We report robust (HC1) standard errors clustered at the substation level.

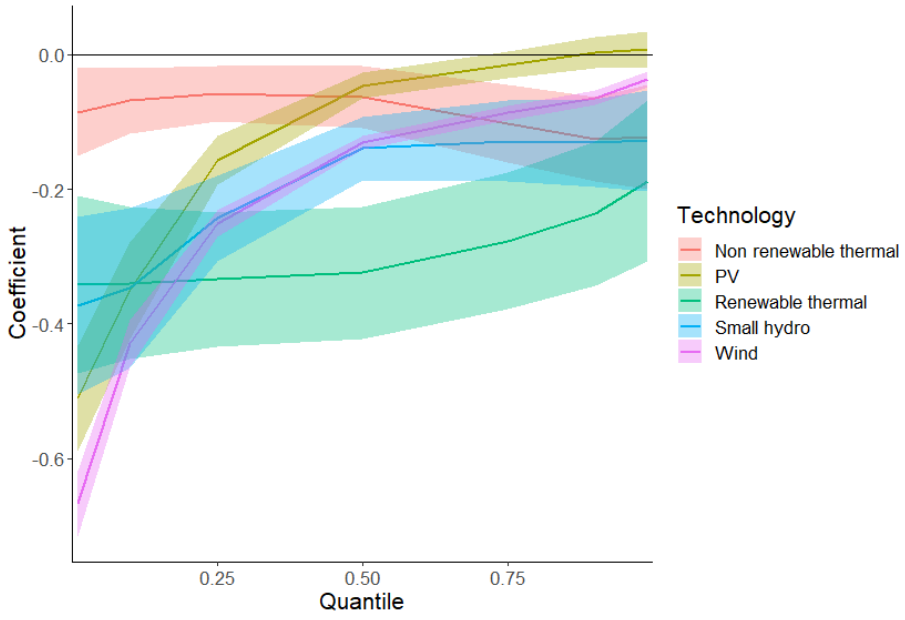
Table 3: Estimated coefficients when regressing the main quantiles of the distribution of hourly net load levels (in a given year for a given substation) on the installed capacities of the different technologies. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	-0.667^{***}	-0.429^{***}	-0.251^{***}	-0.130^{***}	-0.087^{***}	-0.064^{***}	-0.037^{***}
	(0.025)	(0.018)	(0.010)	(0.005)	(0.005)	(0.006)	(0.006)
PV	-0.510^{***}	-0.351^{***}	-0.157^{***}	-0.046^{***}	-0.016	0.003	0.008
	(0.040)	(0.036)	(0.019)	(0.009)	(0.010)	(0.012)	(0.014)
Small hydro	-0.373^{***}	-0.346^{***}	-0.243^{***}	-0.139^{***}	-0.128^{***}	-0.131^{***}	-0.128^{***}
	(0.068)	(0.060)	(0.032)	(0.024)	(0.031)	(0.033)	(0.038)
Renewable thermal	-0.341^{***}	-0.339^{***}	-0.334^{***}	-0.324^{***}	-0.277^{***}	-0.235^{***}	-0.187^{***}
	(0.067)	(0.058)	(0.051)	(0.050)	(0.052)	(0.055)	(0.061)
Non renewable thermal	-0.085^{**}	-0.069^{***}	-0.058^{***}	-0.063^{***}	-0.103^{***}	-0.126^{***}	-0.123^{***}
	(0.033)	(0.025)	(0.021)	(0.023)	(0.029)	(0.032)	(0.040)
Observations	30,091	30,091	30,091	30,091	30,091	30,091	30,091
R^2	0.953	0.958	0.975	0.983	0.983	0.985	0.984
Adjusted R^2	0.949	0.955	0.973	0.981	0.982	0.983	0.983
Note:							*p<0.1; **p<0.05; ***p<0.01

Quite remarkably, we find very contrasted quantile impact functions for the different distributed generation technologies. Two characteristics of these functions are of particular interest: their value for the extreme quantiles and their slope.

First, the higher the coefficients for the highest quantiles (in absolute value), the more a given

Figure 11: Graphical representation of the quantile impact functions for the distribution of hourly net loads. Thick lines correspond to the point estimates. Sleeves delimit (two-sided) 5% confidence intervals from robust standard errors clustered at the substation level.



technology is associated with a decrease in the peak net load supplied by substations, and thus the more likely it is to defer or avoid grid expansions. In France, country-wide annual peak load – and thus annual substation peak load for a large number of distribution grids – is reached during cold winter evenings. Consistently, PV is found to have no significant impact on the highest quantiles of the distribution of hourly net load levels. More surprisingly, the impact of wind on the highest quantiles is also very small, despite the fact that wind generation is typically higher during the winter. Larger installed capacities relative to local load may explain this unexpected result. Indeed, wind farms tend to be relatively large installations, with a median size of 10 MW. Their generation output is thus often significant relative to local consumption, so that local peak net load hours may mechanically shift away from hours with high wind generation. The other technologies are found to have a more sizable impact on local peak net load. For non-renewable thermal units, this finding appears consistent with the fact that public subsidies provide incentives for small natural gas co-

generation units to produce during the winter.[20] Symmetrically, all technologies but non-renewable thermal are found to be associated with a significant downward shift in the lowest quantiles of the distribution of hourly net load levels. This is particularly striking for wind and PV, whose quantile impact functions are very concave. Because both the frequency of occurrence and the magnitude of reverse power flows are increasing (see Figure 9), these large negative impacts on the lowest quantiles of the distribution of hourly net load levels will ultimately impose new constraints on distribution grids.

Second, whether the quantile impact function is upward or downward sloping is also of particular interest. Indeed, a monotone decreasing quantile impact function implies a higher downward shift for the top quantiles than for the bottom quantiles. As a result, it translates into a narrower distribution of hourly net load levels, which increases the utilization rates of grid assets. By contrast, a monotone increasing quantile impact function tends to "stretch" this distribution, expanding the range of hourly net load levels that must be supplied by the substation. In addition, when substations experience reverse power flows, a further decrease in the bottom quantiles increases the local peak injections to the transmission grid. Quite strikingly, quantile impact functions are found to be monotone increasing for all technologies but non-renewable thermal.

We test statistically the main characteristics of the quantile impact functions using the testing framework for seemingly unrelated regressions developed in Wolak (1987, 1989). Appendix D describes in more details the statistical tests performed. Consistently with the graphical intuition from Figure 11, we can reject at the 1% level that the quantile impact function ends at zero for all technologies but PV. In addition, for all technologies but non renewable thermal, we cannot reject (even at the 10% level) that the quantile impact function is increasing. By contrast, this null hypothesis is rejected at the 1% level for non renewable thermal, while the null hypothesis of a decreasing quantile impact function cannot be rejected at the same level of statistical significance.

[20] https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033385467/

6.2 Impact of distributed generation on substation ramp duration curve

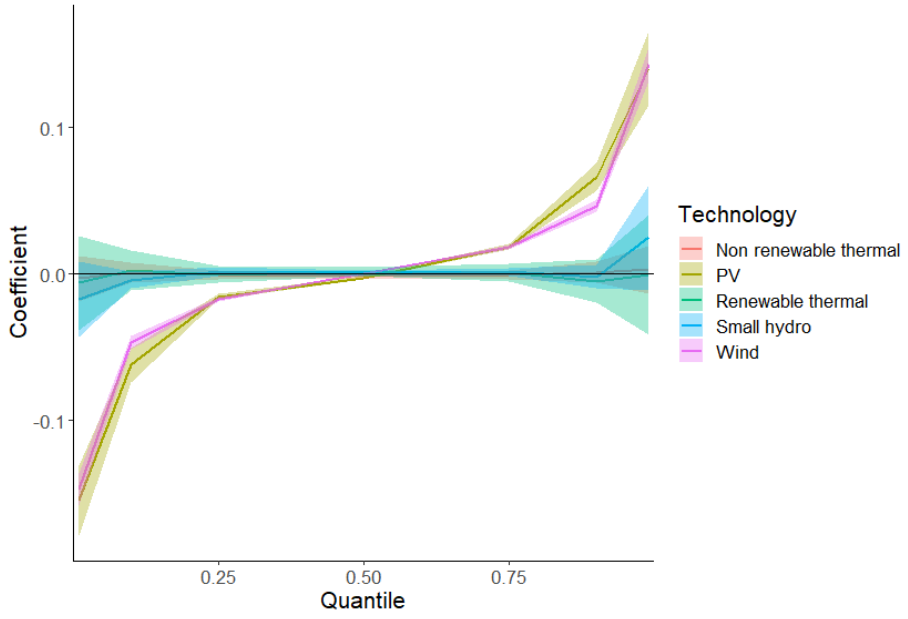
Next, we estimate the impact of the different distributed generation technologies on the ramp duration curve faced by substations. We define hourly ramps as the difference between two consecutive hourly net load levels. Table 4 shows the obtained results, which are presented graphically on Figure 12. We report robust (HC1) standard errors clustered at the substation level.

