Can Distributed Intermittent Renewable Generation Reduce Future Grid Investments? Evidence from France*

Nicolas Astier† Ram Rajagopal‡ Frank A. Wolak§

10 May 2022

Abstract

This paper estimates the relationship between investments in five distributed generation technologies and hourly net withdrawals from over 2,000 electricity distribution networks in France between 2005 and 2018. We find that investments in distributed wind and solar generation have little or no impact on the annual peak of hourly net withdrawals from the distribution grid, while investments in hydroelectric and thermal distributed generation significantly reduce it. An optimistic analysis of the impact of investments in battery storage suggests that high levels are required for distributed wind and solar to deliver similar reductions in the annual peak of hourly net withdrawals. Our results imply that public policies favoring distributed wind and solar generation over utility-scale generation cannot be rationalized by savings in future grid investments.

Keywords: energy transition, 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. We are grateful to seminar participants at the Paris School of Economics, Stanford University, Mines ParisTech, Monash University, Arizona State University, ZEW Mannheim, the 26th Annual Conference of the EAERE, the EMEE Summer workshop, UC Berkeley EI Camp and TPUG session at the AEA ASSA meeting for their feedback and suggestions. Remaining errors are ours.

†Paris School of Economics, Paris, France, nicolas.astier@psemail.eu. This research started when the author was affiliated with the Precourt Institute for Energy (Bits & Watts initiative), Stanford University, Stanford, CA 94305, USA.

‡Department of Civil and Environmental Engineering, Stanford University, 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

Increasing the share of electricity from wind and solar resources and electrifying sectors that traditionally consume fossil fuels such as transportation and space heating is the current consensus path to reducing global greenhouse gas emissions. These wind and solar resources may be deployed as large ("utility-scale") facilities connecting to the high-voltage transmission network as shown in the left panel of Figure 1, or as smaller but more numerous "distributed" units connecting to a distribution network as shown in the right panel of Figure 1. Because both approaches have different strengths and weaknesses, the most cost-efficient way to increase renewable electricity generation is highly debated.

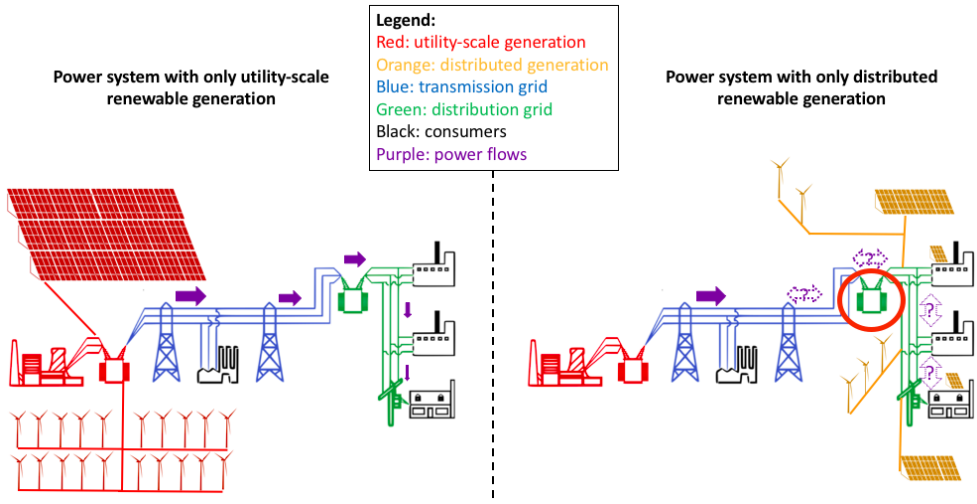


FIGURE 1. Illustration of the difference between utility-scale and distributed generation (adapted from U.S. Department of Energy)

Utility-scale wind and solar generation units can benefit from economies of scale and locating where there are rich wind and solar resources. As a result, they produce electricity at a significantly lower levelized cost than distributed units. Figure 2 plots the global annual capacity-weighted-average of the levelized cost of energy (LCOE) of new utility-scale solar generation

units, along with the annual average LCOE of new commercial and residential distributed solar systems in France from 2010 to 2020.[1] This figure shows that from 2011 onwards the LCOE of distributed solar units is two to three times the LCOE of utility-scale solar units.

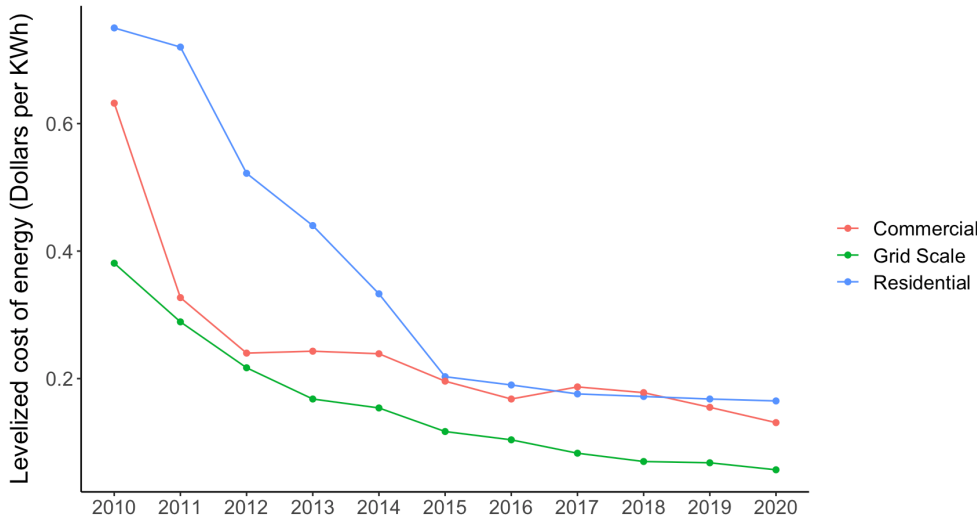


FIGURE 2. Levelized Cost of Energy (LCOE) for utility-scale units (global capacity-weighted average) and distributed solar generation units (average for France) from 2010 to 2020 ($/kWh). Adapted from IRENA (2021).

Because distributed generation units are located closer to end-consumers, as shown in Figure 1, they can reduce electricity transportation costs. These costs consist of energy losses and grid investments. Annual transmission and distribution (T&D) power losses represent less than 10% of the electricity produced in virtually all industrialized countries. Therefore, the potential savings in energy losses are insufficient to close the gap between the LCOE of distributed versus utility-scale wind and solar generation. As a result, unless the savings in future T&D network investments are substantial, utility-scale solar and wind generation units connected to the transmission network are the least-cost sources of renewable electricity.

1. Numbers on Figure 2 are extracted from Figure 3.1 and Table 3.3 of IRENA (2021).

The extent to which investments in distributed generation reduce the need for future T&D network investments is however very difficult to assess. Investments in grid upgrades are indeed triggered by increases in peak rather than average usage. Because wind and solar generation is intermittent, one needs to estimate whether the output of renewable distributed units will coincide with local peak consumption. This assessment is in practice performed by engineering simulations. Such simulations are generally run for a handful of representative distribution networks, and the conclusion reached can be very sensitive to the assumptions made.[2] In addition, residential consumers installing a rooftop PV system may subsequently change their electricity consumption habits as discussed in Qiu et al. (2019).

Many consultant reports claim significant avoided costs associated with distributed solar photovoltaic (PV) investments.[3] Because T&D network costs typically comprise at least one-third of electricity bills, the claimed savings could easily add up to billions of dollars annually for a country like the U.S.[4] Yet, other engineering studies reach less optimistic conclusions (Cohen and Callaway, 2016; Cohen et al., 2016).

We contribute to this debate by taking advantage of a unique panel dataset of distributed generation units and hourly net withdrawals from over 2,000 local distribution networks throughout France for 2005-2018. The hourly net withdrawal from a local distribution grid is the hourly net flow of energy, positive or negative, between the high voltage transmission network and the

2. Key assumptions include the installed capacities of distributed generation, hourly gross consumption patterns, and hourly distributed generation output (possibly accounting for outages and maintenance).

3. Muro and Saha (2016) survey several such studies.

4. According to the United States Energy Information Administration, total annual capital investments in distribution networks alone by major U.S. utilities serving about 70% of national electricity demand were higher than 25 billion dollars in 2017 (www.eia.gov/todayinenergy/detail.php?id=36675). It is often claimed that distributed generation could avoid a significant fraction of these costs: "*on-site production avoids transmission and distribution costs, which otherwise amount to about 30% of the cost of delivered electricity*" (International Energy Agency, 2002).

distribution network measured at the transformer facility circled in red in the right panel of Figure 1.[5]

During our fourteen year sample, over 25 gigawatts (GW) of distributed generation investments occurred. The bulk of these investments were in wind and PV generation capacity, as shown Figure 3. These installations are estimated to have produced roughly 7% of the total electricity supply in France in 2018 (RTE, 2019). We use our dataset to estimate the impact of a marginal increase in the installed capacity of five types of distributed generation technologies (solar PV, wind, small hydro, dispatchable renewable thermal, and dispatchable non-renewable thermal) on the different quantiles of the annual distribution of hourly net withdrawals from distribution networks. Because future T&D investments are primarily driven by peak usage, we pay particular attention to the impact of distributed generation on the highest quantiles of this distribution.

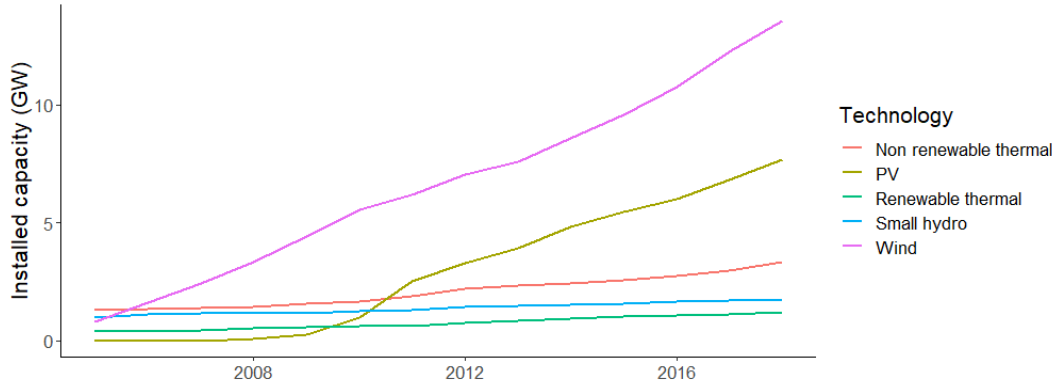


FIGURE 3. Total installed capacities of distributed generation (in GW) by year and technology in France, as observed in our final dataset (which includes the vast majority of distributed generation units in mainland France, see Section 4).

We find that investments in distributed wind and solar generation have little or no impact on the annual peak of hourly net withdrawals from the distribution

5. Net withdrawals from a distribution grid equal gross consumption plus losses minus distributed generation.

grid. More precisely, a marginal 1 MW increase in distributed solar PV capacity has no statistically discernable impact on the highest percentiles of the annual distribution of hourly net withdrawals from the distribution grid. In addition, a marginal 1 MW increase in distributed wind generation capacity is predicted to reduce the 99th percentile of the annual distribution of hourly net withdrawals from distribution networks in France by only 0.037 MW. In contrast, a marginal 1 MW investment in distributed hydro, non-renewable thermal, or renewable thermal generation units predict a reduction in the 99th percentile of the annual distribution of hourly net withdrawals from the distribution grid by more than 0.1 MW.

We then look at the impact of distributed generation on the variability of net withdrawals from the distribution grid. Consistently with industry practice, we assess this variability by looking at hourly ramp rates, defined as differences between two consecutive hourly net withdrawals from the distribution grid. High ramp rates (in absolute value) correspond to rapid increases or decreases in net withdrawals from the distribution grid, which may at some point start raising operational difficulties. We find that investments in either wind or PV distributed generation are associated with significant increases in the magnitude of hourly ramp rates. A 1 MW increase in PV or wind capacity predicts a 0.15 MW.h^{-1} decrease (resp. a 0.14 MW.h^{-1} increase) in the 1st percentile (resp. the 99th percentile) of the annual distribution of the hourly ramp rates. In other words, investments in distributed wind and solar generation are found to increase the largest absolute differences in consecutive hourly net withdrawals from the distribution grid by at least 15% of their nameplate capacity. Investments in the other three distributed generation technologies do not predict a non-zero change in any percentile of the annual distribution of hourly ramp rates.

Finally, we explore two ways in which distributed wind and solar generation may deliver higher savings in future grid investments. We first test whether the

marginal benefits from distributed generation investments vary with the level of existing capacity, which would in turn imply that lower levels of investments in a specific distributed generation technology and/or a more spatially uniform distribution of installations could bring higher benefits. We find that the first MWs of distributed wind capacity connecting to a distribution network are predicted to reduce the 99th percentile of the annual distribution of hourly net withdrawals by 7% of their nameplate capacity, almost double the average effect for all capacity investment levels. The impact of distributed solar investments on the highest percentiles of the annual distribution of hourly net withdrawals from the distribution grid is however not statistically different from zero even at low levels of penetration of distributed PV generation.

We then explore whether installing battery storage along with wind and solar distributed generation capacity could deliver reductions in annual peak net withdrawals similar to those obtained for hydroelectric and thermal distributed generation investments. We show that storage investments substantially higher than existing levels would have to accompany investments in distributed solar and wind generation to reach this outcome. Specifically, we find that roughly one Tesla Powerwall 2 battery would need to be installed for every 3 kW of distributed wind or solar generation in order to achieve reductions in the 99th percentile of the annual distribution of hourly net withdrawals close to 10% of the nameplate capacity of distributed generation units. Because 3 kW is the smallest rooftop solar system typically installed, this means that virtually every PV installation would have to be accompanied by a battery storage system.

Taken together, our results suggest that, at least for the case of France, the benefits from deferring future T&D expansions cannot rationalize policy support for distributed wind and solar generation over utility-scale installations. Investments in distributed wind and solar without any storage capacity are predicted to result in small or zero reductions in the highest percentiles

of the distribution of hourly net withdrawals from distribution networks, which ultimately determine how much grid capacity is built. In addition, we find that investments in distributed wind and solar generation are associated with a significant increase in the variability of net withdrawals from the distribution grid, which may require additional distribution network investments to maintain a reliable supply of energy to customers. Consequently, until substantial transmission network expansions are necessary to connect new utility-scale wind and solar generation capacity in France, the most cost-effective approach to increase renewable electricity production is to rely on utility-scale installations.

The remainder of the paper is organized as follows. The next section reviews the literature and discusses the role of distributed generation in the energy transition. Section 3 explains why investments in distributed generation units may reduce the need for future investments in T&D network capacity. Section 4 describes the data sources we use. Section 5 details our empirical strategy. Section 6 presents our main results. Section 7 discusses non-linear marginal impacts, the potential role for battery storage and the applicability of our results to other jurisdictions. Section 8 concludes.

2. Distributed Generation and the Energy Transition

In recent years, annual global investments in wind and solar electricity generation have reached about $300 billion. According the International Renewable Energy Agency (IRENA, 2021), the magnitude of these annual investments should increase in the future. Distributed generation units represent about half of the investments in solar PV,[6] which is somewhat

6. Because of their higher investment cost per unit of capacity, distributed generation units roughly account for half of the investments in PV generation (in $) but only about one-third of installed capacity (in GW).

surprising because of their significantly higher LCOE in comparison to utility-scale installations as shown in Figure 2.

In practice, distributed generation usually benefits from a stronger policy support, for example through direct (e.g. higher feed-in-tariffs for small installations) or indirect (e.g. net metering) subsidies (see Section 5.2 for the case of France). Because both utility-scale and distributed renewable installations provide zero carbon electricity, distributed generation units must provide other additional benefits in order to be preferred to utility-scale facilities.

Besides greater customer control over their electricity supply, the major additional benefit claimed by distributed generation advocates is the avoided cost of T&D network investments (General Electric Power, 2018). There is a large literature in engineering characterizing and quantifying the potential for distributed generation investments to defer T&D investments. In particular, a number of papers have proposed methodologies to quantify T&D deferral benefits (Hoff, 1996; Feinstein et al., 1997; Gil and Joos, 2006; Mendez et al., 2006). These methodologies typically estimate the extent to which distributed generation reduces peak net withdrawals from a given distribution network.[7]

In the early 2000s, distributed generation mostly consisted of dispatchable thermal units such as back-up diesel generators. Studies assessing the T&D deferral potential of distributed generation therefore assumed that these units were available to produce energy whenever needed (Brown et al., 2001; Piccolo and Siano, 2009; Wang et al., 2009). By contrast, wind and PV output is intermittent. Researchers and utilities thus have to estimate the extent to which these technologies produce electricity during hours of peak demand in the distribution grid. In this context, the external validity of available results is

7. "*Utilities generally make investment decisions for generation and T&D capacity based on peak requirements. Thus, any reduction in peak power requirements provides direct benefits to the utility in the form of deferred capacity upgrade costs*" (U.S. Department of Energy, 2007).

particularly difficult to assess. Engineering simulations indeed typically focus on a single or a handful of distribution networks, and often study an optimized rather than an observed deployment of distributed generation units. They may also poorly capture the influence of hard-to-model real-world factors, such as site-specific irradiation, how quickly equipment failures are detected and repaired, or behavioral responses from end-consumers when they become producers of electricity.

Borenstein (2020) argues qualitatively that the T&D benefits from rooftop PV should be small but notes the lack of academic literature on the topic.[8] On the one hand, studies that have explored the social value of wind and solar PV have focused on environmental benefits (Cullen, 2013; Novan, 2015; Callaway et al., 2018; Fell et al., 2021; Sexton et al., 2021), intermittency (Gowrisankaran et al., 2016), displaced generation costs (Callaway et al., 2018) or transmission congestion (Fell et al., 2021; Sexton et al., 2021). Even though more than a third of PV generation capacity in the U.S. is distributed (EIA, 2021), none of these studies account for the sub-transmission and distribution grids, notably because of the difficulty in gaining access to relevant data over a large service territory. On the other hand, studies that do look specifically at distributed generation, usually in the form of residential PV, have focused on households' decision to install such systems (De Groote and Verboven, 2019; Gillingham and Bollinger, 2021). These studies do not discuss the cost-effectiveness of distributed generation units relative to utility-scale installations in providing additional renewable energy.