Table 4: Estimated coefficients when regressing the main quantiles of the distributions of hourly ramps (in a given year at a given substation) on the installed capacities of the different technologies. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	-0.147^{***}	-0.047^{***}	-0.017^{***}	0.0001^{*}	0.018^{***}	0.047^{***}	0.144^{***}
	(0.006)	(0.002)	(0.001)	(0.0001)	(0.001)	(0.002)	(0.006)
PV	-0.155^{***}	-0.062^{***}	-0.016^{***}	-0.003^{***}	0.019^{***}	0.066^{***}	0.141^{***}
	(0.012)	(0.006)	(0.001)	(0.0004)	(0.001)	(0.005)	(0.013)
Small hydro	−0.017	−0.004	0.001	0.001	0.001	−0.002	0.025
	(0.013)	(0.003)	(0.002)	(0.001)	(0.001)	(0.004)	(0.018)
Renewable thermal	−0.006	0.003	−0.0005	0.001	0.001	−0.005	−0.0005
	(0.016)	(0.007)	(0.003)	(0.002)	(0.003)	(0.008)	(0.021)
Non renewable thermal	−0.003	0.001	−0.001	−0.001	−0.001	0.001	0.003
	(0.008)	(0.003)	(0.002)	(0.001)	(0.002)	(0.004)	(0.008)
Observations	30,091	30,091	30,091	30,091	30,091	30,091	30,091
R^2	0.951	0.963	0.966	0.836	0.966	0.960	0.950
Adjusted R^2	0.947	0.961	0.963	0.823	0.963	0.956	0.947
Note:							*p<0.1; **p<0.05; ***p<0.01

Somewhat strikingly, two distinct groups of distributed generation technologies emerge. On the one hand, thermal and small hydro units are found to have negligible impacts on the ramp duration curve. On the other hand, wind and PV units tend to significantly stretch the distribution of hourly ramps. In other words, increased installed capacities of wind and PV are associated with a significant increase in the absolute value of the most extreme local ramps, both negative and positive. Indeed, we find that a 1 MW increase in installed capacity of either wind or PV is on average associated with an increase of 0.14-0.15 MW in both the 1st and 99th quantiles of the distribution of hourly ramps. Because the total capacities of connected distributed generation may

Figure 12: Graphical representation of the quantile impact functions for the distribution of hourly ramps. Thick lines correspond to the point estimates. Sleeves delimit (two-sided) 5% confidence intervals from robust standard errors clustered at the substation level.



exceed 100 MW (see Table 2), this order of magnitude suggests that very large local hourly ramps could emerge as a result of high installed capacities of distributed wind and PV generation.

7 Discussion

Because our dataset covers the vast majority of both distribution substations (Figure 8) and distributed generation capacities (Table 2) in mainland France, our results represent a very credible assessment for this country of the average impact that different distributed generation technologies have had on both the load and ramp duration curves supplied by distribution substations. Whether these results hold in other countries as well is, at least to some extent, an open question. In particular, France reaches its annual peak load in the winter due to its high reliance on electric heating. By contrast, a number of countries and US States experience their peak load in the summer, due to

a large demand for cooling. One may thus wonder whether our results are valid in such electricity systems as well.

First, in a power system whose demand peaks during the summer, distributed PV may have a higher impact on the top quantiles of the distribution of hourly net load levels. This impact may however be much smaller than one might expect, especially once significant amounts of distribution generation capacities are installed. Indeed, for the case of France, we find that wind has a small impact on the top quantiles of the distribution of hourly net load levels despite the fact that the output from wind generation is significantly higher during the winter (RTE, 2018). This observation suggests that the small or negligible impact of wind and PV on substation peak load levels may be, at least to some extent, driven by the fact that these technologies have reached large amounts of installed capacities. Indeed, when installed capacities start becoming large relative to local peak demand, peak net load is mechanically reached when distributed generation output is small. As a result, even in electricity systems where peak load is driven by cooling demand, the impact of PV generation on the top quantiles of the distribution of hourly net load levels seem likely to decrease with increased penetration of distributed PV, as peak net load hours will progressively shift later during the day.

Second, the beneficial impact of thermal units may not be as clear cut in electricity systems whose peak demand is driven by a large cooling load. Indeed, although back-up generation units may still synchronize to some extent with local peak demand, co-generation units are less likely to do so in such systems. By contrast, the fact that the lowest quantiles of the distribution of hourly net load levels are significantly shifted downwards by wind and PV seems likely to hold in other regions as well as soon as installed distributed generation capacities become large enough to be one of the main reasons for experiencing hours with very low net load levels. For example, the California Independent System Operator is already curtailing significant amounts of PV generation

despite operating a summer peak electricity system.[21]

Third, our finding that distributed wind and PV generation tends to stretch significantly the distribution of hourly ramps seems most likely to be a consequence of the inherent volatility of their electricity output. Indeed, although small hydro units such as run-of-the-river plants are also intermittent, their output is significantly less volatile. Similarly, even when not actively dispatched thermal units are likely to have a steadier generation output. The dichotomy between wind and PV on the one hand, and other technologies on the other hand, thus seems likely to hold in other electricity systems as well.

Importantly, even though we find that distributed wind and PV units are inducing the most adverse changes in the use patterns of the electricity grid, our results should not be interpreted as conclusive evidence that adding distributed wind and PV to the electricity generation mix is an undesirable policy. Indeed, these technologies have other characteristics such as low costs, low carbon emissions or domestic resource availability that can motivate a significant increase in their installed capacities. However, our findings do stress the importance of explicitly taking into account the impact of distributed wind and PV on distribution networks. In the absence of dedicated strategies to mitigate their adverse impacts, it seems unlikely that very high installed capacities of distributed wind and PV generation will be reached in a cost-effective manner.

A corollary is that identifying the most effective strategies to accommodate intermittent distributed generation is of critical importance. Historically, possibilities were largely limited to demand response, excess generation curtailments and network upgrades. Thanks to tremendous technology improvements, battery storage now represents another option, which may act as a complement or a substitute to other alternatives. The extent to which battery storage can help accommodate the adverse changes in grid use patterns induced by distributed wind and PV generation

[21] See http://www.caiso.com/informed/Pages/ManagingOversupply.aspx

will thus be investigated in follow-up work.

8 Conclusion

This paper provides empirical evidence on whether distributed generation is likely or not to defer or avoid network expansions, using data that is rich enough to derive technology-specific conclusions. We study the case of France, where over 28 GW of distributed generation units were connected to the electricity grid as of 2018, which represents a quarter of the maximum consumption ever recorded in the country. These installations belong to five broadly-defined technologies: wind, PV, small hydro, renewable thermal and non renewable thermal. Our analysis combines two main datasets. First, we observe the hourly net load levels at over 2,000 distribution substations from 2005 to 2018. Together, these substations supply the vast majority of mainland France electricity demand from low and medium voltage consumers. Second, we observe detailed information on the universe of distributed generation units. In particular, we know exactly to which substation 75% of PV capacities and over 90% of capacities from other technologies connect. We use a detailed assignment procedure that relies on a division of mainland France into over 45,000 spatial units to match the remaining distributed generation capacities to substations.

We use a seemingly unrelated regressions framework with a two-way fixed effect specification to assess how an increase in the installed capacities of different distributed generation technologies have changed on average the shape of both the load duration curve and the ramp duration curve supplied by substations. Both these duration curves indeed provide very relevant information on respectively needed network capacity and the operating constraints faced by system operators.

We find that wind and PV units are associated with the most adverse changes in both the load and ramp duration curves faced by distribution substations. For the load duration curve, they

indeed have had a small or negligible impact on the highest quantiles of the distribution of net load levels, but have induced a large downward shift in the lowest quantiles. For the ramp duration curve, wind and PV tend to stretch significantly the distribution of hourly ramps, making the largest ramps (in absolute value) more extreme. By contrast, other technologies are found to have had positive (for non-renewable thermal) or less adverse (for small hydro and renewable thermal) impacts on the load duration curve, and to have induced no significant change in the ramp duration curve faced by substations.

Overall, our results suggest that dedicated strategies at the distribution grid level, such as demand response, distributed generation curtailments or battery storage, will likely be needed to mitigate the changes induced by very high levels of wind and PV distributed generation. This question will be explored in follow-up work.

References

Bistline, John E, "Economic and technical challenges of flexible operations under large-scale variable renewable deployment," *Energy Economics*, 2017, *64*, 363–372.

Borenstein, Severin, "What Can Distributed Generation Do For the Grid?," *Energy Institute Blog, UC Berkeley*, 2020. `https://energyathaas.wordpress.com/2020/09/28/what-can-distributed-generation-do-for-the-grid/`.

Burger, Scott P, Jesse D Jenkins, Carlos Batlle, and Ignacio J Perez-Arriaga, "Restructuring revisited part 2: coordination in electricity distribution systems," *The Energy Journal*, 2019, *40* (3).

Callaway, Duncan S, Meredith Fowlie, and Gavin McCormick, "Location, location, location: The variable value of renewable energy and demand-side efficiency resources," *Journal of the Association of Environmental and Resource Economists*, 2018, *5* (1), 39–75.

Cohen, MA and DS Callaway, "Effects of distributed PV generation on California's distribution system, Part 1: Engineering simulations," *Solar Energy*, 2016, *128*, 126–138.