Two studies closely relate to our work. First, Cohen et al. (2016) leverage high resolution data on actual generation from distributed PV units in California to study the impact of distributed solar generation on distribution feeders. They however have to rely on forecasts and simulations for feeder load

8. Similarly, Sexton et al. (2021) acknowledge in their footnote 5: that "*evidence on distribution cost avoidance by distributed solar capacity is scarce*".

data. Second, the on-going work by Ovaere et al. (2020) uses data from a local utility in Connecticut to estimate the effect of distributed solar generation on the distribution network. While both studies provide important findings that complement our results, our work differs from them in several ways. First, rather than focusing only on solar PV, we study five distinct generation technologies. In addition, we rely on a dataset with over 2,000 local distribution networks for fourteen years. Because distribution systems are typically operated by local monopolies, the geographical scope of available data tends to be far more limited in U.S. case studies. In contrast, a single distribution system operator supplies 95% of end-consumers in France, which allows access to internally consistent data on the use of distribution networks over almost the entire country. Finally, observing the hourly net withdrawals from over 2,000 local distribution grids enables us to implement a novel empirical strategy where we study the annual distribution of hourly net withdrawals, their variability (as measured by hourly ramp rates), and simulate the impact of different levels of battery storage investments.

3. Distributed Generation and the Electricity Grid

In virtually all electricity supply industries, the majority of energy is supplied by utility-scale generation facilities that take advantage of economies of scale in production, transmitting the energy produced at high voltage to local distribution grids, where it is transformed to a lower voltage and transferred to final consumers.

Starting in the early 2000s however, many jurisdictions saw significant investments in more environmentally-friendly and smaller-scale generation units, generically referred to as "distributed generation". Due to their smaller size, these units tend to be located closer to final consumers and to connect to distribution grids. Because they inject electricity directly at the distribution grid level, distributed generation units can reduce the magnitude of net

withdrawals from distribution networks.[9] Distributed generation installations could thus reduce both the need for transmission capacity to move energy from large-scale generation units to distribution networks, and the need for distribution network capacity to move the energy from the transmission grid to final consumers.

3.1. Distributed Generation and Future Grid Investments

To understand how distributed generation might affect future investments in network capacity, we build on the power systems literature. The most direct approach to study this relationship is to look at net withdrawals from distribution networks. This interface consists of assets called distribution substations (circled in red in Figure 1).[10] In France, these substations typically host 63/20kV or 90/20kV transformers.[11] Substations hosting these types of transformers would sit at the edge of the sub-transmission network in the United States (U.S. Department of Energy, 2015). Although distribution networks have slightly different designs in Europe and the United States (notably due to differences in typical population density), the main characteristics that are relevant for our analysis are similar in both places.[12]

---

9. With sufficient distributed generation capacity, electricity may however no longer always flow from large-scale generation units to final consumers during some hours. We show in Appendix C that the share of distribution networks in France that are experiencing such situations at least once a year have been steadily increasing over time to reach about 25% in 2018.

10. By focusing on net withdrawals at the substation level, we neglect constraints that may occur at a more granular level such as an individual feeder. To the extent that the substation is operated so as to roughly balance load across the different feeders and distributed generation units spread sufficiently uniformly across the service territory of the distribution network, we expect more granular constraints to correlate to some extent with substation-level constraints.

11. These transformers lower the voltage of electricity that flows in the transmission network to levels closer to the voltage levels at which the energy is ultimately consumed (Kirschen and Strbac, 2018).

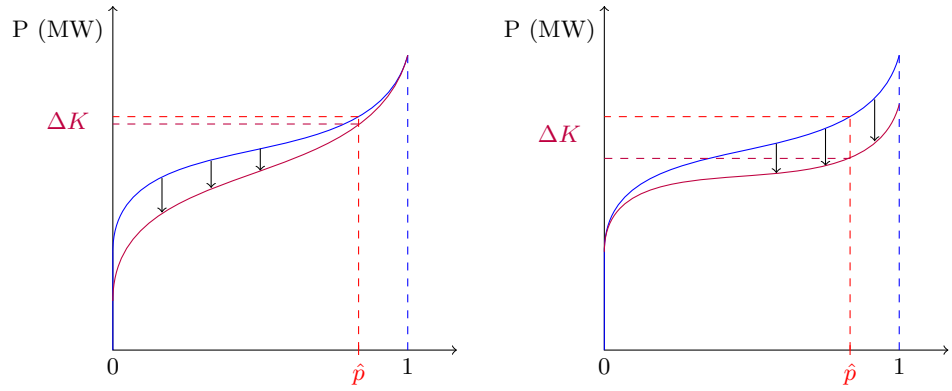
12. In particular, distribution systems are radial, and their voltages and power carrying capabilities are comparable.

We focus our attention on the percentiles of the annual distribution of hourly net withdrawals from a given distribution network. These percentiles map to what is known in the electricity industry as the "load duration curve", which is just another name for the inverse cumulative distribution function[13] of hourly net withdrawals. Because network planning rules typically rely on a reliability threshold $\hat{p}$ to decide on the size of a given grid component, the load duration curve is used in practice to assess the level of net withdrawals that are expected to be exceeded with a probability lower than $1 - \hat{p}$. As an illustration, the computation of distribution network tariffs in France, which builds upon the planning rules reported by utilities, makes use of the 71^{th} percentile for medium-voltage grid components, and the 94^{th} percentile for low-voltage grid components (CRE, 2021). Other countries may use different thresholds for distribution network planning and tariff setting.[14] Some may even choose to consider only the maximum value of net withdrawals. Either way, some notion of peak net withdrawals is always a key input to electricity network planning rules and tariffs.

Whether or not distributed generation decreases the grid capacity needed to ensure a given level of reliability depends on the extent to which this electricity is produced during local peak hours. Figure 4 illustrates this intuition. If the output from distributed generation is not coincident with the highest net withdrawals from the distribution network, grid expansions are unlikely to be deferred (left panel). By contrast, less network capacity is needed if this electricity is produced during peak hours (right panel).

13. Strictly speaking, the "load duration curve" used in the industry corresponds to the inverse cumulative distribution function for which the direction of the x-axis has been inverted. In addition, the unit used for the x-axis is "hours per year", which are just annual frequencies multiplied by 8760.

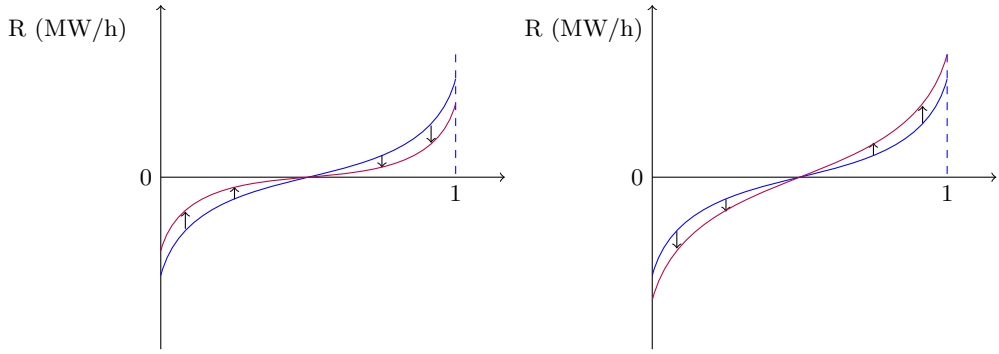
14. The chosen thresholds depend on parameters such as the probability of outages, the value of lost load, and the marginal cost of increasing network capacity.



Small grid capacity savings

Large grid capacity savings

FIGURE 4. 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 threshold.



Ramp magnitude decreases

Ramp magnitude increases

FIGURE 5. Illustration of how distributed generation may shift the ramp duration curve of a substation, decreasing or increasing the magnitude of hourly ramp rates.

Although somewhat technical, our approach for assessing the relationship between distributed generation investments and future grid savings is likely to have a greater level of external validity than alternative methods. In particular, distribution network costs are typically measured with a coarser granularity than a single distribution grid. In addition, because electricity networks are long-lived and lumpy assets, observed accounting costs usually

largely depart from actual economic costs. Finally, realized costs will depend on utility-specific planning rules, which can vary significantly across jurisdictions. Consequently, using distribution network costs or investment expenditures to study our research question would leave many reasons besides the need to manage net demand peaks for year-to-year changes in distribution network costs or investments.

In recent years, power system engineers have paid increasing attention to large variations in net electricity consumption levels over short periods of time, known as "ramps".[15] We thus also look at the impact of distributed generation on the annual distribution of hourly ramps. For a given substation in a given year, the hourly ramp in hour h is defined as the difference between the net withdrawal from the distribution network in hour $h+1$ and the net withdrawal in hour h. Mirroring our approach for net withdrawals, the collection of hourly ramps for a given year can then be sorted in increasing order to build a "ramp duration curve". One can finally estimate how this curve changes with increased levels of distributed generation. By construction, the integral of the ramp duration curve is close to zero.[16] As a result, the ramp duration curve starts at negative values, which correspond to hours during which net withdrawals are decreasing at the highest rates, and ends at positive values, which correspond to hours during which net withdrawals are increasing at the highest rates. The flatter the ramp duration curve, the less severe are the observed ramps.

Figure 5 illustrates two different ways in which distributed generation may impact the ramp duration curve faced by a given substation. In both panels, the

15. The most well-known example is the so-called "duck curve" in California where the rapid decrease in PV output in the evening makes it necessary to ramp up more than 10,000 MW of controllable generation capacity in about three hours (www.caiso.com/documents/flexibleresourceshelprenewables_fastfacts.pdf).

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

blue curve represents the pre-existing ramp duration curve and the purple curve the ramp duration curve after the addition of distributed generation. 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). Even though the former case seems unlikely to occur for intermittent generation technologies given the randomness of their output, it is nonetheless unclear whether they will have a null impact on the ramp duration curve or exacerbate ramps. In the latter case, operating the distribution grid is likely to become more complex and thus potentially more costly, for example due to more frequent tap changes for transformers or violations of operating constraints (e.g. voltage bounds or phase balancing).[17]

3.2. Background on the French Power System

Distributed generation installations generally belong to one of the following categories: wind, solar, small hydro, and thermal units using either renewable (e.g. wood, waste) or non-renewable (e.g. natural gas, diesel) fuels. Figure 3 shows how installed distributed generation capacities have evolved in France between 2005 and 2018. Although all technologies exhibit a significant upward trend, solar and wind have by far experienced the largest growth. As of 2018, there was approximately 28 GW of distributed generation in France.[18] Figures F.1 and F.2 in Appendix F show that distributed wind was mostly developed in the North of the country and distributed PV in the South.

17. For more details on power systems operations, see Kirschen and Strbac (2018).

18. As a comparison, the highest demand for grid-supplied electricity ever recorded in the country is about 100 GW. Intermittent generation units however do not always produce at full capacity, so that wind and solar have supplied less than 10% of the annual electricity consumption in France in 2018. In other words, a future electricity system relying predominantly on wind and solar would need to increase significantly further the installed capacities of these technologies relative to gross consumption.

Figure 6 illustrates the main features of electricity consumption and wind and PV generation (both distributed and utility-scale) in France. Maximum consumption is typically reached during winter evenings due to a high reliance on electric heating. Solar output is low during such hours, although generation during daylight hours may still take place during hours with high consumption levels.[19] In contrast, wind output is higher during the winter but output levels fluctuate significantly across hours and days.[20] We discuss in paragraph 7.3 how our results may extrapolate to power systems with different characteristics.

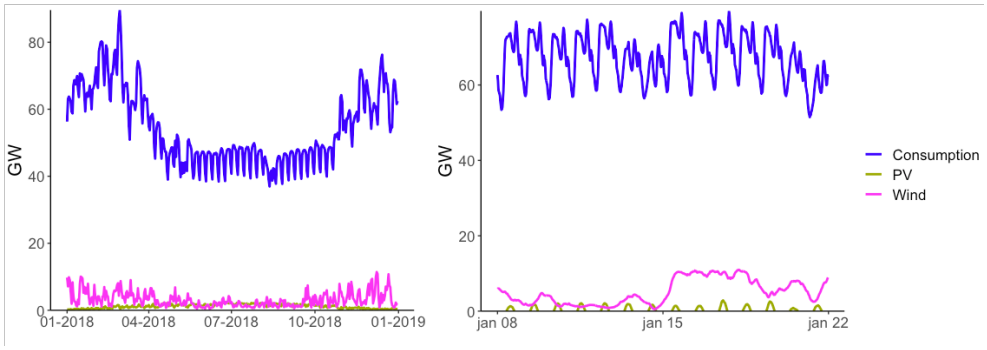


FIGURE 6. Average daily electricity consumption, along with wind and PV generation (GW) in France in 2018 (left panel) ; and half-hourly measurements for two weeks in January 2018 (right panel). Electricity consumption is an estimate by the Transmission System Operator of actual gross consumption, i.e. distributed generation output is not subtracted from it (data source: RTE).

Finally, we observe that the timing of when the maximum hourly net withdrawal is reached can differ across distribution networks. Figure 7 shows the histograms of the month of the year, day of the week and hour of the day when the hourly net withdrawal at a substation reaches its annual peak.

19. For each day in 2018, we can compute the coefficient of correlation (based on publicly available half-hourly measurements at the national level) between gross consumption and PV output. The average of intra-day correlation coefficients was 0.54. Note that the TSO has to perform estimations to compute half-hourly distributed generation output (using a methodology that is not publicly disclosed to the best of our knowledge) and that actual generation at the distribution network level is not observed.

20. The average of intra-day correlation coefficients between hourly gross consumption and wind output (see footnote 19) was -0.11 in 2018.

We restrict attention to 2005 in order to limit the influence of distributed generation. Although the overwhelming majority of substations reach their peak net withdrawal on a working day during winter months (December to March), this peak may occur either during the day (7am to 2pm) or in the evening (after 6pm).

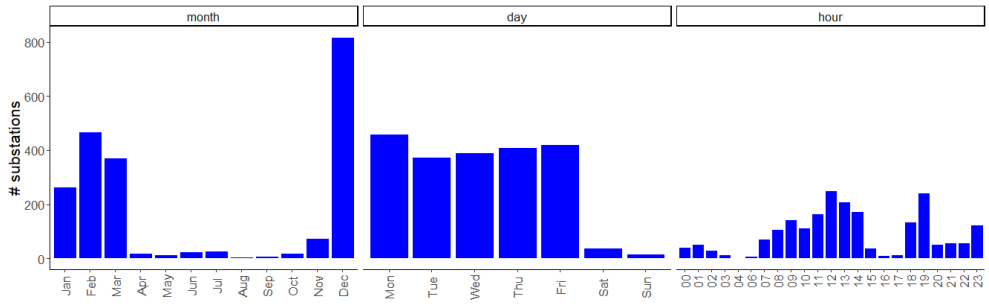


FIGURE 7. Histograms over substations of the month of the year (left), day of the week (middle) and hour of the day (right) when peak hourly net withdrawal was reached in 2005.

4. Data

We observe hourly net withdrawals for 2,216 substations[21] in France (shown in Figure 8) between 1 January 2005 and 31 December 2018.[22] By convention, positive values of net withdrawals correspond to hours where the sum of local consumption and power losses exceeds local generation and negative values of net withdrawals correspond to hours where local generation exceeds the

21. 2,112 substations are observed over the 14-year period. Out of the remaining 114 substations, 90 correspond to substations that were commissioned between 2006 and 2018. We discuss how we account for entry/exit at the end of Appendix A.

22. Figure 8 also depicts in gray in the background all existing or planned substations as of April 2022 (retrieved from https://www.capareseau.fr/). Leaving aside substations that were commissioned between 2018 and 2022, the substations absent from our analysis belong to two main categories. First, the main DSO in France supplies 95% of end-consumers, and we do not observe net withdrawals at substations supplying smaller DSOs (see for example the region around Strasbourg in the North-East). Second, some substations do not supply a distribution network but instead a single end-consumer (e.g. high-speed train railroads). These substations are excluded from the analysis.

sum of local consumption and power losses. We complement this dataset with publicly available information on distributed generation installations compiled as described below.

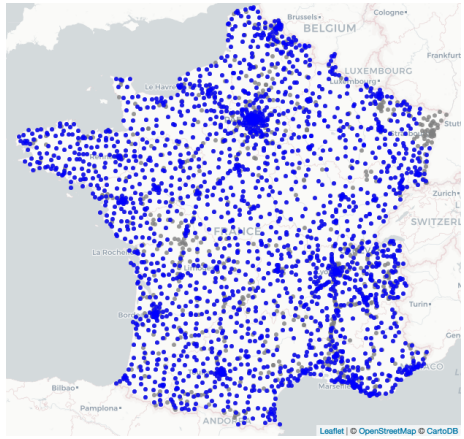


FIGURE 8. Location of the distribution substations for which we observe hourly net withdrawals (blue). All existing or planned substations as of April 2022 are plotted in gray in the background.

4.1. Substation Hourly Net Withdrawals

From our raw dataset of hourly net withdrawals for each substation (about 250 million observations), we can compute summary statistics for the distribution of hourly net withdrawals from a given distribution network in a given year. Appendix C provides detailed information on a number of these summary statistics. In particular, we observe that the most prominent changes that occurred between 2005 and 2018 relate to what is known as "reverse power flows", defined as hours during which local generation exceeds local consumption (i.e. hours with negative net withdrawals). In particular, the fraction of substations that have experienced at least one hour of reverse power flows in a given year has increased from about 6% in 2005 to more than 25% in 2018. In other words, over a quarter of substations now experience hours during which electricity is flowing from the distribution grid to the transmission grid, that is in the opposite direction from the historical pattern shown in the left

panel of Figure 1. In addition, the fraction of substations for which peak usage (in absolute value) was reached during an hour when electricity was flowing from the distribution grid to the transmission grid has increased from under 1% in 2005 to almost 9% in 2018.

As described in Section 5, our empirical strategy uses a given substation in a given year as the unit of observation. For each substation in each year, we observe a time series of hourly net withdrawals. Each time series can be used to build both a load duration curve and a ramp duration curve. We thus observe a load/ramp duration curve for each substation in each year. We keep track of these curves by extracting the 1st, 10th, 25th, 50th, 75th, 90th and 99th percentiles of the annual distribution of hourly net withdrawals (resp. hourly ramps). Our final dataset on observed net withdrawals from distribution networks thus consists of 14 panel variables (7 percentiles for both the load and ramp duration curves) that use substation-year as the unit of observation.

4.2. Distributed Generation Capacity

Information on distributed generation units is extracted from a public inventory which provides detailed data on the universe of power plants in France.[23] As of 31 December 2018, this inventory consisted of 44,000+ observations, out of which 42,000+ referred to installations located in mainland France and connected to the distribution grid. These sites range from a few kW to 50 MW. With a negligible number of exceptions,[24] distributed generation units

23. 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.