_ , PA Kauzmann, and DS Callaway, "Effects of distributed PV generation on California's distribution system, part 2: Economic analysis," *Solar Energy*, 2016, *128*, 139–152.

Commission de Régulation de l'Energie, "Délibération de la CRE du 17 décembre 2020 portant projet de décision sur le tarif d'utilisation des réseaux publics de distribution d'électricité (TURPE 6 HTA-BT)," 2020.

Davis, Lucas, "Does Rooftop Solar Help the Distribution System?," *Energy Institute Blog, UC Berkeley*, 2018. https://energyathaas.wordpress.com/2018/06/25/does-rooftop-solar-help-the-distribution-system/.

Gillingham, Kenneth and Marten Ovaere, "The Heterogeneous Value of Solar and Wind Energy: Empirical Evidence from the United States and Europe," 2020.

Muro, Mark and Devashree Saha, "Rooftop solar: Net metering is a net benefit," 2016. https://www.brookings.edu/research/rooftop-solar-net-metering-is-a-net-benefit/.

Navidi, Thomas, Abbas El Gamal, and Ram Rajagopal, "Co-optimizing Energy, Grid Services and Voltage Support in Networks with Distributed Storage," in "2019 IEEE Power & Energy Society General Meeting (PESGM)" IEEE 2019, pp. 1–5.

Persoz, Henri, Gérard Santucci, Jean-Claude Lemoine, and Paul Sapet, *La planification des réseaux électriques*, Eyrolles, 1984.

Phadke, Amol A., Nikit Abhyankar, Ranjit Deshmukh, Julia Szinai, and Anand R. Gopal, "Cost-effective decarbonization of California's power sector by 2030 with the aid of battery storage," 2020.

RTE, "Electricity report 2018," Technical Report 2018.

Sepulveda, Nestor A, Jesse D Jenkins, Fernando J de Sisternes, and Richard K Lester, "The role of firm low-carbon electricity resources in deep decarbonization of power generation," *Joule*, 2018, *2* (11), 2403–2420.

US Department of Energy, "United States Electricity Industry Primer," Technical Report 2015.

Wolak, Frank A, "An exact test for multiple inequality and equality constraints in the linear regression model," *Journal of the American Statistical Association*, 1987, *82* (399), 782–793.

_ , "Testing inequality constraints in linear econometric models," *Journal of Econometrics*, 1989, *41*, 205–235.

Appendices

A Assignment procedure for distributed generation units whose substation is unknown

This Appendix provides further information on the assignment procedure we implement to infer to which substations(s) distributed generation units with unknown upstream substation are most likely to connect. This procedure largely relies on detailed spatial information. Mainland France is indeed divided into over 30,000 administrative counties.[22] In addition, densely populated counties are further broken down into sub-counties for census purposes. Sub-counties are called "IRIS mesh" and are defined for census purposes in order to split the most highly populated counties (all counties with more than 10,000 inhabitants and most counties with more than 5,000 inhabitants) into smaller geographical units. As of 2019, 1,840 counties in mainland France (about 5%) are further divided into sub-counties. Because the location of distributed generation units is observed down to the county or sub-county level, we divide France into spatial units that correspond to either counties or sub-counties.[23] More precisely, we divide a given county into its sub-counties whenever (i) this decomposition is available; and (ii) the location of at least one distributed generation observation sitting in this county is known down to the sub-county level. We end up using a sub-division of mainland France into 45,508 spatial units, with a mean surface of 11.9 km^2 (4.6 miles squared).

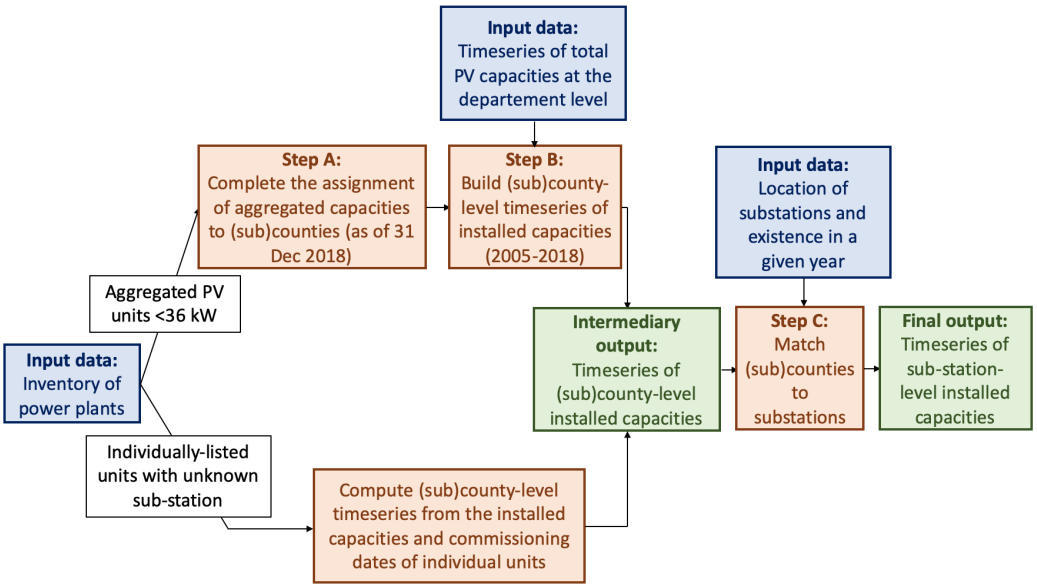
The flow chart of Figure A.1 summarizes the different steps of the procedure. Because we observe the capacity, commissioning date, and (sub)county of the distributed generation units that are listed individually, computing timeseries of installed capacities at the (sub)county level is straightforward for these units.[24] By contrast, observations consisting in aggregated PV units raise

[22] The exact number of counties changes over time due to mergers and boundary updates. We use the definition of administrative boundaries as of 1 January 2019.

[23] The corresponding spatial boundaries were downloaded from: https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html#contoursiris

[24] For 71 individually-listed installations with unknown upstream substation, we only observe the county of location

Figure A.1: Flow chart of the assignment procedure we implement for distributed generation units for which upstream substation information is missing



two challenges,[25] one on the spatial dimension and the other on the temporal dimension. First, the location of a quarter of aggregated PV capacities is only known with a coarser spatial granularity than (sub)counties. This difficulty is dealt with in Step A. Second, we observe capacities as of 31 December 2018. However, by contrast to individually-listed units for which installed capacity most likely remains constant from their commissioning date onward, the composition of aggregated units – and thus their installed capacity – has evolved over time. We address this issue in Step B.

Step A: completing the assignment of aggregated PV units to (sub)counties

In order to respect the privacy of individual owners, most small (<36kW) PV units are aggregated at the finest level of spatial aggregation that makes it possible to group at least 10 installations together. From coarsest to finest, these levels of spatial aggregation are: departement, county and

although we further divided the county into sub-counties. These installations are assumed to be equally likely to be located in each corresponding sub-county.

[25] Due to their very small installed capacities, we neglect aggregated observations for technologies other than PV.

sub-county.[26] Aggregated PV observations are built as follows. First, any sub-county than has more than 10 installations is listed as an observation, whose capacity is the total capacity of these units (as of 31 December 2018). Second, any county than has more than 10 installations not included in one of the sub-county aggregates is then listed as an observation, whose capacity is the sum of the capacity of these units. Finally, remaining PV units that must be aggregated are grouped at the departement level. Mainland France has 94 such departements.

As a result of this aggregation procedure, the majority (74%) of aggregated PV capacities are located down to the county or sub-county level. Most of these observations thus map directly to our spatial division of mainland France. A minority of observations are county-level aggregates located in a county we further divided into sub-counties. The installed capacities of these observations are deemed equally likely to be installed in the pool of sub-counties where they may be located (i.e. the sub-counties within that county with no aggregated PV observation listed in the inventory of power plants). For the remaining 26% of aggregated PV capacities, we only observe in which departement the corresponding units are located. Given the aggregation rule used to build the inventory of power plants, we further know that these units may only be located in (sub)counties where none of the other 74% capacities are located. For simplicity, we thus split the capacity aggregated at the departement level uniformly across the pool of candidate counties where the corresponding units may be located.[27] When a county is itself divided into sub-counties, we subsequently split the capacity that got allocated to it uniformly across its sub-counties with no aggregated observation. Other approaches to allocate to (sub)counties the observations aggregated at the departement level would require additional modeling and/or information sources. They are however very unlikely

[26] Because of idiosyncrasies such as mistakes when entering the fuel type of an installation, a handful of observations are aggregated at the regional level, which is a coarser spatial unit than departements. These observations however add up to less than 1MW and are thus neglected.

[27] For the 5% of counties that are further divided in sub-counties, two situations may arise. First, all sub-counties may each have more than 10 installations, or a total of more than 10 installations may exist in sub-counties that have less than 10 installations each. Second, less than 10 installations in total may exist in sub-counties that have less than 10 installations each. Some installations aggregated at the departement level may then be located in the latter counties, but not in the former counties.

to affect dramatically the obtained outcome. For example, allocating capacities aggregated at the departement level to the remaining (sub)counties using log-population instead of uniform weight yields very similar results. The correlation between the (sub)county-level capacities obtained using uniform vs log-population weights is 0.999 (0.97 when focusing on the subset of (sub)counties where no aggregated observation is directly observed in the inventory of power plants). As a result, we use a uniform allocation for the sake of simplicity.