24. These exceptions are (i) 1 geothermal unit located in a county that is unlikely to be supplied by one of the substations we observe ; (ii) 2 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 were able to infer the technology of 33 out these 51 units based on the fuel used, their name or an internet search of their characteristics (name, location, etc.).

belong to one of the five following categories: wind, solar,[25] small hydro, and thermal units burning either renewable (e.g. biomass) or non-renewable (e.g. gas) fuel. We also observe the installed capacities[26] and commissioning dates[27] of distributed generation units.

Most observations (28,000+) correspond to distributed generation units listed individually. For the vast majority of these units (26,000+), we observe the identifier of the upstream substation to which they connect. We can thus match accurately these installations to distribution substations. The remaining 14,000+ distributed generation observations listed in the inventory however do not correspond to individual units. For privacy reasons, 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 inventory, they add up to a much lower share of total capacity given their small unit-level size (Table 1).

---

25. 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".

26. 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 these concepts are often used interchangeably when entering data into the inventory. For observations that provide both installed and connection capacities, 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 capacity information, but do provide another capacity metric that we use as a proxy. Finally, 141 installations report installed and connection capacities that differ by more than 10%. For these units, we compute the capacity factor implied by their annual energy production (when available) and choose the capacity metric that implies the most credible capacity factor. We extrapolate this choice to similar units when annual energy production is not available.

27. The inventory makes a distinction between the date at which a unit is commissioned and the date at which its grid connection is completed. 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 the commissioning date and the connection date is likely to correspond to a ramping up period during which the unit does not produce at full capacity.

TABLE 1. Installed capacities of distributed generation (MW) as of 31 December 2018 in mainland France by technology and availability of upstream substation information.

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

Note: The last column computes, for each technology, the percentage of total installed capacities for which upstream substation information is directly observed.

For distributed generation installations whose upstream substation is unknown, we design and implement an assignment procedure to infer the substation to which they are most likely to connect. This procedure leverages 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 larger than the number of distribution substations.[28] This very fine spatial granularity allows us to form reasonable guesses about the substation that 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 not directly observed. In other words, our identifying variation primarily comes from distributed generation units for which we do observe the upstream substation (see Appendix B). These installations are usually installed by specialized project developers. Location choices are thus mainly driven by factors such as resource availability (which we control for using substation fixed effects) or policy incentives which are designed at the national level (which we control for using year fixed effects). Figure 9 shows, for each technology, the histograms of

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

the installed capacity of the corresponding units. For non renewable thermal, the median installation size is 3.4 MW. Most installations consist in combined heat and power units. This is also the case for renewable thermal installations, which however tend to be smaller (median size of 500 kW). For PV, the median unit has a capacity of 90 kW, that is about 300 solar panels. Small hydro overwhelmingly refers to run-of-the-river units, with a median size of 400 kW. Most wind units have a size of about 10 MW, corresponding to small installations consisting in 4 to 6 wind turbines.

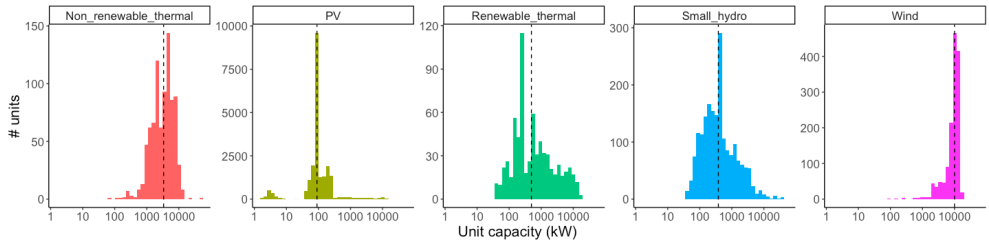


FIGURE 9. Histograms the installed capacities (kW) of distributed generation units (by technology) for which we directly observe the upstream substation. The x-axis is in log10 scale and the vertical dashed line materializes the median of the distribution.

As of 2018, about 90% of total installed wind and PV capacities in France were distributed. This share is higher than in many other developed countries.[29] Because wind and PV installations were not cost-competitive during most of our sample period (relative to say gas-fired power plants), investments in these installations were largely driven by public policies. Support schemes were defined at the national level[30] and have typically consisted of infrequent auctions with limited volume for large installations, and of feed-in-tariffs without volume limitations for small installations. Because early subsidies have

29. For example, distributed PV in the U.S. amounts to less than 40% of total PV capacity (EIA, 2021).

30. See for example www.cre.fr/Transition-energetique-et-innovation-technologique/soutien-a-la-production/dispositifs-de-soutien-aux-enr.

turned out to be fairly generous, this structure has largely favored distributed installations over their utility-scale counterparts. As an illustration, the atom at 100 kW for the size of PV installations on Figure 9 materializes the threshold above which PV installations had to win an auction to benefit from a support scheme.

TABLE 2. Installed capacities of distributed generation at the substation-year level.

Statistic	Mean (MW)	St. Dev. (MW)	Min (MW)	Pctl(25) (MW)	Pctl(75) (MW)	Max (MW)	Total 2018 (% inventory)
Wind	3.11	11.50	0	0	0	189	13,567 (96.8%)
PV	1.39	3.83	0	0.01	1.2	101	7,695 (99.0%)
Small hydro	0.64	2.64	0	0	0	63	1,717 (86.1%)
Renewable thermal	0.35	1.71	0	0	0	35	1,198 (96.0%)
Non renewable thermal	0.97	2.74	0	0	0	45	3,334 (93.3%)

First columns: summary statistics of substation level installed capacities 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 in France.

Our dataset on distributed generation ultimately keeps track, for every distribution substation and each year, of the installed capacities connected to this substation (as of 31 December),[31] broken down by technology. Because the installed capacities of distributed generation were small in 2005 (Figure 3), many observations are at zero.[32] We nonetheless observe significant variation in the installed capacity of distributed generation at the substation level in each year, as shown in Table 2. In addition, the substations for which we observe hourly net withdrawals are hosting the vast majority of distributed generation installations in mainland France.[33] (see Northeastern region on Figure 8).

31. Note that we keep track of installed capacities at the end of each year. This tends to slightly over-estimate the amount of distributed generation capacity that was installed on average during the year, and therefore may under-estimate our coefficients of interest. Keeping instead track of installed capacities at the beginning of each year would by contrast underestimate installed capacities. This latter approach to measuring installed distributed generation capacity yields very similar results to the ones we report.

32. Although distributed generation investments may be biased towards substations with specific unobserved characteristics, such characteristics will be controlled for by substation fixed effects. Appendix B reports the estimation results for the subset of substations that ultimately host in 2018 either some wind installations or more than 1 MW of PV. The estimation results are qualitatively similar, although obviously noisier.

33. The remaining capacities most likely connect to substations for which we do not observe hourly net withdrawals. For example, a number of small hydro units are located on the

5. Empirical Strategy

We characterize the impact of distributed generation on the net load duration curve (resp. ramp duration curve) through a quantile impact function (Figure 10). This function captures the predicted impact that adding 1 MW of a given type of distributed generation capacity has on individual quantiles of the annual distribution of hourly net withdrawals (resp. hourly ramp rates). For a given technology, the quantile impact function maps each quantile index (from 0 to 1) to the predicted impact (in MW or $MW.h^{-1}$) of adding 1 MW of this technology on the corresponding quantile of the annual distribution of hourly net withdrawals (resp. hourly ramp rates). For the load duration curve, this effect may be interpreted as the predicted amount of electricity produced by a distributed unit of a given technology during the hours corresponding to a given quantile index, expressed as a percentage of nameplate capacity.

Importantly, the estimated impacts on duration curves reflect the combination of two effects. First, distributed generation will have differentiated impacts across hours where exogenous factors (e.g. hour of day, temperature, etc.) induce different levels of gross electricity consumption. Second, because this first effect is heterogeneous across hours, the ranking of hourly net withdrawals will differ from the ranking of hourly gross consumption levels. In other words, the 99th quantile of the distribution of hourly net withdrawals will correspond to an hour with high gross consumption in the absence of distributed generation but, after the addition of distributed generation, the 99th quantile may shift to an hour with a slightly lower gross consumption and a very low distributed generation output. Our empirical strategy thus underestimates the first "direct" effect of distributed generation on reducing peak hourly net withdrawals. However, as discussed in Section 3, system operators ultimately care about

Northern part of the Rhine river, which is one of the few areas where we lack information on substation hourly net withdrawals from distribution networks

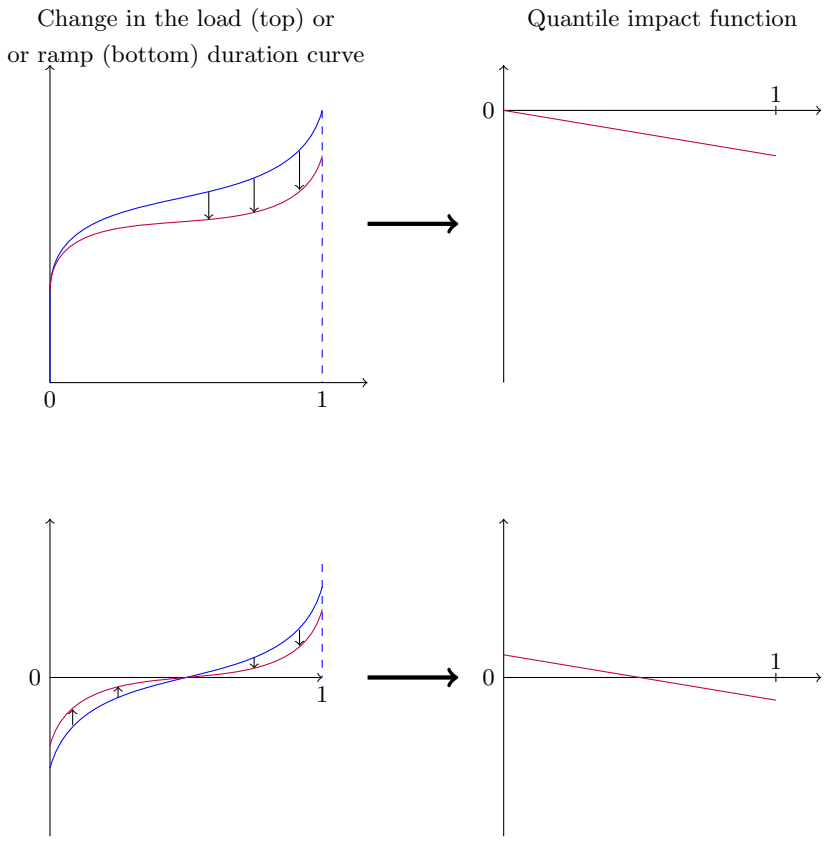


FIGURE 10. Illustration of the intuition behind quantile impact functions for the load duration curve (top) and the ramp duration curve (bottom). Adding 1 MW of a given distributed generation technology changes the duration curve of interest. The quantile impact function captures the difference between the duration curve after vs before a marginal 1 MW increase in distributed generation capacity.

supplying the new net load duration curve after the addition of distributed generation, which results from the combination of both a "direct" effect and an additional "peak shifting" effect. Because we want to assess the potential for distributed generation to defer future grid investments, our empirical strategy estimates the sum of these two effects.

From a forward-looking perspective, estimating the "direct" effect of distributed generation in isolation would be helpful to better understand how changes in the annual distribution of hourly gross consumption may affect

our conclusions. Indeed, if the extreme weather events brought by climate change predominantly occur during hours where gross consumption is currently highest, then the ability of renewables to defer future grid investments would significantly depend on whether they generate electricity during such hours. Unfortunately, this information cannot be inferred from historical patterns of hourly gross demand and renewables production. In addition, we do not observe gross consumption and distributed generation output separately. As a result, disentangling the relative importance of the "direct" effect and the "peak shifting" effect would require further assumptions and is beyond the scope of this paper. We however provide in Appendix F.2 evidence suggesting that the "peak shifting" effect is an important driver of our results.

5.1. Model Specification

We use a seemingly unrelated regressions framework with a two-way fixed effect model to estimate quantile impact functions.[34], justifying our use of a more parsimonious model. Non-constant marginal impacts are discussed in Section 7. Finally, we estimate in Appendix B spatial regression models. In other words, we run the following regression for the main percentiles $q \in \{0.01, 0.1, 0.25, 0.5, 0.75, 0.9, 0.99\}$ of the annual distribution of hourly net withdrawals from distribution networks on the one hand, and the annual distribution of hourly ramp rates on the other hand:

$$Y_{q,s,y} = \sum_t \beta_{q,t} K_{t,s,y} + \delta_{q,s} + \delta_{q,y} + \varepsilon_{q,s,y} \tag{1}$$

where $Y_{q,s,y}$ denotes the q-th quantile of either the annual distribution of hourly net withdrawals (in MW) or the annual distribution of hourly ramp rates (in MW.h^{-1}) for substation s in year y, $K_{t,s,y}$ the installed capacity

34. We discuss in Appendix B the results from estimating alternative specifications. In particular, neither adding covariates interacting the installed capacities of different distributed technologies nor including region-by-year fixed effects were not found to change our results

of distributed generation technology t connected to substation s as of year y (in MW), and $\delta_{q,s}$ and $\delta_{q,y}$ are respectively substation and year fixed effects of the regression for quantile q. 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. For a given technology t, 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 MW or MW.h^{-1}) that adding 1 MW of technology t has on the q-th quantile of the distribution of interest. For example, for the load duration curve, $\hat{\beta}_{0.5,PV} = -0.2$ means that adding 1 MW of distributed PV generation decreases on average the median ($q = 0.5$) hourly net withdrawals at distribution substations by 0.2 MW. Although our discussion of results will mostly focus on wind and PV, which have the broadest policy implications, including distributed thermal and hydro capacities in our model enhances the credibility of these results. In particular, we expect investments in thermal technologies, which can generate electricity on demand, to reduce peak net withdrawals from the distribution grid.

5.2. Determinants of Distributed Generation Investments

Over our period of interest, support mechanisms for distributed generation in France were defined nationally and did not provide incentives to develop projects in specific locations. In addition, grid connection charges in France are only mildly differentiated at the level of administrative regions, and these regions aggregate on average about two hundred distribution systems (de Lagarde, 2018). This means that grid connection charges do not provide significant locational incentives at the substation level for distributed generation investments. Project developers can face additional delays or costs if the distribution system operator deems that grid hosting capacity is scarce, but this information is not known *ex ante* by project developers and not publicly

available.[35] If anything, any unobserved locational incentive of this sort would induce our analysis to conclude that distributed generation is more grid-friendly than it would be in the absence of any intervention by the system operator. Finally, the electricity tariffs faced by small consumer are not linked in any way to local distributed generation output.[36]

Land and resource availability considerations do create incentives to locate distributed generation units non-uniformly across space. We control for this factor and any other unobserved factor that varies across distribution substations but not across years with substation fixed effects.[37] Our model specification also controls for national trends such as load growth and differences over time in distributed generation support mechanisms with year-of-sample fixed effects.

Although distributed PV is arguably the most challenging technology for a causal interpretation of our results, we however believe that our approach provides credible estimates even for this technology. Indeed, our identifying variation primarily comes from units for which we directly observe the upstream substation, which means that our identification does not primarily rely on residential PV capacities.[38] These PV installations tend to be relatively large and, although the vast majority of them connect to the low voltage network, installations connected to the medium voltage network represent 63% of their total capacity. In the absence of spatially differentiated investment incentives,

35. From 2015 onward, the website https://www.capareseau.fr/ made this information publicly available. However, it takes a few years to develop a distributed generation project so that this change is very unlikely to affect our results given our sample period.

36. The largest industrial consumers, some of which face dynamic prices that may correlate with renewable output through the wholesale market, usually connect to the grid at higher voltage levels.

37. We also show in Appendix B that our results are robust when we restrict attention to substations where investments in wind or PV projects actually occurred. In addition, we find that our results are robust to estimating a richer spatial regression model.

38. Our estimates however speak to the deployment of residential PV systems since, from the point of view of a substation, and in the absence of a strong rebound effect, a thousand 3 kW PV systems are observationally equivalent to a single 3 MW system.

the development of ground-level installations seems in turn unlikely to depend on local trends such as population or economic growth.

As robustness checks, we show in Appendix B that we obtain similar results when we replace year-of-sample fixed effects with a larger set of fixed effects that interact the year of sample with a spatial unit coarser than substations (regions or departments). We also run a placebo test applying our model to night-time hours (11pm-5am) where PV output is known to be zero. We find no evidence of a bias for our analysis of hourly ramps, and a small bias for our analysis of hourly net withdrawals, whose magnitude would not affect our main conclusions. In addition, this bias is consistent with the behavioral response or "solar rebound" from the installation of rooftop PV units identified in Qiu et al. (2019) for residential PV. The combined effect of distributed PV generation and of the solar rebound would then be the most relevant metric to inform future grid planning.

Finally, our specification assumes linear and constant "treatment effects". We discuss in Section 7 possible non-linearities in the impact of installed capacities on the load and ramp duration curves.[39]

6. Main Results

We first estimate the impact of the different distributed generation technologies on the quantiles of the annual distribution of net withdrawals from distribution networks. Table 3 shows our estimation results, which are also represented graphically on Figure 11. We find substantially different quantile impact functions for the five distributed generation technologies. Two characteristics of these functions are of particular interest: their value for the extreme quantiles and their slope.

39. Appendix B also reports the results from splitting the sample into an "early" period (2005-2011) and a "late period" (2012-2018). Our results for wind and PV remain qualitatively unchanged.