Step B: building (sub)county-level timeseries of aggregated PV installed capacities

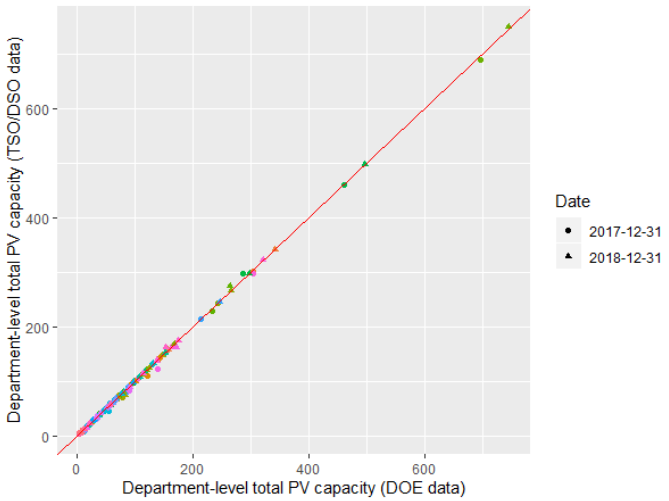
The output of Step A is a cross-section of installed capacities $K_{c,d,2018}$ from aggregated PV units in (sub)county c of departement d as of 31 December 2018. However, because aggregated observations are not individual installations, their composition and thus their installed capacity has changed over time. In order to infer how (sub)county-level installed capacities are likely to have evolved between 2005 and 2018, we proceed in two steps.

First, we use a third dataset from the French Department of Energy (DOE) that provides panel data at the departement level of total installed PV capacities between 2006 and 2018 (installed capacities being virtually zero in 2005).[28] These capacities include all PV installations, from small residential units to large-scale farms connected to the transmission grid. As shown in Figure A.2, this third dataset appears to be consistent with the information available in the public inventories of power plants. Because we observe the location, installed capacity and commissioning date of installations that are listed individually in the public inventory of power plants, we can compute departement-level timeseries of PV capacity from individually-listed units. Subtracting these timeseries to the timeseries of departement-level total PV capacity from the DOE dataset yields departement-level timeseries of PV capacity from aggregated units. Figures A.3 and A.4 show the

[28]This information is published quarterly by the Service des données et études statistiques (e.g. https://www.statistiques.developpement-durable.gouv.fr/tableau-de-bord-solaire-photovoltaique-quatrieme-trimestre-2018 for the fourth quarter of 2018). We are grateful to the Department of Energy for having shared the corresponding historical data (updated of as of July 2020).

obtained results, and compare them to the capacities observed in the public inventories for 2017 and 2018. Overall, both sources of information agree very well. In the very few cases where some discrepancies are observed, we use the maximum of both metrics, since it generally appears to be more consistent with the rest of the timeseries. We further impose monotonicity which is (mildly) violated on three occasions.

Figure A.2: Departement-level installed PV capacities (in MW) as of 31 December 2017 and 2018 (i) in the DOE dataset (x-axis) and (ii) in the public inventories of power plants 2017 and 2018 (y-axis)



Second, for each year and each departement, we need to dispatch the departement-level capacity $K_{d,y}$ from aggregated units to the different (sub)counties. In other words, we want to define capacities $K_{c,d,y}$ for each year y and (sub)county c (located in departement d) such that:

$$\forall y, \forall d, \sum_{c \in d} K_{c,d,y} = K_{d,y} \tag{2}$$

To do so, we leverage the cross-section $\{W_{c,d}\}_c$ computed in Step A, where $W_{c,d}$ is the obtained capacity from aggregated units in (sub)county c of departement d. We implement four different methodologies to build $K_{c,d,y}$:

Figure A.3: Inferred departement-level timeseries of PV capacity from aggregated units (first 48 departements). Green dots represent actual capacities as reported in the public inventories of power plants 2017 and 2018.

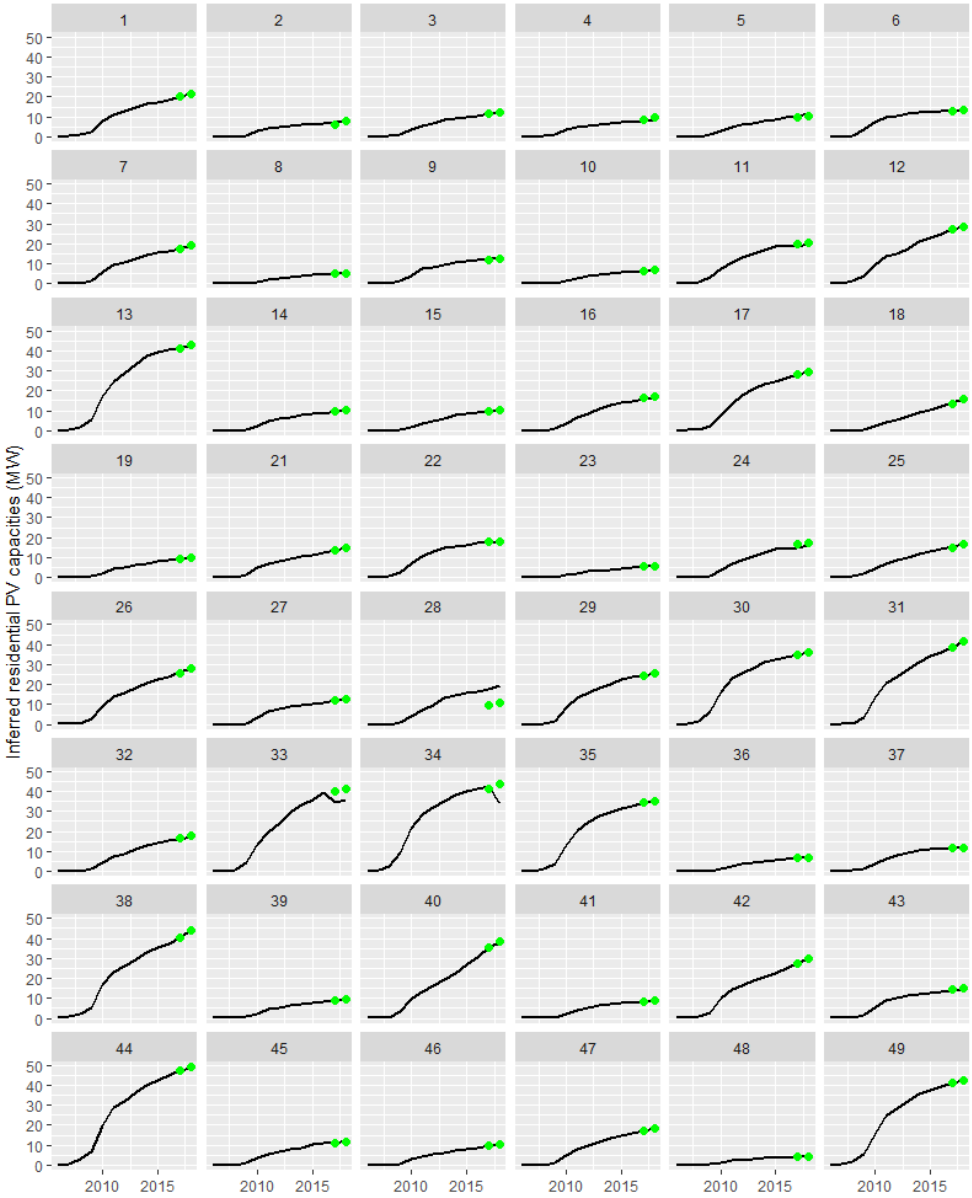
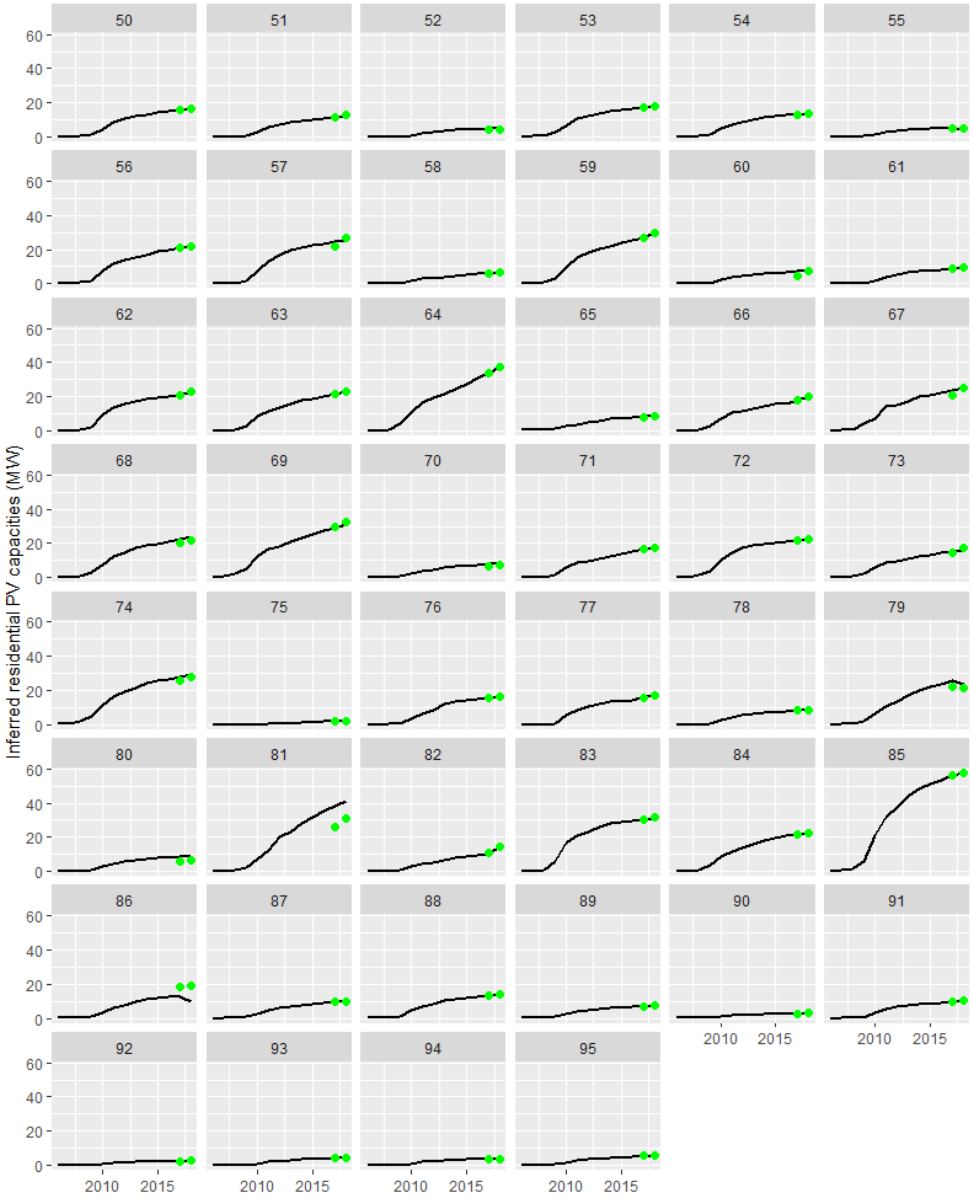


Figure A.4: Inferred departement-level timeseries of PV capacity from aggregated units (remaining 46 departements). Green dots represent actual capacities as reported in the public inventories of power plants 2017 and 2018.



1. **Homothetic static approach:** this method considers that the probability to see a given amount of installed capacities in a given county is proportional to the installed capacity $W_{c,d}$ in this county as of 2018. In other words, we postulate that the installed capacity in (sub)county c of departement d as of year y was:

$$K^{HS}_{c,d,y} \equiv \frac{W_{c,d}}{\sum_{c' \in d} W_{c',d}} K_{d,y} \tag{3}$$

2. **Sequential static approach:** this method considers that new PV units get installed first in the counties with the highest remaining capacity to be installed. In other words, knowing that $W_{c,d}$ must be installed by 2018,[29] we compute $K^*_{d,y}$ such that:

$$\sum_{c \in d} \max(W_{c,d} - K^*_{d,y}, 0) \equiv K_{d,y} \tag{4}$$

We then postulate that the installed capacity in county c of departement d as of year y was:

$$K^{SS}_{c,d,y} \equiv \max(W_{c,d} - K^*_{d,y}, 0) \tag{5}$$

3. **Homothetic dynamic approach:** although their exact meaning is somewhat ambiguous, the public inventory of power plants does provide "commissioning dates" for aggregated PV observations. We interpret such dates as the date at which the 10th unit got installed in the corresponding spatial unit, and further assume that the 1st unit was installed shortly before that. As a result, these commissioning dates put additional restrictions on the set of (sub)counties that may have host aggregated PV units as of a given year. We denote $\mathcal{C}_{d,y}$ the set of (sub)counties in departement d whose "commissioning date" is anterior to 31 December of year y.[30] We then define the installed capacity in county c of departement d as of year y

[29] More precisely, $W_{c,d}$ is normalized within each department in order to sum to $K_{d,2018}$ and thus be more consistent with the rest of the timeseries. As shown on Figures A.3 and A.4, there is barely any difference between $K_{d,2018}$ and $\sum_{c \in d} W_{c,d}$ for virtually all departements.

[30] (Sub)counties for which no aggregated observation exist in the inventory are assumed to belong to $\mathcal{C}_{d,y}$ for all y.

using an iterative approach:

$$K^{HD}_{c,d,y} \equiv \mathbf{1}_{c \in \mathcal{C}_{d,y}} \left(K^{HD}_{c,d,y-1} + \frac{W_{c,d} - K^{HD}_{c,d,y-1}}{\sum_{c' \in \mathcal{C}_{d,y}} W_{c',d} - K^{HD}_{c',d,y-1}} \left(K_{d,y} - K_{d,y-1} \right) \right) \tag{6}$$

with $K^{HD}_{c,d,2005} = 0$ for all c, d.

4. **Sequential dynamic approach:** as for the homothetic method, we also implement a dynamic version of the sequential approach. Formally, we first compute $K^*_{d,y}$ such that:

$$\sum_{c \in \mathcal{C}_{d,y}} \max(W_{c,d} - K^{SD}_{c,d,y-1} - K^*_{d,y}, 0) \equiv K_{d,y} - K_{d,y-1} \tag{7}$$

with $K^{SD}_{c,d,2005} = 0$ for all c, d. We then postulate that the installed capacity in county c of departement d as of year y was:

$$K^{SD}_{c,d,y} \equiv \mathbf{1}_{c \in \mathcal{C}_{d,y}} \left(K^{SD}_{c,d,y-1} + \max(W_{c,d} - K^{SD}_{c,d,y-1} - K^*_{d,y}, 0) \right) \tag{8}$$

The outcome of steps A and B is four alternative (sub)county-level timeseries of installed capacities from aggregated PV units that are consistent with their observed evolution at the departement level. The results reported in the main text are derived using the homothetic static approach.

The last step of the assignment procedure consists in mapping (sub)counties to upstream distribution substations. This step also applies to distributed generation units that are individually listed in the inventory of power plants but for which the upstream substation is unknown.

Step C: matching (sub)counties to substations and deriving substation level timeseries of installed capacities

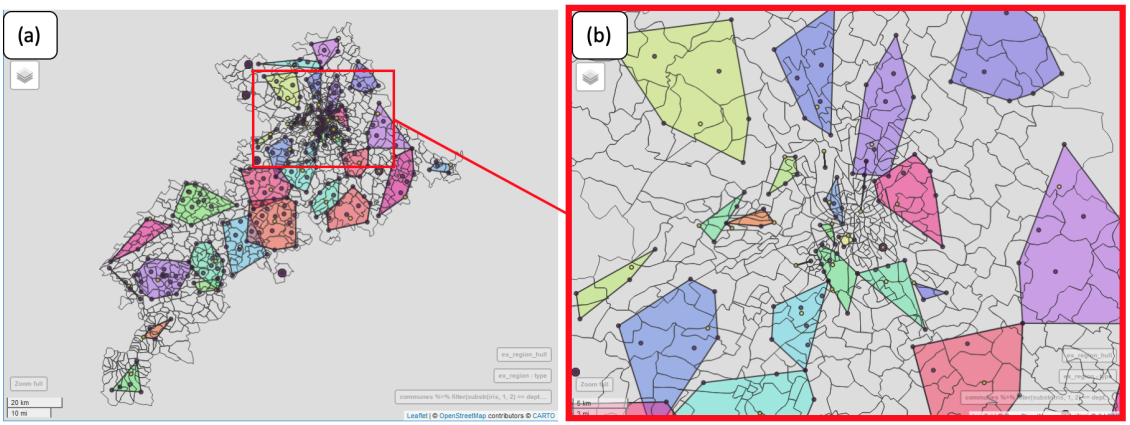
In order to match (sub)counties to substations, we rely on two sources of information. First, we know the GPS coordinates of the substations.[31] Second, we know the (sub)county where a

[31] This information is for example available from: https://www.data.gouv.fr/en/datasets/postes-electriques-rte-au-6-juin-2020-1/ (last accessed on 31 August 2020).

large number of individually-listed distributed generation units with observed upstream substation are located. We use the public inventory as of 31 December 2019 (restricting attention to the substations that are known to exist as of 31 December 2018) in order to maximize the number of observed (sub)county – substation pairs. We observe 14,000+ such pairs, as well as the location of 2,000+ substations.

For each substation, we first compute the convex hull of both its location and the centroids of the (sub)counties where one or several distributed generation units that are known to connect to this substation are located. When building the convex hulls, we exclude (sub)counties whose centroid is located more than 40km away from the connected substation to filter potential mistakes in the public inventory (this procedure screens out 138 (sub)county – substation pairs). Panel (a) on Figure A.5 shows the outcome of this procedure for one of the 94 departements. Panel (b) further zooms on a densely populated area where we further divided counties into sub-counties. Even in urban areas, the spatial units we use appears to be granular enough relative to the spatial density of distribution substations.