TABLE 3. Estimated coefficients when regressing the main quantiles of the distribution of hourly net withdrawals (for a given substation in a given year) 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

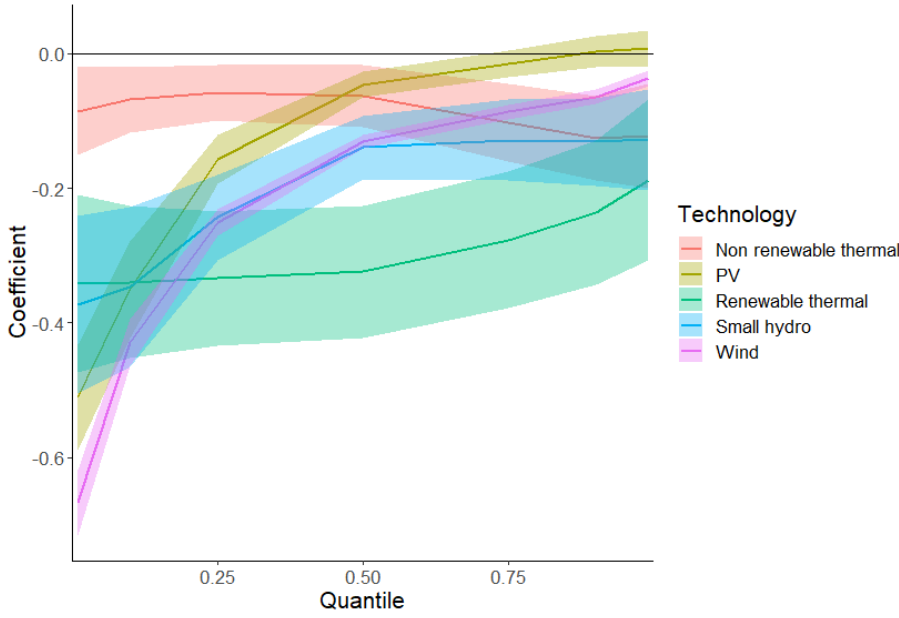


FIGURE 11. Graphical representation of the quantile impact functions for the distribution of hourly net withdrawals. Thick lines correspond to the point estimates. Sleeves delimit (two-sided) 95% confidence intervals from robust standard errors clustered at the substation level.

First, the higher (in absolute value) the coefficients for the highest quantiles, the more a given technology is associated with a decrease in the peak net

withdrawals from distribution networks, and thus the more likely it is to defer or avoid grid expansions. In France, the national annual peak electricity consumption is reached during cold winter evenings. Consistently, PV is found to have no significant impact on the highest quantiles of the annual distribution of hourly net withdrawals. The impact of wind on the top quantiles is also very small even though, as reported in RTE (2019), average wind generation is higher during the winter. The other technologies are found to have a sizable impact on peak hourly net withdrawals, with an average impact on the 99th quantile of at least 0.12 MW per MW of distributed generation.[40] Conversely, large negative impacts on the bottom quantiles are likely to result in reverse power flows as distributed generation capacities increase. The largest reverse power flows may ultimately reach levels comparable (in absolute value) to local peak consumption, potentially compromising any network capacity savings enabled by a decrease in the top quantiles of the distribution of hourly net withdrawals.

Second, whether the quantile impact function is upward or downward sloping is also of particular interest. Indeed, a monotone decreasing quantile impact function means that the corresponding distributed generation technology tends to narrow the range of hourly net withdrawals from distribution networks, increasing the utilization rates of future grid assets. In contrast, a monotone increasing quantile impact function means that the corresponding distributed generation technology tends to "stretch" the distribution of hourly net withdrawals, that is to expand the range of net withdrawals supplied by substations. The subsequent decrease in the utilization rates of grid assets seems in turn likely to be associated with higher long-term costs on a per-MWh basis. Figure 11 suggests that quantile impact functions are monotone increasing for all technologies but non-renewable thermal. We

40. For non-renewable thermal units, this finding appears consistent with the fact that public subsidies provide incentives for small natural gas combined heat and power units to produce during the winter (www.legifrance.gouv.fr/jorf/id/JORFTEXT000033385467/).

test this observation statistically in Appendix D using the testing framework developed by Wolak (1987, 1989) for seemingly unrelated regressions. 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 nonrenewable thermal, and the null hypothesis of a decreasing quantile impact function cannot be rejected at the same level of statistical significance.

TABLE 4. Estimated coefficients when regressing the main quantiles of the distributions of hourly ramps (for a given substation in a given year) 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

Next, we estimate the impact of the different distributed generation technologies on the ramp duration curve faced by substations. Table 4 shows our results, which are represented graphically in Figure 12. Two distinct groups of distributed generation technologies emerge. On the one hand, thermal and small hydro units are found to have a negligible impact on any quantile of the annual distribution of hourly ramps. On the other hand, wind and PV units tend to significantly stretch this distribution. In other words, increased installed capacities of distributed wind and PV are associated with a significant increase in the absolute value of the most extreme hourly ramps, both positive and negative. More precisely, we find that a 1 MW increase in the installed

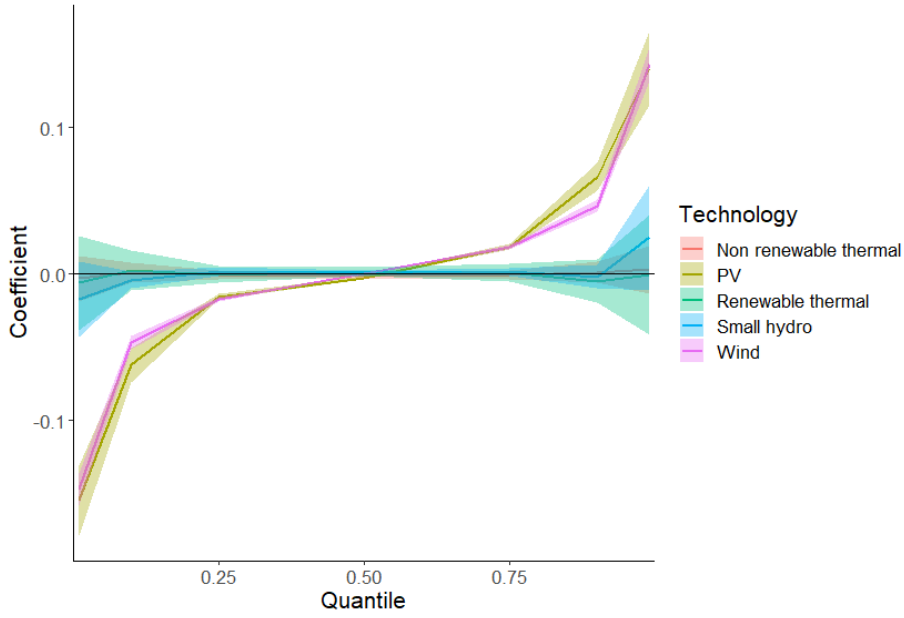


FIGURE 12. Graphical representation of the quantile impact functions for the distribution of hourly ramps. Thick lines correspond to the point estimates. Sleeves delimit (two-sided) 95% confidence intervals from robust standard errors clustered at the substation level.

capacity of either wind or PV is associated with an average increase of 0.14-0.15 MW.h^{-1} in the absolute value of both the 1st and 99th percentiles of the distribution of hourly ramps. Large installed capacities of distributed wind and PV are thus associated with very substantial and rapid variations in hourly net withdrawals from distribution networks. Such increases in the magnitude of extreme hourly ramps may put more stress on network components (e.g. more frequent tap changes for transformers) and make it harder for system operators to meet operational constraints (e.g. voltage regulation or phase balancing). If anything, network costs are more likely to increase rather than decrease as a result of these larger ramps.

7. Discussion

This section explores various reasons why distributed wind and PV may bring higher T&D savings under circumstances that would depart significantly from

the ones underlying our main estimates. We first investigate whether our marginal effects for investments in distributed generation depend on the pre-existing level of these investments.[41] Next we study how the installation of battery storage along with investments in distributed solar and wind could change our estimated marginal effects. Finally, we consider the applicability of our empirical results to other electricity supply industries.

7.1. Non-Linear Quantile Impacts

Although we estimate a distinct regression for each quantile, our specification is linear in installed capacities of distributed generation (cf. Equation (1)). In this section we explore whether the marginal impacts of distributed generation on the duration curves change with the level of installed generation capacities. More specifically, we estimate the following quadratic specification:

$$Y_{q,s,y} = \sum_t \alpha_{q,t} K_{t,s,y} + \sum_t \beta_{q,t} K_{t,s,y}^2 + \delta_{q,s} + \delta_{q,y} + \varepsilon_{q,s,y} \tag{2}$$

where notations are the same as in Equation (1). Tables E.1 and E.2 in Appendix E report the our results. For the load duration curve, we can reject the null hypothesis that the marginal effect of distributed solar and wind capacities on hourly net withdrawals is constant ($\beta_{q,t} = 0$) for some but not all quantiles. This null hypothesis is rejected at different quantiles for the two technologies. PV is found to have a negligible impact on the top quantiles even at low levels of installed capacities, and to have an increasingly negative impact on the bottom quantiles as more capacity gets installed. The former observation is consistent with the fact that demand usually peaks during winter evenings, when PV output is negligible. The latter observation suggests that

41. We also explore the potential interaction effects of investments in the different distributed generation technologies in Appendix B (e.g. complementarity between wind and PV investments). We found that a 0.1 size test of the null hypothesis that the coefficients on the ten interaction terms in each of the seven quantile impact functions are zero could not be rejected.

distributed solar generation is increasingly responsible for the lowest hourly net withdrawals as installed capacities grow.

By contrast, the marginal effect of distributed wind capacities on the bottom quantiles is found to be fairly constant. However, the magnitude of the negative impact of distributed wind capacities on the top quantiles decreases with installed capacities. In particular, we find that the first distributed wind units connecting to a distribution network are predicted to reduce the 99th percentile of the annual distribution of hourly net withdrawals by 7% of their nameplate capacity, almost double the average effect for all capacity investment levels. In other words, the marginal benefit of distributed wind in terms of potential grid deferrals is decreasing in the installed capacity of distributed wind generation. For the ramp duration curve, we do not find evidence of a non-linear impact of distributed wind on the quantiles of the annual distribution of hourly ramps. By contrast, distributed PV is found to have a marginal impact on the extreme ramp rates that increases (in absolute value) with installed capacities. In other words, distributed PV capacities are found to amplify the magnitude of extreme hourly ramps at an increasing rate.

7.2. Battery Storage

This section explores the extent to which battery storage may enhance the grid deferral benefits from distributed wind and solar generation. More precisely, we assume that for each kW of distributed wind or solar generation connected to a substation in a given year, a proportional amount of storage is also installed. We then simulate the operation of the batteries and re-estimate the marginal impact of distributed generation investments on the resulting load duration curve.[42] In other words, battery charging and discharging alter the

42. In other contexts, such as wildfire events in California, battery storage is often argued to provide "reliability benefits" to consumers. These consideration are not as relevant in our context. First, most distributed generation capacity come from installations connected to the medium voltage grid rather than installed at residential consumers' premises. Second,

left hand-side variable in our regressions by subtracting, prior to computing duration curves, the hourly operations of batteries from actually observed net withdrawals during each hour of the year. We assume the batteries can charge both from the grid and local distributed generation. Conceptually, this stylized exercise is consistent with a mandate requiring the installation of storage capacity along with intermittent renewables so that the installed capacity of storage grows in proportion to investments in intermittent renewable generation.

We assume that batteries are operated in a weekly peak-shaving mode. In other words, the storage operator aims at minimizing either weekly peak net withdrawal or peak net injection, depending on which metric is initially the highest in absolute value over the course of the week. We use the peak-shaving algorithm described in Pimm et al. (2018) with two modifications. First, in order to ensure energy conservation, we add the constraint that storage is half-full at the beginning of the week and must finish the week at the same state-of-charge. Because substation peak net demand is rarely reached at midnight on a Sunday evening (see Figure 7), this constraint is unlikely to bias significantly battery operations. Second, when the battery has enough idle time to fully charge, we assume that charging is spread uniformly across available hours rather than occurring at its maximum rate until full charge. Figure 13 illustrates how battery operations change the time series of net hourly loads over four consecutive weeks at a given substation.

Assuming that storage is operated in peak-shaving mode is arguably an optimistic approach because storage owners may choose different and less grid-friendly operating rules, as for example in Green and Staffell (2017). We make

consumers connected to the low-voltage grid are disconnected about 1 hour/year on average, so that potential reliability benefits seem likely to be smaller than in California. Third, a residential PV + battery system will provide sizable reliability benefits only if PV generates significant amounts of electricity during the most severe outages. Such outages tend to occur in France during cold winter spells, that is when PV output is at its lowest.

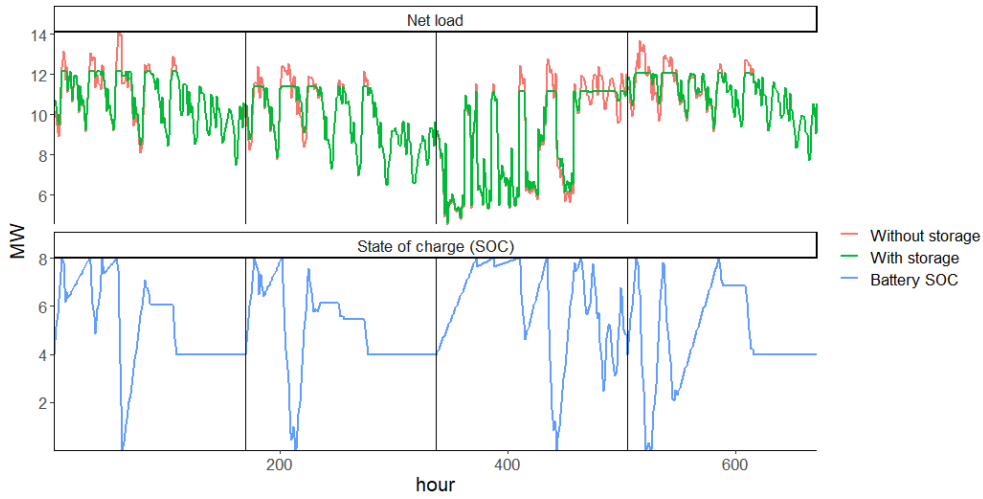


FIGURE 13. Hourly net withdrawals from a given substation over four weeks before and after the addition of 8 MWh of storage capacity with a maximum discharge rate of 4 MW (top panel). The amount of energy available in the battery, known as state of charge (SOC), is also displayed (bottom panel).

three additional assumptions in favor of the effectiveness of batteries at reducing the highest quantiles of the annual distribution of net demands. First, we assume that the storage operator has a perfect foresight of substation hourly net withdrawals for the upcoming week. Second, we neglect for simplicity power losses when charging/discharging electricity. Third, we do not model battery degradation or the opportunity cost of having a limited number of available refresh cycles. Overall, our aim is to get a sense of the upper bound of the potential grid deferral benefits from combining distributed wind and solar with battery storage. Operating batteries in a peak-shaving mode seems likely to be the most critical assumption.[43] The perfect forecast assumption is also quite optimistic. Building a credible hourly net load forecast model is however beyond the scope of this paper. Relaxing the other assumptions seems unlikely to drastically change our results: charge/discharge efficiencies are in the 80-90%

43. For example, running a similar analysis with storage responding to a time-of-use tariff, we ended up in situations where distributed generation bundled with storage significantly *increased* the top quantiles of the distribution of hourly net withdrawals.

range, and peak-shaving would in practice be needed during a limited number of hours to shave only the annual peak of hourly net withdrawals.

Battery storage "capacity" is generally understood as the maximum energy E (in kWh) that can be stored,[44] rather than the maximum (dis)charge rate P (in kW).[45] The charge rate or "C-rate" (in h^{-1}) is the ratio of P and E. It captures how fast a battery can be (dis)charged. For example, it takes 2 hours to fully charge an initially empty battery whose C-rate is 0.5. To fix ideas, the Tesla Powerwall 2 can store up to 13.5 kWh with a maximum (dis)charge rate of 7 kW, corresponding to a C-rate of 0.52.

We define "storage penetration" as the ratio of installed storage capacity (in kWh) and installed distributed generation (in kW). For example, a storage penetration of 50% means that for each kW of distributed wind/solar generation connected to a given substation in a given year, we assume that 0.5 kWh of storage is also installed. We explore storage penetration levels ranging from 50% to 500%. Using figures from Tesla products to fix ideas, a single solar panel has a capacity of about 340 W. As a result, a 500% penetration would correspond to a situation where a Tesla Powerwall 2 is installed for every 8 solar panels connected to the grid.[46] Battery storage sales in California in 2020 were expected to be 50,000 systems[47] along with expected investments in about 1.1 GW of residential rooftop PV for that year.[48] This suggests that the

---

44. Note that for other storage technologies, and notably reservoir or pumped hydro, "capacity" is defined as the maximum (dis)charge rate rather than as the maximum energy that can be stored.

45. We assume for simplicity that the maximum charge and discharge rates are equal although this need not be the case in practice.

46. Note that 8 solar panels amount to about 3 kW, which is in the lower end of residential rooftop PV systems in terms of installed capacity.

47. https://about.bnef.com/blog/california-household-battery-sales-to-quadruple-in-2020/

48. www.seia.org/sites/default/files/2021-03/California.pdf

current battery storage penetration for new residential PV installations is 60% assuming each new battery storage system is a Tesla Powerwall 2.[49]

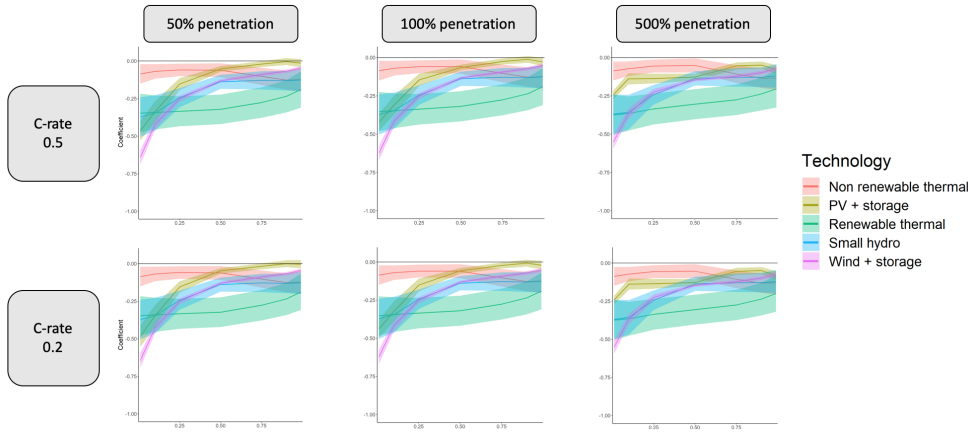


FIGURE 14. Quantile impact functions for the distribution of hourly net withdrawals under the different storage scenarios.

TABLE 5. Estimated coefficients for the 90th and 99th percentiles for wind and PV (bundled with storage investments) under the different storage scenarios. Robust standard errors clustered at the substation level are reported.

	0.5C - 50%	0.5C - 100%	0.5C - 500%	0.2C-50%	0.2C-100%	0.2C-500%
	Dependent variable: Q90					
PV	-0.004	-0.012	-0.049	0.000	-0.007	-0.049
	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.012)
Wind	-0.068	-0.072	-0.099	-0.067	-0.072	-0.099
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
	Dependent variable: Q99					
PV	-0.014	-0.029	-0.079	-0.004	-0.020	-0.082
	(0.014)	(0.014)	(0.015)	(0.014)	(0.014)	(0.015)
Wind	-0.046	-0.051	-0.074	-0.045	-0.052	-0.076
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)

Figure 14 shows the quantile impact functions for the distribution of hourly net withdrawals under different storage investment scenarios. The corresponding estimates for the highest quantiles are reported in Table 5. We find that very substantial levels of storage penetration would be needed to decrease significantly the highest percentiles of the load duration curve. At 50% penetration, adding storage to distributed wind and solar investments barely

49. Note that 60% ≈ (50,000 batteries x 13.5 kWh/battery)/(1,100,000 kW of new solar PV).

increases the magnitude of their cumulative impact on the highest quantiles, and mildly attenuates their impact on the lowest quantiles. By contrast, at 500% penetration of battery storage, distributed wind and PV with storage are able to decrease significantly substation peak hourly net withdrawals, and would thus prove helpful to defer future grid expansions.

To understand the intuition behind our results, one should note that battery storage systems can operate at full (dis)charge rate only for a limited amount of time. For example, a fully charged 0.5C battery can only discharge at full power during two consecutive hours before ending up empty. As a result, if net withdrawals remain high for a significant number of consecutive hours, battery storage has a limited ability to reduce the magnitude of the highest net withdrawals. For example, imagine that net withdrawals are flat during ten consecutive hours. Then, under the 100% storage penetration, the best the batteries can do is to reduce net withdrawals by 5% of the nameplate capacity of distributed generation. Indeed, 1 kWh of battery storage gets by assumption installed for each kW of distributed wind/PV which, for a C-rate of 0.5C, correspond to a maximum discharge power of 0.5 kW. Because this capacity has to be spread over ten hours, battery storage can only decrease net withdrawal by 0.05 kW per kW of distributed generation.

The left panel on Figure 15 shows the histogram of the maximum number of consecutive hours during which hourly net withdrawals at a given substation in a given year are higher than the 99^{th} percentile of the distribution. The median value is 10 hours, which is consistent with the results presented in Table 5. This median value decreases to 3 hours when looking at the 999^{th} quantile. In other words, battery storage is more effective at shaving net peak demand when focusing on a more limited number of hours than the 99^{th} quantile.

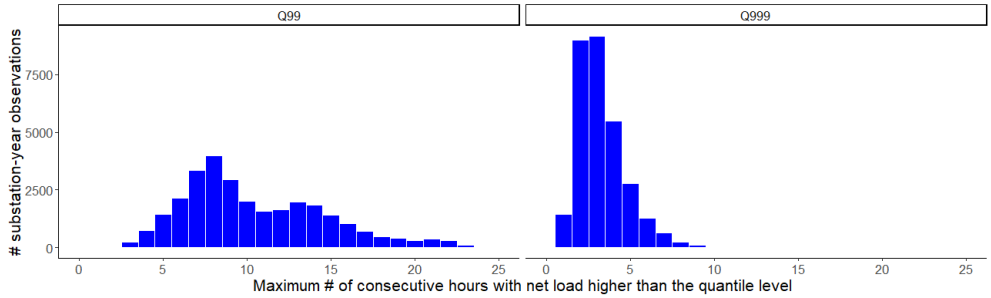


FIGURE 15. Histograms of the maximum number of consecutive hours during which net load is higher than the 99th quantile (left panel) and the 999th quantile (right panel). The unit of observation is the distribution of hourly net withdrawals (before storage operation) at a given substation in a given year (for more clarity, the right tail is censored on the graph).

7.3. External Validity

Because our data covers the vast majority of both distribution substations and distributed generation installations in France, our results provide a comprehensive assessment of the impact that investments in different distributed generation technologies have had on average on hourly net withdrawals and ramp rates throughout mainland France. Appendix F takes advantage of the sample size of our cross-section to explore the extent to which the impact of distributed generation on the grid is heterogeneous across space. More precisely, we divide France into four "macro regions" (North-West, North-East, South-West and South-East) and run our main model separately for each region. We find qualitatively similar results with two caveats. First, estimates are noisier or imprecise in regions where a given technology is not sufficiently deployed (e.g. PV in the North-East). Second, consistently with the results of paragraph 7.1, wind has a larger impact on the highest quantiles of the distribution of hourly net withdrawals in the South, where its installed capacity is lower than in the North.

Whether our results carry over to other countries is an open question. One potential issue is that France has its annual peak electricity consumption in the winter due to its high reliance on electric space heating. By contrast, a number

of countries and U.S. regions experience their peak electricity consumption in the summer, due to a large demand for cooling. There are a number of reasons to believe that our results would carry over to a significant extent to these electricity systems. Although distributed solar will of course have a non-zero impact on the highest quantiles of the annual distribution of hourly net withdrawals in summer-peak systems, the magnitude of this impact may prove smaller than one might expect, especially once significant amounts of distribution generation capacity are installed. In the case of France, we indeed find that distributed wind has a small impact on the highest quantiles of the distribution of hourly net withdrawals even though wind output is highest during the winter. In addition, the impact of distributed wind investments on peak hourly net withdrawals is found to decrease in installed capacity. This suggests that our results largely stem from the intermittent nature and the high level of contemporaneous correlation between the output of distributed generation units, which imply that, in the absence of storage, some hours with high demand and low distributed generation output persist even with high levels of distributed generation investments. Peak net load then mechanically shifts to hours when distributed generation output is small. Appendix F discusses this point in more detail. Finally, it is worth noting that increasing the role of electricity for space heating is one of the key policies to reduce greenhouse gases emissions. As a result, a growing number of electricity systems may experience peak-consumption events during the winter, as illustrated by the electricity crisis of February 2021 in Texas.

8. Conclusion

Public policies supporting renewable electricity generation typically favor distributed units over utility-scale installations despite their significantly higher levelized cost of generating electricity. As a result of these policies, among other incentives, about half of the annual $100+billion global investments in PV

electricity generation over the past decade were directed towards distributed solar units rather than utility-scale installations. This paper presents empirical evidence suggesting that, at least for the case of France, potential savings in future T&D network investments cannot rationalize favoring distributed wind and solar generation units over their utility-scale counterparts.

We estimate the relationship between investments in five distributed generation technologies (wind, PV, small hydro, renewable thermal and non-renewable thermal) and the hourly net withdrawals of electricity by distribution networks. We study the case of France, where distributed generation capacity has grown significantly over the past two decades. Our analysis combines two very comprehensive datasets. First, we observe the hourly net withdrawals at over 2,000 distribution substations between 2005 and 2018. Second, we use detailed information on the universe of electricity generation units to determine how much capacity of each distributed generation technology was connected to each substation in each year. We then use a seemingly unrelated regressions framework with a two-way fixed effect specification to estimate the impact of a marginal increase in the installed capacity of the different distributed generation technologies on both the hourly net withdrawals from distribution networks, and the hourly ramp rates these networks experience.

Distributed wind and PV investments are found (i) to have little to no impact on the highest quantiles of the distribution of hourly net withdrawals, (ii) to induce large downward shifts in the lowest quantiles of the distribution of hourly net withdrawals, and (iii) to exacerbate the magnitude of the most extreme ramps. However, we show that bundling investments in distributed wind and PV with investments in battery storage has the potential to increase significantly their grid deferral benefits, although storage investment rates much larger than current rates are required.

The main policy implication that emerges from our results is that support mechanisms that provide unconditionally higher subsidies to distributed wind

and solar installations over their utility-scale counterparts are very unlikely to represent the least-cost path towards a low-carbon electricity supply. Relying instead on installation-size-neutral subsidies would represent a more cost-effective approach. Such leveled policy incentives would not necessarily prevent the development of distributed generation in the long run. Indeed, utility-scale installations require large areas of land, whose opportunity cost and/or distance to the existing transmission grid will increase as more installations develop. Ultimately, such an increase in costs will lower the LCOE gap between utility-scale and distributed generation installations. In addition, battery storage could improve the business case for distributed generation installations. However, very significant investments and a high level of coordination with distribution system operators would be needed for battery storage to unlock substantial benefits from deferring T&D investments.

Acknowledgements

We are very grateful to the French energy regulator, the Commission de régulation de l'énergie (CRE), for granting us access to the substation hourly net withdrawals dataset 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.

References

Bivand, Roger S., Edzer Pebesma, and Virgilio Gomez-Rubio, *Applied spatial data analysis with R, Second edition*, Springer, NY, 2013.

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/.

Brown, Richard E, Jiuping Pan, Xiaorning Feng, and Krassimir Koutlev, "Siting distributed generation to defer T&D expansion," in "2001 IEEE/PES Transmission and Distribution Conference and Exposition. Developing New Perspectives (Cat. No. 01CH37294)," Vol. 2 IEEE 2001, pp. 622–627.

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.

CRE, "Délibération de la CRE du 21 janvier 2021 portant décision sur le tarif d'utilisation des réseaux publics de distribution d'électricité (TURPE 6 HTA-BT)," 2021.

Cullen, Joseph, "Measuring the environmental benefits of wind-generated electricity," *American Economic Journal: Economic Policy*, 2013, *5* (4), 107–33.

de Lagarde, Cyril, "Network connection schemes for renewable energy in France: a spatial analysis," *CEEM Working paper*, 2018.

EIA, "Electric Power Annual 2020," Technical Report, U.S. Energy Information Administration 2021.

Feinstein, Charles D, Peter A Morris, and Stephen W Chapel, "Capacity planning under uncertainty: Developing local area strategies for integrating distributed resources," *The Energy Journal*, 1997, *18* (Special Issue).

Fell, Harrison, Daniel T Kaffine, and Kevin Novan, "Emissions, transmission, and the environmental value of renewable energy," *American Economic Journal: Economic Policy*, 2021.

General Electric Power, "The Distributed Energy Transformation," Technical Report 2018.

Gil, Hugo A and Geza Joos, "On the quantification of the network capacity deferral value of distributed generation," *IEEE Transactions on Power Systems*, 2006, *21* (4), 1592–1599.

Gillingham, Kenneth T and Bryan Bollinger, "Social learning and solar photovoltaic adoption," *Management Science*, 2021.

Gowrisankaran, Gautam, Stanley S Reynolds, and Mario Samano, "Intermittency and the value of renewable energy," *Journal of Political Economy*, 2016, *124* (4), 1187–1234.

Green, Richard and Iain Staffell, ""Prosumage" and the British electricity market," *Economics of Energy & Environmental Policy*, 2017, *6* (1), 33–50.

Groote, Olivier De and Frank Verboven, "Subsidies and time discounting in new technology adoption: Evidence from solar photovoltaic systems," *American Economic Review*, 2019, *109* (6), 2137–72.

Hoff, Thomas E, "Identifying distributed generation and demand side management investment opportunities," *The Energy Journal*, 1996, *17* (4).

International Energy Agency, "Distributed Generation in Liberalized Electricity Markets," 2002.

IRENA, "Renewable Power Generation Costs in 2020," Technical Report, International Renewable Energy Agency, Abu Dhabi 2021.

Kirschen, Daniel S and Goran Strbac, *Fundamentals of power system economics*, John Wiley & Sons, 2018.

Mendez, VH, J Rivier, JI De La Fuente, T Gomez, J Arceluz, J Marin, and A Madurga, "Impact of distributed generation on distribution investment deferral," *International Journal of Electrical Power & Energy Systems*, 2006, *28* (4), 244–252.

Millo, Giovanni and Gianfranco Piras, "splm: Spatial Panel Data Models in R," *Journal of Statistical Software*, 2012, *47* (1), 1–38.

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/.

Novan, Kevin, "Valuing the wind: renewable energy policies and air pollution avoided," *American Economic Journal: Economic Policy*, 2015, *7* (3), 291–326.

Ovaere, Marten, Bryan Bolinger, and Kenneth Gillingham, "The Value of Distributed Solar: Evidence from a Field Experiment," 2020.

Piccolo, Antonio and Pierluigi Siano, "Evaluating the impact of network investment deferral on distributed generation expansion," *IEEE Transactions on Power Systems*, 2009, *24* (3), 1559–1567.

Pimm, Andrew J, Tim T Cockerill, and Peter G Taylor, "The potential for peak shaving on low voltage distribution networks using electricity storage," *Journal of Energy Storage*, 2018, *16*, 231–242.

Qiu, Yueming, Matthew E Kahn, and Bo Xing, "Quantifying the rebound effects of residential solar panel adoption," *Journal of Environmental Economics and Management*, 2019, *96*, 310–341.

RTE, "Electricity report 2018," Technical Report 2019.

Sexton, Steven, A Justin Kirkpatrick, Robert I Harris, and Nicholas Z Muller, "Heterogeneous Solar Capacity Benefits, Appropriability, and the Costs of Suboptimal Siting," *Journal of the Association of Environmental and Resource Economists*, 2021.

U.S. Department of Energy, "The potential benefits of distributed generation and rate-related issues that may impede its expansion," 2007.

_ , "United States Electricity Industry Primer," Technical Report 2015.

Wang, David T-C, Luis F Ochoa, and Gareth P Harrison, "DG impact on investment deferral: Network planning and security of supply," *IEEE Transactions on Power Systems*, 2009, *25* (2), 1134–1141.

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.

_ , "Level versus Variability Trade-offs in Wind and Solar Generation

Investments: The Case of California," *The Energy Journal*, 2016, *37* (Bollino-Madlener Special Issue).

_ , "Long-term resource adequacy in wholesale electricity markets with significant intermittent renewables," *Environmental and Energy Policy and the Economy*, 2022, *3* (1), 155–220.

For Online Publication - Appendices

Appendix A: Assignment Procedure for Distributed Generation Units whose Upstream Substation is Unknown

This Appendix provides further information on the assignment procedure we implement in order to infer to which substation distributed generation units whose upstream substation is unknown are most likely to connect. This procedure largely relies on very detailed spatial information. Mainland France is divided into over 30,000 administrative counties.[50] In addition, counties with a high population (all counties with more than 10,000 inhabitants and most counties with more than 5,000 inhabitants) are further broken down into sub-counties (called the "IRIS mesh") for census purposes. As of 2019, 1,840 counties in mainland France were 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.[51] 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 unit 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 area of 11.9 km^2 (4.6 miles squared).

The flow chart of Figure A.1 summarizes the different steps of our assignment procedure. Because we observe the capacity, commissioning date, and (sub)county of the distributed generation units that are listed individually, we can directly compute the time series of installed capacities at the

50. 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.

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

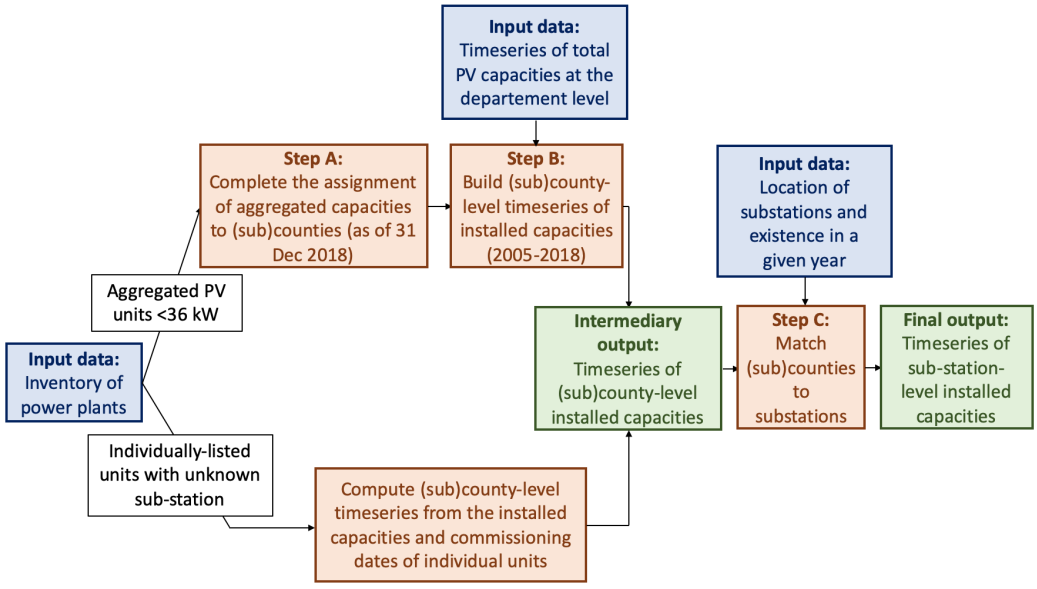


FIGURE A.1. Flow chart of the assignment procedure we implement for distributed generation units for which upstream substation information is missing.

(sub)county level for these units.[52] By contrast, observations consisting in aggregated PV units raise two challenges,[53] 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, in contrast to individually-listed units for which installed capacity usually 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

52. For 71 individually-listed installations with unknown upstream substation, we observe the county but not the sub-county where they are located. These installations are assumed to be equally likely to be located in the different sub-counties of the county.

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

In order to respect the privacy of individual owners, most small (<36 kW) 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: department, county and sub-county.[54] Aggregated PV observations are built as follows. First, any sub-county that hosts more than 10 installations is listed as an observation, whose capacity is the sum of the capacities 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 aggregated at the department level. Mainland France has 94 such departments.

Despite this aggregation procedure, the location of the majority (74%) of aggregated PV capacities is observed 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 that 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 know the department in which these units are located. Given the aggregation procedure used to build the inventory of power plants, we further know that these units can only be located in (sub)counties where none of the other 74% capacities are located. We thus split the capacity aggregated at the department level uniformly across the pool of candidate counties where the corresponding

54. 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 departments. These observations however add up to less than 1 MW and are thus neglected.

units may be located.[55] 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 the observations aggregated at the department level to (sub)counties would require additional modeling and/or information sources. These alternative approaches are unlikely to affect significantly our empirical results. For example, allocating capacities aggregated at the department level to the remaining (sub)counties using log-population instead of a uniform weight yields very similar results. The correlation between the (sub)county-level capacities using uniform versus 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 time series 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 department 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 department level of total installed PV capacities between 2006 and 2018 (installed capacities being virtually zero in 2005).[56]

55. 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 department level may then be located in the latter counties, but not in the former counties.