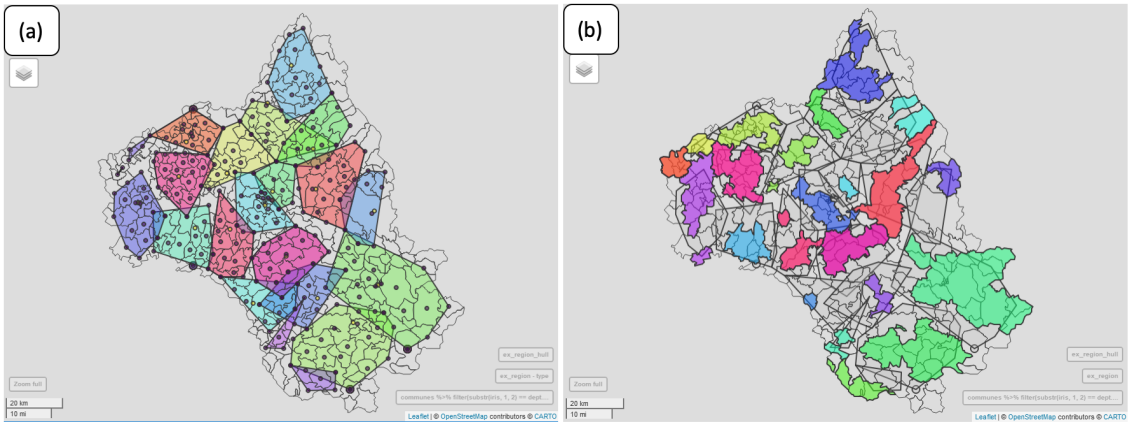
Figure A.5: Panel (a): obtained convex hulls for the Haute-Garonne departement. Panel (b): zoom on the urban area of the city of Toulouse.



We then use the computed convex hulls and the knowledge of the spatial boundaries of (sub)counties

to build a mapping from (sub)counties to substations. First, a (sub)county c that intersects with the convex hull of substation s is assumed to connect to this substation. If a given (sub)county intersects with several convex hulls, a distributed generation unit located in this (sub)county is deemed equally likely to connect to the corresponding substations. This first step maps almost two thirds of our spatial units (29,330 out of 45,508). In addition, over half of our spatial units (24,585 out of 45,508) intersect with a single convex hull. Figure A.6 illustrate this first step by showing, for one of the 94 departements, the counties that intersect with a single substation convex hull. In a second step, we isolate remaining (sub)counties that are adjacent to one or several (sub)counties that were all matched in step 1 to the same substation. These (sub)counties are assumed to also connect to the corresponding substation. This second step, which maps 6,523 additional spatial units, aims at expanding in a sensible way the service territory of substations in areas where we initially observe a relatively small number of distributed generation units. Third, we focus on remaining (sub)counties that are adjacent to one or several (sub)counties matched in either step 1 or 2. A user located in these (sub)counties is assumed to be equally likely to connect to either of the substations that were matched to the neighbor (sub)counties. This third step further maps 8,569 (sub)counties to substations.

Figure A.6: Panel (a): obtained convex hulls for the Aveyron departement. Panel (b): (sub)counties that intersect with a single convex hull, grouped by corresponding substations.



In the end, our spatial matching procedure allows us to map over 97% of our spatial units (44,422 out of 45,508) to substations. Reassuringly, the vast majority of unmatched spatial units are located outside of the main DSO's service territory, and are thus very likely to be supplied by distribution substations that we do not observe.

Finally, we use our mapping from (sub)counties to substations to build timeseries of installed capacities by technology at the substation level. When doing so, we account for the entry/exit of the 114 substations (5% of total) that are not observed for the full 14-year period. Indeed, the spatial matching of Step C is done in a static fashion, meaning it takes into account all known substations irrespective of their (de)commissioning date. However, when computing substation level installed capacities for a given year, we restrict attention to the substations that are known to exist in that year. In particular, although we may observe the upstream substation of a distributed generation unit as of 2018, this substation may not be commissioned yet in the early years of our period. In such (rare) case, the distributed generation unit is treated as an observation with unknown upstream substation for that year, and we use our mapping from (sub)counties to substations to assign the corresponding capacities to substations that existed in that year.

B Sensitivity analyses

This Appendix discusses a number of robustness checks and sensitivity analyses.

B.1 Sensitivity to our assignment procedure for units with unknown substations

As discussed in Appendix A, the results reported in the main text implement an assignment procedure to form guesses about the substation(s) to which distributed generation units with unobserved upstream substation are most likely to connect. This Appendix shows that our results hold irrespective of the details of the assignment procedure, and are even robust to ignoring distributed generation units with unknown substation altogether.

Sensitivity to how we match aggregated PV units (Steps A and B of the assignment procedure described in Appendix A)

About a quarter of installed PV capacities (as of 2018) consist in small (<36kW) units, for the most part aggregated at the (sub)county level. As discussed in Appendix A, their aggregated nature prevents us from accurately observing their evolution over time at a finer spatial granularity than departements. We thus implemented four contrasted methodologies to infer how (sub)county-level installed capacities may have evolved over time, under the constraint of being consistent with known departement-level capacities.

(Sub)county-level PV capacities from small aggregated units are then added up at the substation level using our spatial matching methodology (Step C of the assignment procedure described in Appendix A) and further added to substation level PV capacity from units that are listed individually in the public inventory of power plants. In the end, we obtain four different measures of total

installed PV capacity connected to a given substation in a given year, depending on the methodology used in Step B of our assignment procedure. These measures include both individually-listed and aggregated PV units.

Figure B.1: Top-right corner: scatter plots of pair-wise relationship between the different installed PV capacity metrics (unit of observation: substation by year). Darker colors correspond the later years. Bottom-left corner: corresponding coefficients of correlation.

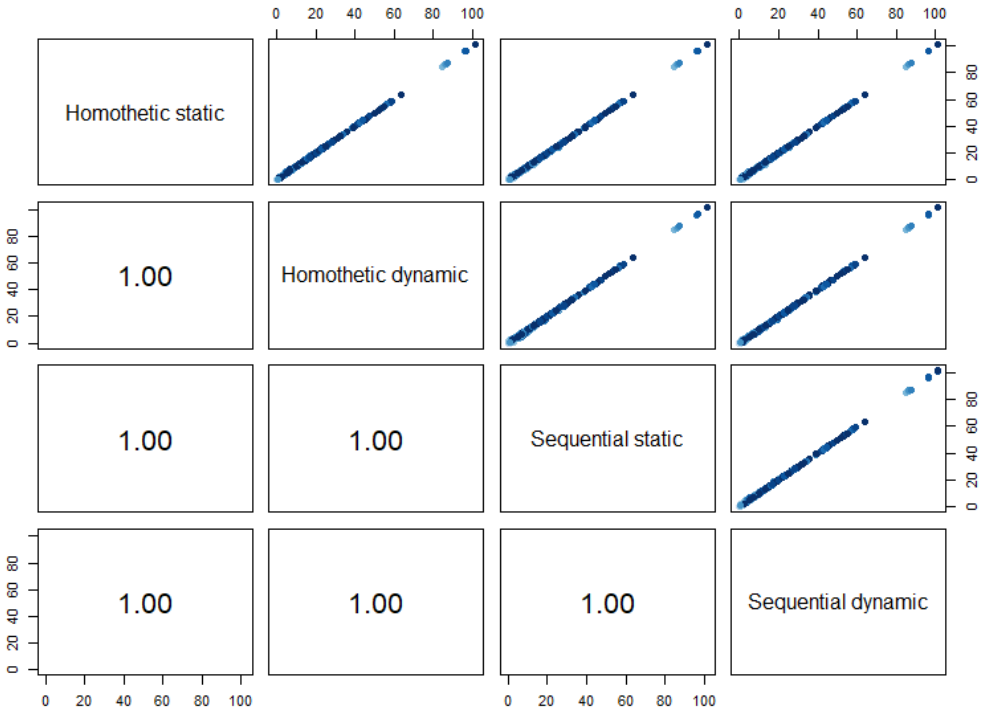


Figure B.1 shows that the four PV metrics we obtain are virtually identical. This is so for several reasons. First, three quarter of installed PV capacities come from individually-listed units. Second, we observe precisely where small aggregated units are as of 2018, which has the highest amount of installed PV capacities since the number of installations has grown over time. Third, even if we use contrasted approaches to assign installations to (sub)counties, these spatial units are then aggregated when we infer the territories served by the different substations. This aggregation

step tends to smooth any difference between the different allocation methods.

In the absence of any significant difference between our four metrics for installed PV capacities, the main text and the rest of the sensitivity analyses use the homothetic static approach in Step B of the assignment procedure. This method is indeed both the most parsimonious.

Sensitivity to including matched capacities (Step C of our assignment procedure)

In Equation (1), the installed capacity $K_{t,s,y}$ of distributed generation technology t connected to substation s in year y is the sum of two terms:

$$K_{t,s,y} \equiv K^0_{t,s,y} + \hat{K}_{t,s,y} \tag{9}$$

where:

- $K^0_{t,s,y}$ is the capacity from distributed generation units which are *known* to connect to this substation, as directly observed in the public inventory of power plants;
- $\hat{K}_{t,s,y}$ is the capacity from distributed generation units which are *assumed to* connect to this substation, as inferred from our assignment procedure.

Table 5: Decomposition of the variance (in MW^2) of the installed capacities of each technology between known and inferred capacities.

Technology	Var($K^0_{t,s,y}$)	Var($\hat{K}_{t,s,y}$)	Var($K_{t,s,y}$)
Wind	117.2	6.0	132.2
PV	12.8	0.5	14.7
Small hydro	6.7	0.2	7.0
Renewable thermal	2.8	0.1	2.9
Non renewable thermal	7.2	0.3	7.5

Table 5 reports the variances of $K^0_{t,s,y}$, $\hat{K}_{t,s,y}$ and $K_{t,s,y}$ for all five technologies, across all years and substations. We observe that our identifying variation almost exclusively comes from installed

capacities for which we directly observed the upstream substation in the public registry of power plants. Accordingly, our results are robust to ignoring altogether capacities that were inferred from our assignment procedure. Tables 6 and 7 report the results of our quantile regressions when using $K^0_{t,s,y}$ instead of $K_{t,s,y}$ as our independent variables. These results are almost identical to the results reported in the main text.

Table 6: Estimated coefficients when regressing net load distribution quantiles on the installed capacities $K^0_{t,s,y}$ from units for which upstream substation is directly observed. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	-0.723^{***}	-0.463^{***}	-0.271^{***}	-0.139^{***}	-0.091^{***}	-0.066^{***}	-0.035^{***}
	(0.018)	(0.014)	(0.008)	(0.004)	(0.005)	(0.005)	(0.006)
PV	-0.557^{***}	-0.384^{***}	-0.173^{***}	-0.052^{***}	-0.020^{*}	-0.0004	0.004
	(0.035)	(0.034)	(0.017)	(0.010)	(0.010)	(0.012)	(0.014)
Small hydro	-0.394^{***}	-0.359^{***}	-0.249^{***}	-0.139^{***}	-0.126^{***}	-0.127^{***}	-0.127^{***}
	(0.066)	(0.061)	(0.033)	(0.024)	(0.031)	(0.034)	(0.037)
Renewable thermal	-0.390^{***}	-0.373^{***}	-0.352^{***}	-0.332^{***}	-0.279^{***}	-0.234^{***}	-0.182^{***}
	(0.065)	(0.057)	(0.051)	(0.051)	(0.052)	(0.056)	(0.062)
Non renewable thermal	-0.083^{**}	-0.069^{***}	-0.060^{***}	-0.065^{***}	-0.107^{***}	-0.130^{***}	-0.127^{***}
	(0.034)	(0.024)	(0.021)	(0.024)	(0.029)	(0.032)	(0.040)
Observations	30,091	30,091	30,091	30,091	30,091	30,091	30,091
R^2	0.958	0.960	0.976	0.983	0.983	0.985	0.984
Adjusted R^2	0.955	0.957	0.974	0.981	0.982	0.983	0.983
Note:							*p<0.1; **p<0.05; ***p<0.01

B.2 Sensitivity to our model specification

Our econometric model is a two-way fixed effect specification. This very parsimonious approach accounts for both unobserved substation specific unobserved characteristics (as long as they are constant over time) and year specific unobserved characteristic (as long as they are uniform across space). While these controls account for a large number of possible con-founders, explanatory variables of electricity load that have experienced both significant and spatially contrasted changes over our period of interest, could introduce some estimation bias if they are correlated with installed distributed generation capacities. For example, although residential PV represents a small fraction

Table 7: Estimated coefficients when regressing hourly ramps distribution quantiles on the installed capacities $K^0_{t,s,y}$ from units for which upstream substation is directly observed. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	-0.160^{***}	-0.051^{***}	-0.019^{***}	0.0001	0.020^{***}	0.051^{***}	0.156^{***}
	(0.004)	(0.002)	(0.001)	(0.0001)	(0.001)	(0.002)	(0.004)
PV	-0.168^{***}	-0.067^{***}	-0.016^{***}	-0.003^{***}	0.020^{***}	0.071^{***}	0.153^{***}
	(0.011)	(0.006)	(0.001)	(0.0004)	(0.001)	(0.005)	(0.012)
Small hydro	-0.025^{**}	-0.006^{**}	0.001	0.001	0.002	0.001	0.035^{**}
	(0.012)	(0.003)	(0.001)	(0.001)	(0.001)	(0.004)	(0.017)
Renewable thermal	−0.017	−0.001	−0.002	0.001	0.002	−0.002	0.010
	(0.016)	(0.007)	(0.003)	(0.002)	(0.003)	(0.007)	(0.021)
Non renewable thermal	−0.002	0.002	−0.0002	-0.001^{*}	−0.001	0.001	0.002
	(0.008)	(0.003)	(0.002)	(0.001)	(0.002)	(0.004)	(0.008)
Observations	30,091	30,091	30,091	30,091	30,091	30,091	30,091
R^2	0.957	0.966	0.968	0.836	0.968	0.963	0.955
Adjusted R^2	0.954	0.964	0.965	0.823	0.965	0.960	0.952
Note:							*p<0.1; **p<0.05; ***p<0.01

of total PV capacities,[32] population growth could both drive electricity consumption up and be correlated with PV adoption, to the extent that solar panels are more likely to be installed on new buildings. To test for this possibility, we compare the obtained results from our main specification to the results obtained when adding year by region (mainland France has 12 regions) and year by departement (mainland France has 94 departements) fixed effects. Table 8 reports the obtained results when using mean net load as our dependent variable, which enables a direct interpretation of the estimated coefficients as capacity factors. Our estimates appear to be robust to the inclusion of these additional controls. Similar conclusions were reached with specifications using the values of the different quantiles as dependent variables.

[32]Residential PV installations are included in the small aggregated PV units, and thus represent at most a quarter of total capacities.

Table 8: Estimated coefficients when regressing mean net load on installed capacities (i) using our main specification (column (1)) ; (ii) adding year by region fixed effects (column (2)) ; and (iii) adding year by departement fixed effects (column (3))

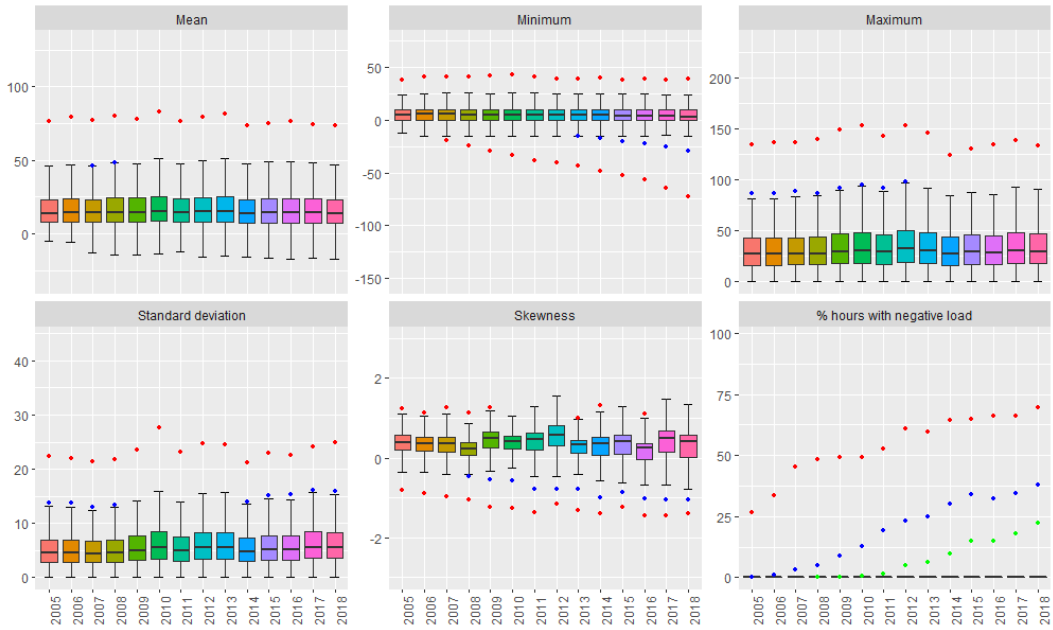
	Dependent variable:		
	Mean net load		
	(1)	(2)	(3)
Wind	-0.192^{***}	-0.187^{***}	-0.180^{***}
	(0.006)	(0.007)	(0.007)
PV	-0.109^{***}	-0.118^{***}	-0.120^{***}
	(0.012)	(0.012)	(0.011)
Small hydro	-0.192^{***}	-0.209^{***}	-0.200^{***}
	(0.025)	(0.025)	(0.024)
Renewable thermal	-0.302^{***}	-0.298^{***}	-0.296^{***}
	(0.050)	(0.052)	(0.051)
Non renewable thermal	-0.081^{***}	-0.082^{***}	-0.088^{***}
	(0.024)	(0.024)	(0.026)
year FE	Y	N	N
substation FE	Y	Y	Y
year by region FE	N	Y	N
year by departement FE	N	N	Y
Observations	30,091	30,091	30,091
R^2	0.983	0.984	0.985
Adjusted R^2	0.982	0.983	0.983
Note:		*p<0.1; **p<0.05; ***p<0.01	

C Impact of the different distributed generation technologies on other summary statistics

Restricting attention to a given substation s in a given year y defines a distribution $\{L_{s,y}(h)\}_h$, where $L_{s,y}(h)$ is the net load for substation s for the hour h of year y. Numerous summary statistics can then be derived from these distributions. Figure C.2 shows the evolution, between 2005 and 2018, of the cross-sectional distribution of the following summary statistics: mean, minimum, maximum, standard deviation, skewness, and percentage of hours with negative net load, that is hours during which power was flowing from the distribution to the transmission grid. Interestingly, we observe that the summary statistics that exhibit the most significant changes relate to reverse power flows, that is to hours where local generation exceeds local consumption. For example, the left tail of the distribution of minimum net load and the right tail of the distribution of the percentage of hours with negative net load have been expanding significantly.