56. This information is published quarterly by the Service des données et études statistiques (e.g. www.statistiques.developpement-durable.gouv.fr/tableau-de-bord-solaire-photovoltaique-quatrieme-trimestre-2018 for the fourth quarter of 2018). We are grateful

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 department-level time series of PV capacity from individually-listed units. Subtracting these time series from the time series of department-level total PV capacity from the DOE dataset yields a department-level time series for PV capacity from aggregated units. Figures A.3 and A.4 show the results for each department, and compare them to the capacities observed in the public inventories of power plants of 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 time series. We further impose monotonicity which is (mildly) violated on only three occasions.

Second, for each year and each department, we need to dispatch the total (department-level) capacity $K_{d,y}$ of 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 department d) such that:

$$\forall y, \forall d, \sum_{c \in d} K_{c,d,y} = K_{d,y} \tag{A.1}$$

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

1. **Homothetic static approach:** this method assumes that the probability of observing a given amount of installed capacity in a given county is

to the Department of Energy for having shared with us the corresponding historical data (updated of as of July 2020).

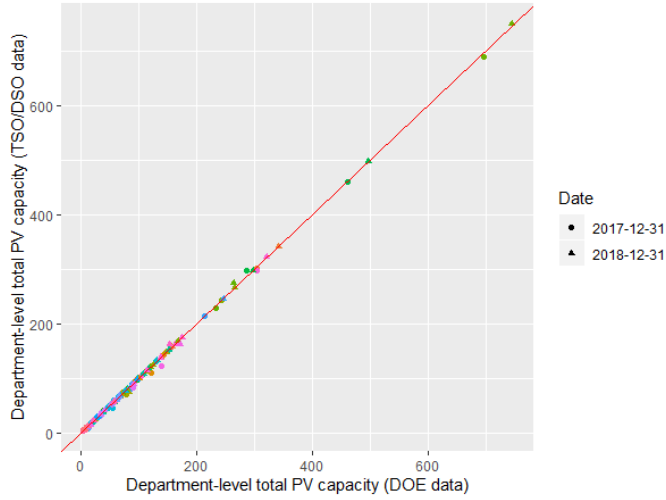


FIGURE A.2. department-level installed PV capacities (in MW) as of 31 December 2017 and 2018 (i) in the DOE dataset (x-axis) and (ii) in the public inventories of power plants 2017 and 2018 (y-axis).

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 department d as of year y was:

$$K^{HS}_{c,d,y} \equiv \frac{W_{c,d}}{\sum_{c' \in d} W_{c',d}} K_{d,y} \tag{A.2}$$

2. **Sequential static approach:** this method assumes 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,[57] we compute $K^*_{d,y}$ such that:

$$\sum_{c \in d} \max(W_{c,d} - K^*_{d,y}, 0) \equiv K_{d,y} \tag{A.3}$$

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

$$K^{SS}_{c,d,y} \equiv \max(W_{c,d} - K^*_{d,y}, 0) \tag{A.4}$$

57. 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 time series. As shown on Figures A.3 and A.4, $K_{d,2018}$ and $\sum_{c \in d} W_{c,d}$ are virtually identical for almost every department.

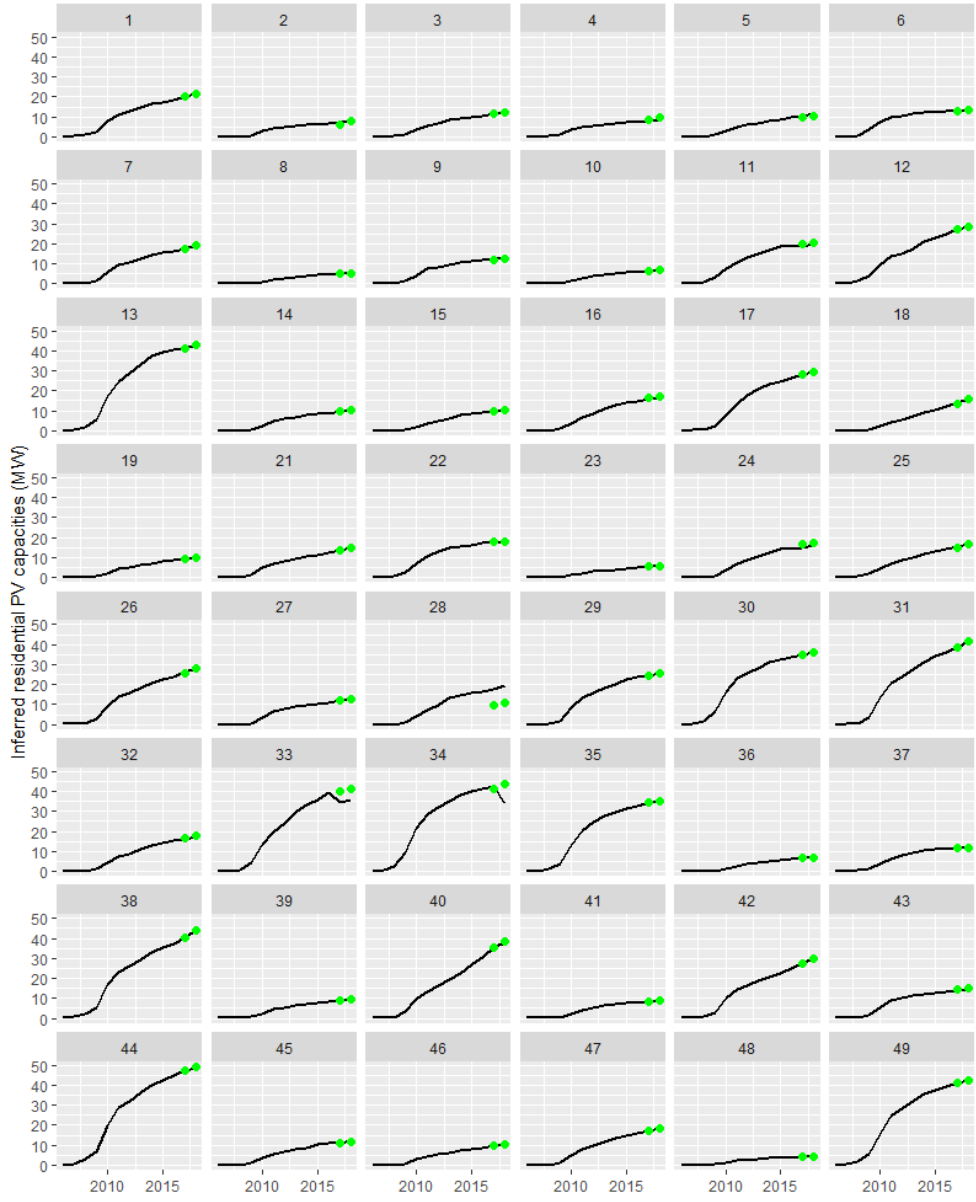


FIGURE A.3. Inferred department-level time series of PV capacity from aggregated units (first 48 departments). Green dots represent actual capacities as reported in the public inventories of power plants of 2017 and 2018.

3. **Homothetic dynamic approach:** although their exact meaning is somewhat ambiguous, the public inventory of power plants does provide

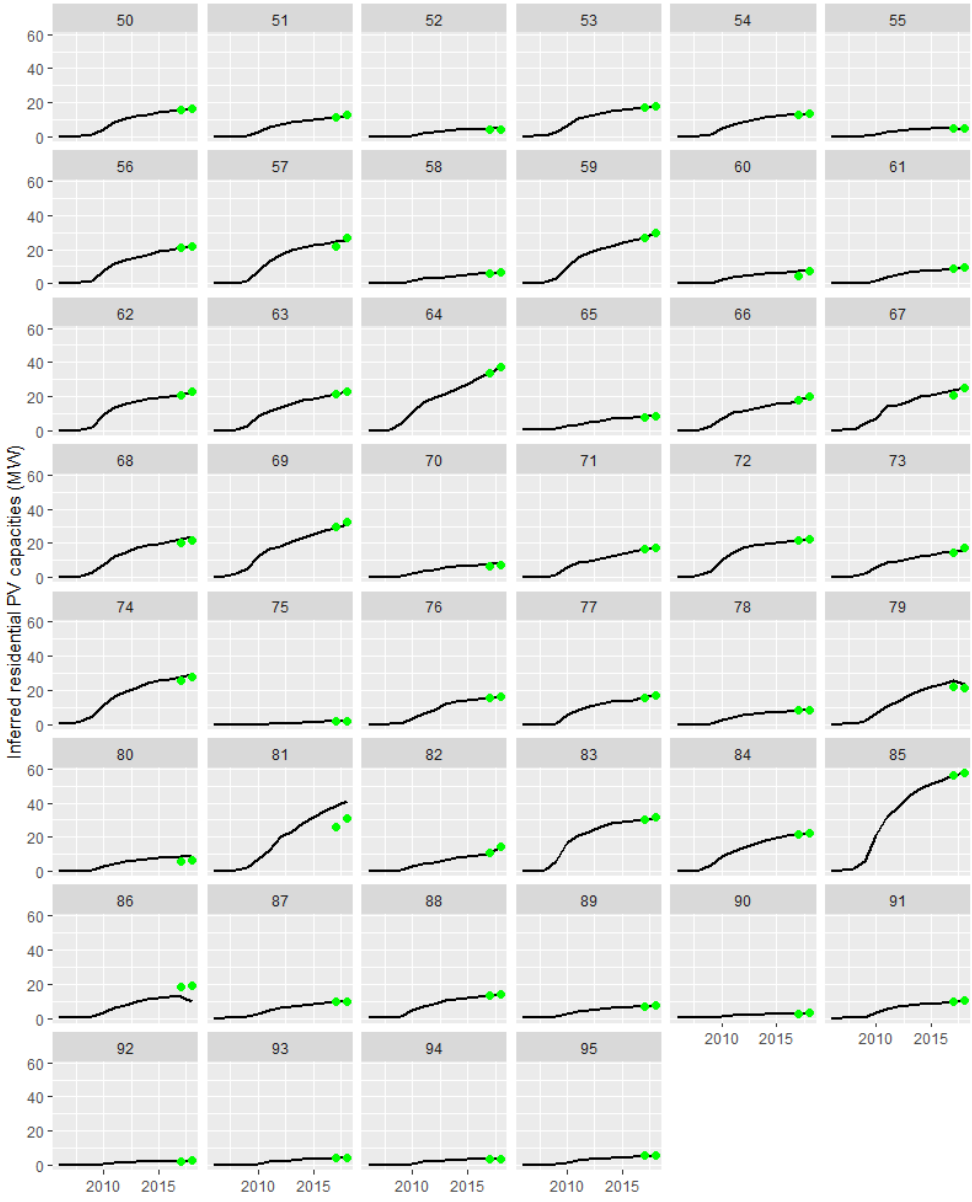


FIGURE A.4. Inferred department-level time series of PV capacity from aggregated units (remaining 46 departments). Green dots represent actual capacities as reported in the public inventories of power plants 2017 and 2018.

"commissioning dates" for aggregated PV observations. We interpret these 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 host aggregated PV units as of a given year. We denote $\mathcal{C}_{d,y}$ the set of (sub)counties in department d whose "commissioning date" is anterior to 31 December of year y.[58] We then define the installed capacity in county c of department d as of year y using an iterative approach:

$$K^{HD}_{c,d,y} \equiv \mathbf{1}_{c\in\mathcal{C}_{d,y}} \left(K^{HD}_{c,d,y-1} + \frac{W_{c,d} - K^{HD}_{c,d,y-1}}{\sum_{c'\in\mathcal{C}_{d,y}} W_{c',d} - K^{HD}_{c',d,y-1}} \left(K_{d,y} - K_{d,y-1}\right) \right) \tag{A.5}$$

where $K^{HD}_{c,d,2005} = 0$ for all c, d and $\mathbf{1}_{c\in\mathcal{C}_{d,y}}$ is a dummy variable that takes the value 1 if county c belongs to $\mathcal{C}_{d,y}$ (and 0 otherwise).

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

$$\sum_{c\in\mathcal{C}_{d,y}} \max(W_{c,d} - K^{SD}_{c,d,y-1} - K^*_{d,y}, 0) \equiv K_{d,y} - K_{d,y-1} \tag{A.6}$$

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

$$K^{SD}_{c,d,y} \equiv \mathbf{1}_{c\in\mathcal{C}_{d,y}} \left(K^{SD}_{c,d,y-1} + \max(W_{c,d} - K^{SD}_{c,d,y-1} - K^*_{d,y}, 0)\right) \tag{A.7}$$

The outcome of steps A and B is four alternative (sub)county-level time series of installed capacities from aggregated PV units that are consistent with their observed evolution at the department level. The results reported in the main text use the time series derived from 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

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

generation units that are listed as individual observations in the inventory of power plants but for which the upstream substation information is missing.

Step C: matching (sub)counties to substations and deriving substation level time series 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.[59] Second, we observe both the upstream substation and the (sub)county of a large number of individually-listed distributed generation units. 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 40 km away from the substation in order 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 departments. Panel (b) further zooms in 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.

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

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

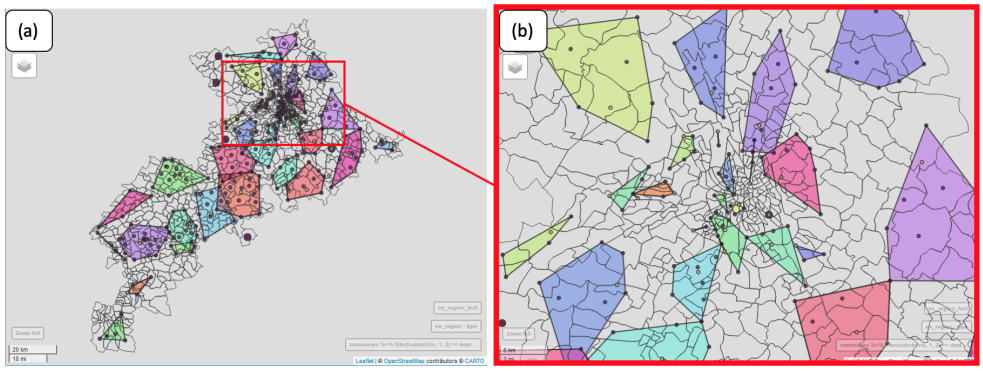


FIGURE A.5. Panel (a): convex hulls for the Haute-Garonne department. Panel (b): zoom on the urban area of the city of Toulouse.

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 departments, 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 to substations, 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 unit 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.

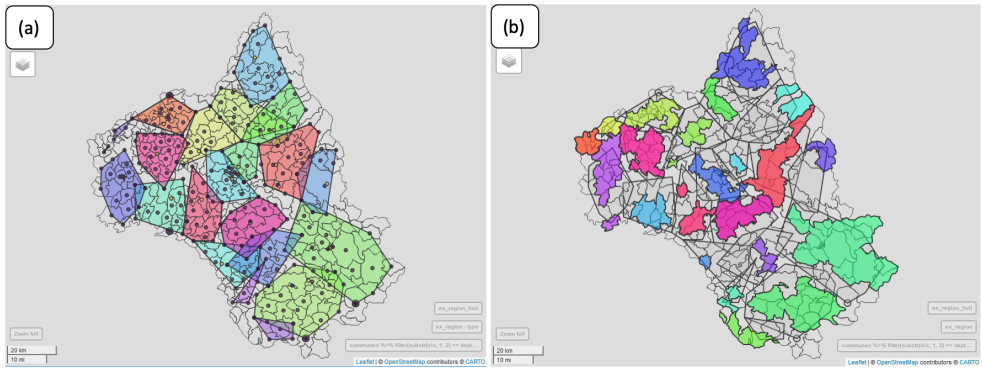


FIGURE A.6. Panel (a): convex hulls for the Aveyron department. Panel (b): (sub)counties that intersect with a single convex hull, and corresponding service territory of the substations (outcome of step 1).

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 service territory DSO in France. They 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 time series 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 installed capacities at the substation-level in 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 the period of our study. In such (rare) cases, 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.

Appendix B: Robustness checks

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, we implement an assignment procedure to make informed guesses about the substation(s) to which distributed generation units whose upstream substation is unknown are most likely to connect. This paragraph shows that our results hold irrespective of the details of the assignment procedure.

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 of small (<36kW) units, for the most part aggregated at the (sub)county level. As discussed in Appendix A, the aggregation procedure prevents us from observing the evolution of aggregated capacities over time at a finer spatial granularity than departments. We thus implement four contrasted methodologies to infer how (sub)county-level installed capacities may have evolved over time. All four approaches are consistent with the known trajectories of department-level capacities.

The time series of (sub)county-level PV capacities from small aggregated units are then aggregated at the substation level using our spatial matching methodology (Step C of the assignment procedure described in Appendix A), where they are added to capacities from individually-listed units. 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.

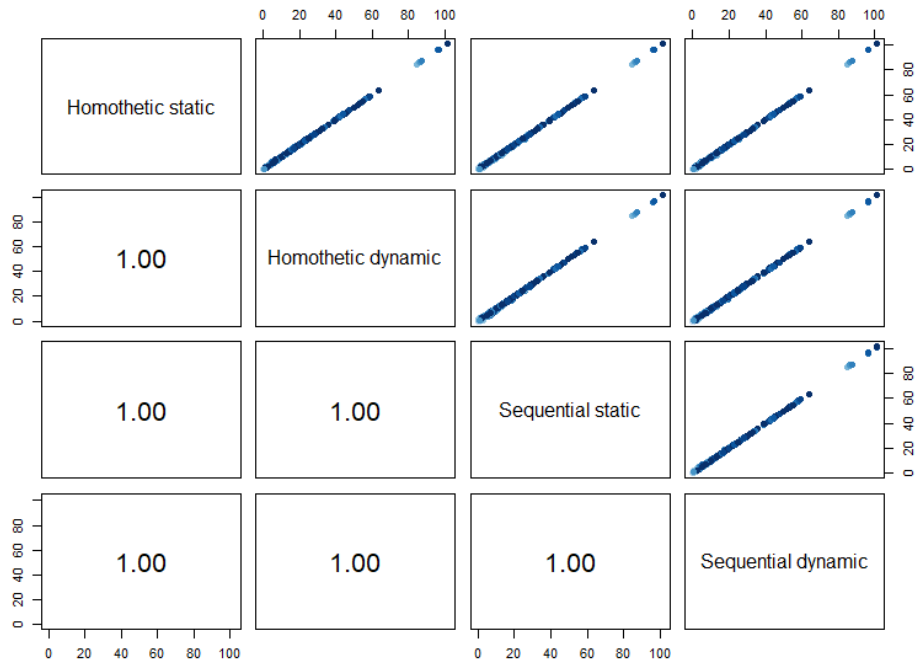


FIGURE B.1. Top-right corner: scatter plots of pair-wise relationship between the different installed PV capacity metrics (unit of observation: substation by year). Darker colors correspond the later years. Bottom-left corner: corresponding coefficients of correlation.