We can then estimate the impact that different distributed generation technologies have on these main summary statistics represented. We use the specification of Equation (1) with the summary statistics plotted on Figure C.2 as dependent variables. Table 9 shows the obtained results. First, we note that the coefficients estimated when using mean net load as the dependent variable may be interpreted as capacity factors, that is as the ratio of average generation over installed capacity. For example, our results suggest a capacity factor of 11% for PV and 19% for wind. These estimates are close to but somewhat smaller than publicly reported capacity factors of 14 and 21% respectively (RTE, 2018). Power losses between the generation site and the substation, as well as the fact that installed capacities are measured as of 31 December, are possible rationales for getting smaller estimates. In addition, small-scale PV installations, whose output is typically not observed by the TSO, seem likely to have lower capacity factors than larger units due to less efficient technologies

Figure C.2: Evolution of a sample of summary statistics for the distribution of hourly net loads in a given year at a given substation. "Minimum" and "maximum" net load are actually the 1st and 999th 1000-quantiles to account for the possibility of idiosyncratic measurement errors. Minimum, maximum, mean and standard deviation statistics are expressed in MW. Boxes locate the first, second and third quartiles of the distributions. Top whiskers (resp. bottom whiskers) are drawn at a distance of 1.5 interquartile range above the third quartile (resp. below the first quartile). When they fall outside of the interval delimited by whiskers, the 1st, 5th and 10h (resp. the 99th, 95th and 90th) centiles are respectively depicted as red, blue and green dots. For more clarity, the tails of the distributions are censored for the skewness metric.



and more frequent outages.[33] Second, we observe that different technologies have very contrasted impacts on minimum and maximum net load. While non-renewable thermal units impact minimum and maximum net load in a similar way as they impact mean net load, PV and wind seem to have a much lower impact on peak load. We discuss these differences in more details below. Third, an increase in PV, wind and small hydro capacities is found to be associated with more negatively skewed and more volatile distributions of net loads. By contrast, thermal units have much milder impacts. Again, we discuss these observations further when we look at the impact of the different

[33]Consistently, estimating the same model when replacing total installed PV capacity by installed PV capacity from units for which we observe the upstream substation (which tend to be larger installations) yields a capacity factor of 12%.

technologies on the distribution of hourly ramps. Finally, we observe that reverse power flows seem to be driven by wind, PV and small hydro capacities.

Table 9: Estimated impact of the different distributed generation technologies on a set of summary statistics for the distribution of substation net hourly loads. When relevant, variables are expressed in MW. "Minimum" and "maximum" statistics are actually the 1st and 999th 1000-quantiles to account for the possibility of idiosyncratic measurement errors. Robust standard errors clustered at the substation level are reported.

	Dependent variable:					
	Mean	Minimum	Maximum	Standard Deviation	Skewness	% hours net load <0
Wind	-0.192^{***}	-0.735^{***}	-0.030^{***}	0.146^{***}	-0.019^{***}	0.525^{***}
	(0.006)	(0.025)	(0.006)	(0.008)	(0.001)	(0.026)
PV	-0.109^{***}	-0.565^{***}	0.020	0.130^{***}	-0.028^{***}	0.385^{***}
	(0.012)	(0.039)	(0.015)	(0.014)	(0.002)	(0.035)
Small hydro	-0.192^{***}	-0.363^{***}	-0.143^{***}	0.072^{***}	-0.015^{***}	0.817^{***}
	(0.025)	(0.058)	(0.039)	(0.025)	(0.005)	(0.180)
Renewable thermal	-0.302^{***}	-0.350^{***}	-0.171^{***}	0.037^{**}	0.001	0.053
	(0.050)	(0.068)	(0.061)	(0.017)	(0.004)	(0.069)
Non renewable thermal	-0.081^{***}	-0.091^{**}	-0.107^{**}	-0.019^{**}	-0.005	-0.021
	(0.024)	(0.038)	(0.045)	(0.009)	(0.004)	(0.034)
Observations	30,091	30,091	30,091	30,091	30,091	30,091
R^2	0.983	0.952	0.983	0.957	0.361	0.885
Adjusted R^2	0.982	0.948	0.981	0.953	0.311	0.875
Note:						*p<0.1; **p<0.05; ***p<0.01

D Statistical tests

This Appendix provides details on the statistical tests we ran to assess the characteristics of the quantile impact functions for the load duration curve. We follow the approach derived in Wolak (1987, 1989). More specifically, we stack the 7 quantiles regressions (for quantiles 1, 10, 25, 50, 75, 90 and 99) into a single model. We then compute the variance $\text{Var}(\hat{\beta})$ of the ordinary least square estimator. For tractability reasons, residuals for the stacked model are obtained by estimating each regression separately. More precisely, we first regress the dependent and independent variables on our set of fixed effects, and then use the residuals from these regressions to estimate

the regressions for each quantile level. We obtain 35 coefficients $\hat{\beta}_{t,q}$ where t indexes distributed generation technologies and q indexes quantiles. We denote $\hat{\beta}$ the corresponding vector of estimated coefficients.

Our test statistic τ is then the optimized value of the following problem:

$$\begin{gathered} \tau \equiv \min_{\delta} (\hat{\beta} - \delta)^T \mathrm{Var}(\hat{\beta})^{-1} (\hat{\beta} - \delta) \\ \text{s.t.} \\ \text{HC0-h} \end{gathered} \tag{10}$$

The constraint HC0-h formalizes the different null hypotheses we test in terms of linear equality or inequality constraints on δ. More specifically, for a given technology t, these constraints are:

- **HC0-peak:** the coefficient for the impact on the 99th quantile of the distribution of hourly net load is zero

$$\delta_{t,99} = 0$$

- **HC0-inc:** the quantile impact function is increasing

$$\delta_{t,1} \leq \delta_{t,10} \leq \delta_{t,25} \leq \delta_{t,50} \leq \delta_{t,75} \leq \delta_{t,90} \leq \delta_{t,99}$$

- **HC0-inc-peak:** the quantile impact function is increasing and the coefficient for the impact on the 99th quantile of the distribution of hourly net load is zero

$$\delta_{t,1} \leq \delta_{t,10} \leq \delta_{t,25} \leq \delta_{t,50} \leq \delta_{t,75} \leq \delta_{t,90} \leq \delta_{t,99} \text{ and } \delta_{t,99} = 0$$

- **HC0-dec:** the quantile impact function is decreasing

$$\delta_{t,1} \geq \delta_{t,10} \geq \delta_{t,25} \geq \delta_{t,50} \geq \delta_{t,75} \geq \delta_{t,90} \geq \delta_{t,99}$$

We run a total of 20 statistical tests (5 technologies times 4 null hypotheses). Table 10 reports the obtained test statistics.

Table 10: Obtained test statistics

Technology	HC0-peak	HC0-inc	HC0-inc-peak	HC0-dec
Wind	145.92	0	145.92	46,154.49
PV	1.26	0	1.26	7,388.58
Small hydro	18.36	0.12	18.36	144.89
Renewable thermal	42.67	0	42.67	54.83
Non renewable thermal	59.39	90.32	97.21	6.06

As described in Wolak (1987, 1989), the null distribution of the test statistics is a weighted sum of chi-square distributions ranging from zero to P degrees of freedom (where P is the number of constraints). Because the weights sum to one, bounds for the exact critical values for the test statistic can be obtained from the critical values of the chi-square distribution with the most unfavorable number of degrees of freedom. In our application, these bounds appear to be sufficient to infer the result of the statistical tests. For example, HC0-inc simultaneously tests for 6 inequalities. Since $\Pr\left[\chi_1^2 \geq 2.706\right] = 0.1$, we cannot reject the null hypothesis even at the 0.1 level whenever the test statistic is lower than 2.706. Conversely, since $\Pr\left[\chi_6^2 \geq 16.812\right] = 0.01$, a test statistic higher than 16.812 rejects the null hypothesis at the 0.01 level (the critical value for the 0.01 level being weakly less stringent). To fix ideas about the ranges of critical values, the upper-tail critical values of χ^2 distribution with 1 (resp. 7) degrees of freedom are 2.706 and 6.635 (resp. 12.017 and 18.475) for probabilities 0.1 and 0.01.