Figure B.1 shows that the four PV metrics we obtain are virtually identical. This result has several explanations. First, 74% of installed PV capacities correspond to individually-listed units. Second, we observe precisely where small aggregated units are as of 2018, which by construction is the year with the highest amount of total installed PV capacity. Third, even if we use contrasted approaches to assign installations to (sub)counties, these spatial units are then aggregated spatially when we infer the territories supplied 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.

Sensitivity to including units whose upstream substation is inferred and not directly observed (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:

$$K_{t,s,y} \equiv K^0_{t,s,y} + \hat{K}_{t,s,y} \tag{B.1}$$

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 B.1. Decomposition of the variance (in MW^2) of the installed capacities of each technology between known and inferred capacities.

Technology	$\text{Var}(K^0_{t,s,y})$	$\text{Var}(\hat{K}_{t,s,y})$	$\text{Var}(K_{t,s,y})$
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 B.1 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 observe the upstream substation in the public inventory of power plants. Consistently, our results are robust to ignoring altogether capacities for which we had to infer the upstream substation. Tables B.2 and B.3 report the results of our quantile regressions when using $K^0_{t,s,y}$ instead of $K_{t,s,y}$ as independent variables. These results are almost identical to the results reported in the main text.

TABLE B.2. Estimated coefficients when regressing the main quantiles of the annual distribution of hourly net withdrawals on the installed capacities $K^0_{t,s,y}$ from units whose upstream substation is known. 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	−0.083	−0.069	−0.060	−0.065	−0.107	−0.130	−0.127
thermal	(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

TABLE B.3. Estimated coefficients when regressing the main quantiles of the annual distribution of hourly ramps on the installed capacities $K^0_{t,s,y}$ from units whose upstream substation is known. 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	−0.002	0.002	−0.0002	−0.001	−0.001	0.001	0.002
thermal	(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

B.2. Potential endogeneity of installed capacities

Substation selection when investing in distributed generation

As shown in Table 2, investments in distributed generation are skewed towards a subset of substations. Substation fixed effects are included to control for any unobserved characteristic of substations that is constant over time and explains net withdrawals to the distribution grid. To check the robustness of this approach, we ran our regressions on two subsets. First, in Table B.4,

we restrict attention to the sample of substations where some investments in wind distributed generation have occurred. The estimated coefficients for wind remains unchanged and the coefficients for other technologies are less precisely estimated but not inconsistent with our main estimates. Second, in Table B.5, we focus on the subset of substations with at least 1 MW of distributed PV generation as of 2018. The coefficient estimates for PV are again consistent with our main results.

TABLE B.4. Estimated coefficients when regressing the main quantiles of the distribution of hourly net withdrawals (for a given substation in a given year) on the installed capacities of the different technologies with substation and year-by-region fixed effects, restricting attention to the subset of substations for which some investment in wind distributed generation have occurred. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	−0.677	−0.434	−0.249	−0.122	−0.073	−0.051	−0.024
	(0.030)	(0.021)	(0.013)	(0.006)	(0.006)	(0.006)	(0.006)
PV	−0.129	−0.101	−0.118	−0.092	−0.065	−0.048	−0.060
	(0.081)	(0.064)	(0.051)	(0.033)	(0.029)	(0.033)	(0.039)
Small hydro	−0.248	−0.298	−0.306	−0.267	−0.263	−0.271	−0.295
	(0.137)	(0.108)	(0.090)	(0.078)	(0.068)	(0.068)	(0.091)
Renewable thermal	−0.414	−0.410	−0.388	−0.388	−0.375	−0.325	−0.274
	(0.201)	(0.143)	(0.089)	(0.067)	(0.065)	(0.074)	(0.101)
Non renewable thermal	−0.172	−0.118	−0.087	−0.040	−0.065	−0.103	−0.135
	(0.088)	(0.072)	(0.052)	(0.037)	(0.044)	(0.048)	(0.058)
Observations	6,061	6,061	6,061	6,061	6,061	6,061	6,061

Unobserved characteristics causing differential trends

Our econometric model is a two-way fixed effect specification. This very parsimonious approach accounts for both substation-specific unobserved characteristics (as long as they are constant over time) and year-specific unobserved characteristics (as long as they are uniform across space). While these controls account for a large number of possible confounders, our estimates could still be biased if there exist unobserved variables that explain electricity consumption that are correlated with installed distributed generation capacities and have experienced both significant and spatially contrasted changes over our period of interest.

TABLE B.5. Estimated coefficients when regressing the main quantiles of the distribution of hourly net withdrawals (for a given substation in a given year) on the installed capacities of the different technologies with substation and year-by-region fixed effects, restricting attention to the subset of substations for which host at least 1 MW of distributed PV at the end of 2018. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	−0.592	−0.384	−0.229	−0.121	−0.086	−0.065	−0.043
	(0.039)	(0.025)	(0.016)	(0.009)	(0.008)	(0.007)	(0.008)
PV	−0.580	−0.400	−0.178	−0.052	−0.020	−0.001	0.008
	(0.037)	(0.036)	(0.019)	(0.011)	(0.011)	(0.012)	(0.015)
Small hydro	−0.207	−0.200	−0.197	−0.174	−0.164	−0.152	−0.130
	(0.106)	(0.075)	(0.069)	(0.060)	(0.061)	(0.065)	(0.073)
Renewable thermal	−0.402	−0.372	−0.370	−0.382	−0.360	−0.317	−0.271
	(0.085)	(0.079)	(0.071)	(0.071)	(0.067)	(0.068)	(0.083)
Non renewable thermal	−0.131	−0.109	−0.108	−0.123	−0.180	−0.210	−0.216
	(0.060)	(0.048)	(0.040)	(0.042)	(0.050)	(0.059)	(0.077)
Observations	12,113	12,113	12,113	12,113	12,113	12,113	12,113

To further alleviate endogeneity concerns, we run our model with larger sets of fixed effects. More precisely, instead of using year fixed effects, we include year-by-region (mainland France has 12 regions, corresponding to NUTS 1 level) and year-by-department (mainland France has 94 departments, corresponding to NUTS 3 level) fixed effects. This set of fixed effects will absorb any unobserved differential trend at a spatial granularity coarser than substations, such as population migrations. Tables B.6 and B.7 show the results when using respectively year-by-region and year-by-department fixed effects. Our results remain qualitatively unchanged.

Finally, potential endogeneity concerns are probably most acute for PV. On the hand, our identifying variation primarily comes from large installations of larger sizes, for which considerations independent from electricity demand (e.g. local acceptability, land availability, the ability to navigate administrative requirements, etc.) have a large influence on location choices. On the other hand, although residential PV represents a small fraction of the total PV capacities in France,[60] larger rooftop PV units could be predominantly installed

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

TABLE B.6. Estimated coefficients when regressing the main quantiles of the distribution of hourly net withdrawals (for a given substation in a given year) on the installed capacities of the different technologies with substation and year-by-region fixed effects. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	−0.657	−0.421	−0.245	−0.126	−0.083	−0.061	−0.033
	(0.026)	(0.018)	(0.011)	(0.005)	(0.006)	(0.006)	(0.006)
PV	−0.541	−0.373	−0.167	−0.053	−0.024	−0.0001	0.019
	(0.040)	(0.037)	(0.019)	(0.011)	(0.011)	(0.013)	(0.015)
Small hydro	−0.391	−0.362	−0.257	−0.154	−0.149	−0.156	−0.140
	(0.067)	(0.058)	(0.031)	(0.026)	(0.037)	(0.042)	(0.045)
Renewable thermal	−0.328	−0.330	−0.330	−0.322	−0.274	−0.233	−0.187
	(0.068)	(0.058)	(0.052)	(0.052)	(0.054)	(0.057)	(0.062)
Non renewable thermal	−0.080	−0.069	−0.059	−0.063	−0.102	−0.129	−0.133
	(0.034)	(0.025)	(0.021)	(0.024)	(0.030)	(0.033)	(0.040)
Observations	30,091	30,091	30,091	30,091	30,091	30,091	30,091
R^2	0.954	0.959	0.976	0.983	0.984	0.986	0.985
Adjusted R^2	0.950	0.955	0.974	0.982	0.983	0.984	0.984

TABLE B.7. Estimated coefficients when regressing the main quantiles of the distribution of hourly net withdrawals (for a given substation in a given year) on the installed capacities of the different technologies with substation and year-by-department fixed effects. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	−0.647	−0.411	−0.239	−0.121	−0.077	−0.054	−0.025
	(0.026)	(0.018)	(0.011)	(0.006)	(0.006)	(0.006)	(0.006)
PV	−0.545	−0.376	−0.170	−0.055	−0.025	−0.001	0.019
	(0.038)	(0.036)	(0.018)	(0.010)	(0.011)	(0.012)	(0.014)
Small_hydro	−0.368	−0.343	−0.239	−0.143	−0.152	−0.160	−0.130
	(0.063)	(0.056)	(0.032)	(0.024)	(0.034)	(0.037)	(0.039)
Renewable_thermal	−0.335	−0.329	−0.327	−0.319	−0.272	−0.230	−0.185
	(0.067)	(0.058)	(0.052)	(0.052)	(0.053)	(0.057)	(0.063)
Non_renewable_thermal	−0.087	−0.074	−0.066	−0.069	−0.109	−0.134	−0.141
	(0.034)	(0.025)	(0.023)	(0.026)	(0.033)	(0.036)	(0.043)
Observations	30,091	30,091	30,091	30,091	30,091	30,091	30,091
R^2	0.956	0.961	0.977	0.984	0.985	0.987	0.987
Adjusted R^2	0.951	0.956	0.974	0.982	0.983	0.985	0.985

on new buildings and thus correlate with local load growth. To assess the magnitude of this concern, we run a placebo test. More specifically, we restrict attention to night-time hours (11pm-5am), for which we know PV output to be equal to zero. We then compute the annual load and ramp duration curves for this subset of hours and estimate our model on this dataset. Because PV output is known to be zero, PV capacity should be found to have a statistically insignificant impact. Tables B.8 and B.9 report the results for hourly net withdrawals and hourly ramps, respectively. First, we do find evidence of a small bias for the impact of PV on the load duration curve. This bias is

however small: net load is on average 0.03 kWh higher per kW of installed distributed PV. A bias of this magnitude would not be sufficient to change our main conclusions. In addition, it is consistent with a solar rebound effect documented in Qiu et al. (2019), in which case the combined effect of distributed PV generation and a solar rebound would be the most relevant metric to inform future grid investments. Second, the placebo impact of PV on hourly ramps is not statistically different from zero.

TABLE B.8. Estimated coefficients when regressing the main quantiles of the distribution of hourly net withdrawals during night-time hours 11pm-5am (for a given substation in a given year) on the installed capacities of the different technologies with substation. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	−0.672	−0.425	−0.261	−0.150	−0.094	−0.066	−0.038
	(0.024)	(0.017)	(0.010)	(0.005)	(0.005)	(0.005)	(0.006)
PV	0.041	0.029	0.027	0.032	0.033	0.031	0.028
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.014)
Small_hydro	−0.384	−0.355	−0.244	−0.149	−0.149	−0.161	−0.142
	(0.062)	(0.060)	(0.031)	(0.024)	(0.036)	(0.038)	(0.039)
Renewable_thermal	−0.383	−0.383	−0.364	−0.320	−0.271	−0.233	−0.192
	(0.065)	(0.058)	(0.053)	(0.049)	(0.050)	(0.050)	(0.057)
Non_renewable_thermal	−0.112	−0.084	−0.073	−0.072	−0.086	−0.098	−0.103
	(0.034)	(0.025)	(0.021)	(0.022)	(0.027)	(0.031)	(0.038)
Observations	30,091	30,091	30,091	30,091	30,091	30,091	30,091
R^2	0.953	0.958	0.972	0.980	0.981	0.982	0.981
Adjusted R^2	0.950	0.955	0.969	0.978	0.980	0.981	0.979

TABLE B.9. Estimated coefficients when regressing the main quantiles of the distribution of hourly ramps during night-time hours 11pm-5am (for a given substation in a given year) on the installed capacities of the different technologies with substation. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	−0.141	−0.041	−0.014	0.003	0.020	0.047	0.144
	(0.005)	(0.002)	(0.001)	(0.0003)	(0.001)	(0.002)	(0.005)
PV	−0.006	−0.006	−0.004	−0.004	−0.001	−0.002	−0.007
	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.004)
Small_hydro	−0.025	−0.004	−0.002	−0.003	0.002	−0.004	−0.013
	(0.011)	(0.005)	(0.003)	(0.003)	(0.004)	(0.007)	(0.014)
Renewable_thermal	−0.014	−0.003	−0.005	0.0004	0.004	0.001	0.004
	(0.014)	(0.010)	(0.008)	(0.004)	(0.005)	(0.009)	(0.012)
Non_renewable_thermal	−0.001	0.001	−0.001	−0.007	−0.007	0.003	0.014
	(0.007)	(0.005)	(0.003)	(0.002)	(0.002)	(0.005)	(0.009)
Observations	30,091	30,091	30,091	30,091	30,091	30,091	30,091
R^2	0.959	0.973	0.975	0.956	0.881	0.901	0.915
Adjusted R^2	0.956	0.971	0.973	0.952	0.872	0.893	0.908

B.3. Heterogeneous treatment effects

Our main specification assumes uniform treatment effects. This simplifying assumption may bias estimates, although one can think of competing stories for the direction of the bias. On the one hand, technological progress seems likely to increase over time the output per MW of newly installed distributed generation. On the other hand, the best sites are likely to be equipped first so that later units will be less productive. As a robustness check, we thus estimate our main model on the periods 2005-2011 and 2012-2018 separately. Tables B.10 and B.11 report the estimates. Although the results for small hydro and non renewable thermal appears to be unstable, our results for wind and PV remain qualitatively similar.

TABLE B.10. Estimated coefficients when regressing the main quantiles of the distribution of hourly net withdrawals (for a given substation in a given year) on the installed capacities of the different technologies, restricting the sample to 2005-2011. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	−0.596	−0.346	−0.208	−0.119	−0.093	−0.073	−0.051
	(0.026)	(0.016)	(0.010)	(0.006)	(0.007)	(0.006)	(0.006)
PV	−0.341	−0.208	−0.081	−0.009	0.016	0.029	0.059
	(0.094)	(0.069)	(0.032)	(0.014)	(0.018)	(0.022)	(0.037)
Small hydro	−0.295	−0.239	−0.194	−0.146	−0.175	−0.203	−0.268
	(0.105)	(0.076)	(0.066)	(0.051)	(0.047)	(0.050)	(0.065)
Renewable thermal	−0.439	−0.348	−0.303	−0.283	−0.262	−0.226	−0.138
	(0.109)	(0.079)	(0.058)	(0.055)	(0.060)	(0.072)	(0.106)
Non renewable thermal	0.062	0.001	0.003	−0.004	−0.022	−0.027	0.037
	(0.054)	(0.053)	(0.052)	(0.052)	(0.059)	(0.062)	(0.079)
Observations	14,907	14,907	14,907	14,907	14,907	14,907	14,907
R^2	0.956	0.969	0.991	0.991	0.991	0.991	0.991
Adjusted R^2	0.948	0.963	0.990	0.990	0.990	0.989	0.990

B.4. Interaction terms

Our main specification does not include interaction terms between the installed capacities of distinct distributed generation technologies. As shown in Table B.12, adding interaction terms does not change our results. In particular, using the statistical test framework described in Appendix D, we cannot reject at the 10% level the null hypothesis that all interaction terms are all equal to zero

TABLE B.11. Estimated coefficients when regressing the main quantiles of the distribution of hourly net withdrawals (for a given substation in a given year) on the installed capacities of the different technologies, restricting the sample to 2012-2018. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	−0.607	−0.375	−0.214	−0.107	−0.066	−0.045	−0.019
	(0.031)	(0.025)	(0.015)	(0.007)	(0.008)	(0.008)	(0.008)
PV	−0.557	−0.369	−0.158	−0.054	−0.027	−0.015	−0.028
	(0.031)	(0.027)	(0.016)	(0.011)	(0.011)	(0.012)	(0.017)
Small hydro	−0.549	−0.463	−0.294	−0.125	−0.081	−0.099	−0.079
	(0.096)	(0.059)	(0.045)	(0.044)	(0.055)	(0.068)	(0.066)
Renewable thermal	−0.335	−0.365	−0.340	−0.312	−0.264	−0.241	−0.261
	(0.083)	(0.076)	(0.072)	(0.078)	(0.072)	(0.074)	(0.097)
Non renewable thermal	−0.167	−0.114	−0.101	−0.104	−0.161	−0.187	−0.179
	(0.053)	(0.040)	(0.035)	(0.035)	(0.046)	(0.054)	(0.053)
Observations	15,184	15,184	15,184	15,184	15,184	15,184	15,184
R^2	0.972	0.972	0.981	0.988	0.989	0.990	0.989
Adjusted R^2	0.967	0.968	0.978	0.986	0.987	0.989	0.987

(a joint test of this hypothesis yields a test-statistic of 1.41). The main text of the article thus reports a more parsimonious specification without interaction terms.

B.5. Spatial regression models

Our main specification considers each distribution network as an independent observation. This approach is motivated by the fact that distribution networks are relatively passive components of the electricity grid: they have a stable radial topology and are not monitored in real-time. One may however be concerned about the potential influence of spatial interactions on our results (e.g. local shocks in temperatures or wind speed).

To assess whether these spatial interactions may bias our estimates, we need to formally define a spatial structure. Because we do not observe the precise topology of the electricity grid, we instead rely on geographical proximity. For simplicity, we restrict attention to the balanced panel of 2,112 substations that we observe throughout the 14-year period. We then define a weight matrix W based on whether two substations are "neighbors", represented as a graph on Figure B.2. More precisely, we first define a matrix with coefficients $w_{i,j}$ that equal to 1 if substations i and j are considered as neighbors, and 0 otherwise.

TABLE B.12. Results for the load duration curve when including interactions terms. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Non_renewable_thermal	−0.100	−0.087	−0.064	−0.059	−0.104	−0.126	−0.130
	(0.041)	(0.032)	(0.025)	(0.028)	(0.035)	(0.038)	(0.046)
PV	−0.546	−0.375	−0.160	−0.040	−0.013	0.004	0.011
	(0.041)	(0.040)	(0.021)	(0.011)	(0.012)	(0.013)	(0.016)
Renewable_thermal	−0.389	−0.377	−0.350	−0.296	−0.245	−0.198	−0.176
	(0.086)	(0.073)	(0.065)	(0.064)	(0.068)	(0.078)	(0.091)
Small_hydro	−0.398	−0.360	−0.226	−0.114	−0.113	−0.124	−0.125
	(0.070)	(0.064)	(0.037)	(0.023)	(0.029)	(0.033)	(0.037)
Wind	−0.695	−0.440	−0.249	−0.126	−0.083	−0.061	−0.032
	(0.028)	(0.021)	(0.012)	(0.006)	(0.006)	(0.006)	(0.007)
PV_x_Wind	0.010	0.005	0.0001	−0.001	−0.001	−0.0004	−0.001
	(0.004)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
PV_x_Hydro	0.018	0.014	0.0002	−0.005	−0.0004	0.002	0.001
	(0.009)	(0.007)	(0.004)	(0.003)	(0.002)	(0.003)	(0.004)
PV_x_Ren_thermal	−0.0005	0.001	0.0004	−0.004	−0.005	−0.005	−0.004
	(0.011)	(0.009)	(0.006)	(0.005)	(0.004)	(0.004)	(0.006)
PV_x_Non_Ren_thermal	0.006	0.008	0.004	0.002	0.003	0.002	0.002
	(0.010)	(0.008)	(0.004)	(0.003)	(0.004)	(0.004)	(0.005)
Wind_x_Hydro	−0.023	−0.021	−0.018	−0.013	−0.009	−0.007	−0.007
	(0.017)	(0.013)	(0.009)	(0.006)	(0.004)	(0.004)	(0.006)
Wind_x_Ren_thermal	0.007	0.005	0.004	0.0001	−0.004	−0.004	−0.004
	(0.011)	(0.008)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Wind_x_Non_Ren_thermal	−0.0004	0.0005	0.0004	0.001	0.001	0.001	0.002
	(0.004)	(0.004)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Hydro_x_Ren_thermal	0.040	0.049	0.039	0.025	0.030	0.030	0.046
	(0.041)	(0.037)	(0.033)	(0.035)	(0.041)	(0.047)	(0.057)
Hydro_x_Non_Ren_thermal	−0.003	−0.006	−0.010	−0.012	−0.016	−0.013	−0.005
	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.014)	(0.016)
Ren_thermal_x_Non_Ren_thermal	0.001	−0.001	−0.003	−0.004	−0.002	−0.004	0.0003
	(0.008)	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.007)
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	0.984	0.985	0.984
Adjusted R^2	0.950	0.955	0.973	0.981	0.982	0.983	0.983

We then row-normalize the matrix so that applying W to a cross-sectional variable yields the vector of the average value of this variable across neighbors. In practice, we use the R package "spdep" (Bivand et al., 2013) to explore two contrasted ways to define neighbors: (i) Delaunay triangulation (removing the few links that are drawn partially outside the territory of mainland France) ; and (ii) a relative neighbors approach that considers that two points p_i and p_j are neighbours if, and only if:

$$d(p_i, p_j) \leq \max\left[d(p_i, p_k),\, d(p_j, p_k)\right] \forall k = 1, ..., N,\, k \neq i, j$$

Both approaches are found to yield similar results for our coefficients of interest. We thus report in what follows relative neighbors approach which is more parsimonious in terms of assumed connections.

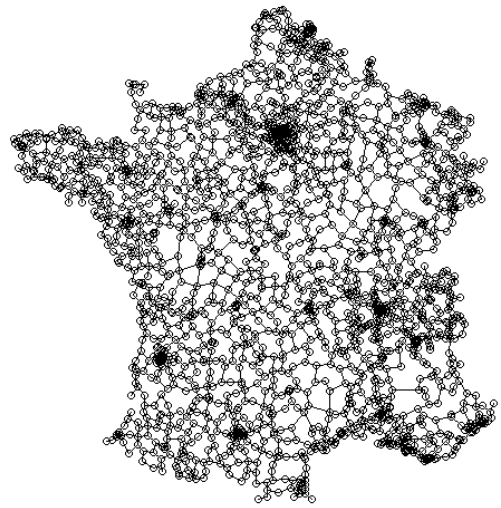


FIGURE B.2. Spatial structure used to build the weight matrix (relative neighbors approach) in our spatial regressions.

We consider a spatial regression model with the following structure (to simplify notations, we express the model in its "instantaneous" form):

$$\begin{cases} Q_y = \rho W Q_y + X_y \beta + \nu + \delta_y \iota_N + u_y \\ u_y = \lambda W u_y + \varepsilon_y \end{cases}$$

where Q_y is the cross-sectional vector of our dependent variable (the value of given percentile for each substation in year y), W the weight matrix (defined above), X_y the matrix of distributed generation capacities in year y (with technologies indexing column and substations indexing rows), ν is the vector of substation fixed effects and δ_y is the year y fixed effect (ι_N is a vector of as many ones as substations). The components of the error term ε_y are assumed i.i.d. normally distributed. The parameters we are interested in are the 5 coefficients in β, which capture the average impact of each distributed generation technology on the quantile of interest.

This model allows for two types of spatial interactions: residual (λ) and endogenous (ρ). First, we expect errors to be spatially correlated because net

withdrawals from nearby substations are influenced by a number of common factors (e.g. local wind speed, temperature, etc.). This concern would affect our inference conclusions. Second, the existence of endogenous spatial interaction could bias our estimates. In particular, even though the topology of the distribution grid is passive, it can be altered by opening or closing electrical switching in outage situations. One may thus wonder whether the distribution system operator does not use this technical possibility to also alter the prevalent topology of the distribution grid, for example to account for changes in local distributed generation capacities.

We estimate the spatial regression model using R package "splm" (Millo and Piras, 2012). Tables B.13 and B.14 report the results, respectively for the load duration curve and the ramp duration curve.

TABLE B.13. Results for the load duration curve when estimating our spatial regression models.

	Dependent variable:						
	Q01	Q10	Q25	Q50	Q75	Q90	Q99
ρ	0.011	0.024	0.018	0.377	0.524	0.587	0.640
	(0.006)	(0.008)	(0.012)	(0.013)	(0.010)	(0.008)	(0.007)
λ	0.058	0.016	-0.006	-0.358	-0.488	-0.525	-0.553
	(0.009)	(0.011)	(0.014)	(0.013)	(0.010)	(0.009)	(0.008)
PV	-0.512	-0.349	-0.155	-0.036	-0.011	0.001	0.004
	(0.006)	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
Wind	-0.661	-0.422	-0.247	-0.102	-0.059	-0.040	-0.022
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Small hydro	-0.367	-0.345	-0.243	-0.136	-0.112	-0.109	-0.100
	(0.026)	(0.024)	(0.019)	(0.017)	(0.018)	(0.019)	(0.022)
Renewable thermal	-0.346	-0.344	-0.338	-0.287	-0.220	-0.177	-0.136
	(0.025)	(0.023)	(0.019)	(0.017)	(0.018)	(0.019)	(0.022)
Non renewable thermal	-0.093	-0.074	-0.064	-0.057	-0.080	-0.089	-0.077
	(0.014)	(0.013)	(0.010)	(0.009)	(0.009)	(0.010)	(0.012)
Observations	29,568	29,568	29,568	29,568	29,568	29,568	29,568
R^2	0.951	0.957	0.975	0.982	0.982	0.984	0.984

Overall, although the estimates from the spatial regressions do suggest the existence of spatial interaction, our coefficients of interest (β) are robust to accounting for them. We however note that the spatial interaction terms (ρ and λ) take high values of opposite signs for some regressions, notably for the 50th, 75th, 90th and 99th percentiles of the load duration curve (Table B.13). Because the joint estimation of these parameters is sometimes reported to

TABLE B.14. Results for the ramp duration curve when estimating our spatial regression models.

	Dependent variable:						
	Q01	Q10	Q25	Q50	Q75	Q90	Q99
ρ	0.015	0.014	0.076	0.676	0.058	0.051	0.001
	(0.006)	(0.007)	(0.009)	(0.007)	(0.009)	(0.008)	(0.007)
λ	0.121	0.110	0.039	-0.496	0.102	0.063	0.168
	(0.009)	(0.010)	(0.012)	(0.009)	(0.011)	(0.010)	(0.009)
PV	-0.155	-0.063	-0.015	-0.001	0.018	0.066	0.142
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)
Wind	-0.145	-0.046	-0.017	0.000	0.018	0.045	0.142
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Small hydro	-0.018	-0.004	0.001	0.001	0.001	-0.002	0.025
	(0.006)	(0.003)	(0.001)	(0.000)	(0.001)	(0.003)	(0.007)
Renewable thermal	-0.006	0.003	-0.001	0.001	0.001	-0.005	-0.001
	(0.006)	(0.002)	(0.001)	(0.000)	(0.001)	(0.003)	(0.006)
Non renewable thermal	-0.003	0.001	0.000	-0.001	-0.001	0.001	0.001
	(0.003)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.004)
Observations	29,568	29,568	29,568	29,568	29,568	29,568	29,568
R^2	0.951	0.964	0.967	0.850	0.967	0.960	0.950

be unstable, we also estimate simpler spatial models with either only spatial lags (Table B.15) or only spatial interactions in the error term (Table B.16). Although the estimates for λ and ρ are indeed found to be take different values, our main coefficients of interest stay by contrast stable.

TABLE B.15. Results for the load duration curve when estimating our spatial regression models with only spatial lags.

	Dependent variable:						
	Q01	Q10	Q25	Q50	Q75	Q90	Q99
ρ	0.039	0.034	0.013	0.042	0.068	0.110	0.157
	(0.004)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
PV	-0.506	-0.348	-0.155	-0.045	-0.015	0.003	0.008
	(0.006)	(0.006)	(0.004)	(0.004)	(0.005)	(0.006)	(0.007)
Wind	-0.655	-0.420	-0.248	-0.128	-0.085	-0.062	-0.035
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Smal hydro	-0.376	-0.346	-0.242	-0.138	-0.126	-0.128	-0.118
	(0.026)	(0.024)	(0.019)	(0.019)	(0.022)	(0.024)	(0.029)
Renewable thermal	-0.347	-0.345	-0.338	-0.328	-0.278	-0.236	-0.187
	(0.025)	(0.023)	(0.019)	(0.018)	(0.021)	(0.023)	(0.028)
Non renewable thermal	-0.094	-0.074	-0.064	-0.067	-0.109	-0.132	-0.126
	(0.014)	(0.013)	(0.010)	(0.010)	(0.012)	(0.013)	(0.016)
Observations	29,568	29,568	29,568	29,568	29,568	29,568	29,568
R^2	0.951	0.957	0.975	0.983	0.984	0.985	0.985

TABLE B.16. Results for the load duration curve when estimating our spatial regression models with only spatially correlated errors

	Dependent variable:						
	Q01	Q10	Q25	Q50	Q75	Q90	Q99
λ	0.070	0.040	0.013	0.044	0.069	0.111	0.157
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
PV	-0.513	-0.352	-0.156	-0.045	-0.016	0.002	0.009
	(0.006)	(0.006)	(0.005)	(0.004)	(0.005)	(0.006)	(0.007)
Wind	-0.663	-0.425	-0.249	-0.129	-0.087	-0.064	-0.036
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Small hydro	-0.365	-0.342	-0.241	-0.136	-0.125	-0.126	-0.111
	(0.026)	(0.024)	(0.019)	(0.019)	(0.022)	(0.025)	(0.030)
Renewable thermal	-0.345	-0.344	-0.338	-0.328	-0.279	-0.237	-0.188
	(0.025)	(0.023)	(0.019)	(0.018)	(0.021)	(0.023)	(0.028)
Non renewable thermal	-0.093	-0.074	-0.064	-0.068	-0.110	-0.134	-0.132
	(0.014)	(0.013)	(0.010)	(0.010)	(0.012)	(0.013)	(0.016)
Observations	29,568	29,568	29,568	29,568	29,568	29,568	29,568
R^2	0.951	0.957	0.975	0.983	0.984	0.985	0.984

Appendix C: Impact of 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 hourly net withdrawals or substation s during hour h of year y. Various summary statistics can then be derived from these distributions. Figure C.1 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.

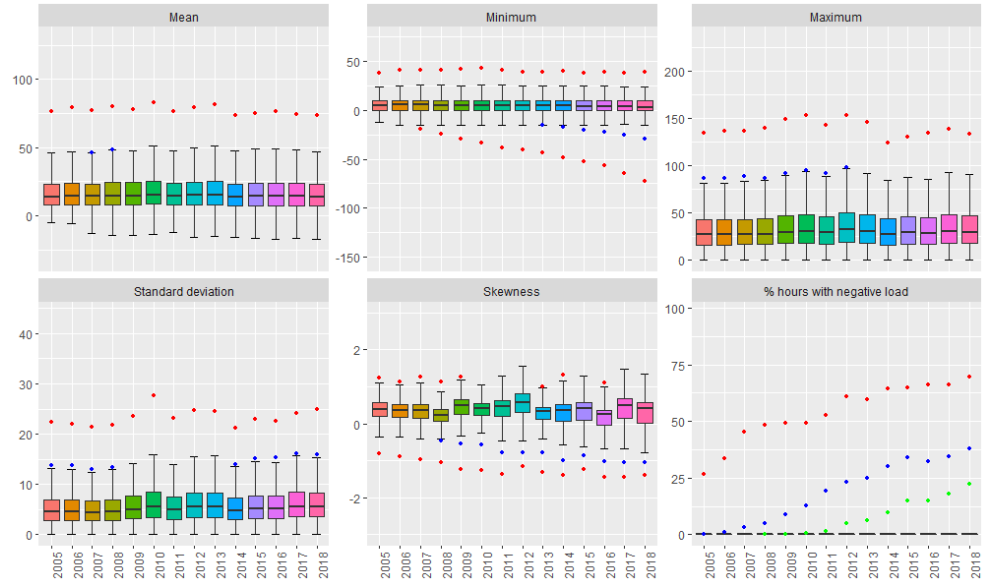


FIGURE C.1. Evolution of a sample of summary statistics for the distribution of hourly net loads in a given year at a given substation. "Minimum" and "maximum" net load are actually the 1st and 999th 1000-quantiles to account for the possibility of idiosyncratic measurement errors. Minimum, maximum, 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.

We observe that the summary statistics that exhibit the most significant changes relate to reverse power flows, that is to hours during which 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 expanded significantly between 2005 and 2018. Figure C.2 further 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 under 1% in 2005 to almost 9% in 2018.

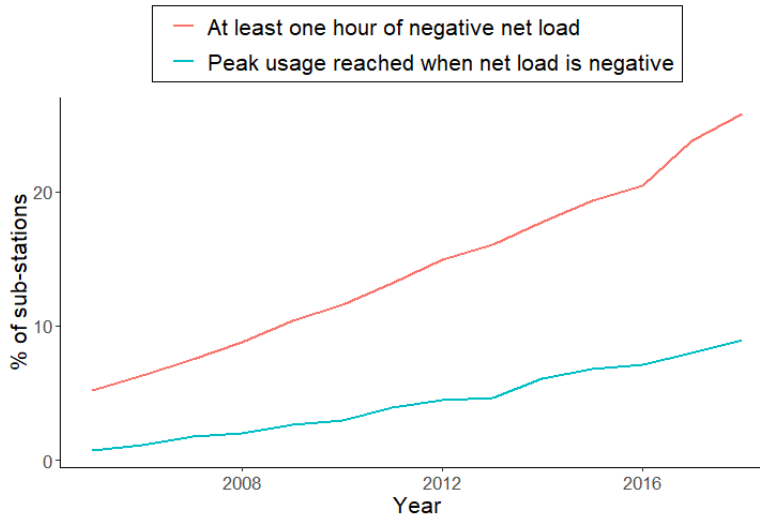


FIGURE C.2. Evolution of the percentage of substations that (i) experienced reverse power flows in a given year; (ii) reached their peak usage (in absolute value) when net load was negative, that is during an hour where they were moving electricity from the distribution to the transmission grid.

We can also estimate the impact that different distributed generation technologies have had on the summary statistics plotted on Figure C.1. To

do so, we use the specification of Equation (1) using these summary statistics as dependent variables. Table C.1 shows our results. First, we note that the estimated coefficients when the dependent variable is mean hourly net load 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 (respectively 14 and 21% according to RTE (2019)). 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 explanations for why we are 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 and more frequent outages.[61] Second, we observe that different technologies have very different impacts on minimum and maximum net load. While non-renewable thermal units impact minimum and maximum net load in a similar way to they impact mean net load, PV and wind have a much lower impact on peak load. 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. In contrast, thermal units have much milder impacts. Finally, we observe that reverse power flows seem to be driven by wind, PV, and small hydro investments.

Appendix D: Statistical Tests

This Appendix provides details on the statistical tests we performed 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

61. Consistent with this hypothesis, 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%.

TABLE C.1. 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	−0.081	−0.091	−0.107	−0.019	−0.005	−0.021
thermal	(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

stack the 7 quantiles regressions 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:

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

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 D.1 reports these test statistics.

TABLE D.1. 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 statistic 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 $\Pr\left[\chi^2_6 \geq 16.812\right] = 0.01$, 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.

Appendix E: Quadratic Specification Detailed Results

This Appendix reports the results from a quadratic specification in installed capacities of distributed generation. Figures E.1 and E.2 plot the corresponding marginal effects for wind and PV, defined as:

$$\hat{\alpha}_{q,t} + 2\hat{\beta}_{q,t} K_{t,s,y}$$

Confidence intervals are built from the variance-covariance matrix with errors clustered at the substation level.

TABLE E.1. Estimated coefficients for a quadratic specification of the impact of distributed generation on the load duration curve. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q01	Q10	Q25	Q50	Q75	Q90	Q99
PV	−0.227	−0.116	−0.058	−0.051	−0.010	0.011	0.014
	(0.050)	(0.033)	(0.017)	(0.015)	(0.018)	(0.020)	(0.024)
PV^2	−0.007	−0.006	−0.002	0.0001	−0.0001	−0.0002	−0.0001
	(0.002)	(0.001)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0004)
Wind	−0.607	−0.366	−0.235	−0.151	−0.133	−0.105	−0.069
	(0.055)	(0.040)	(0.023)	(0.010)	(0.007)	(0.007)	(0.008)
Wind^2	−0.001	−0.001	−0.0002	0.0002	0.001	0.001	0.0004
	(0.001)	(0.001)	(0.0004)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Small_hydro	−0.092	−0.117	−0.147	−0.160	−0.191	−0.182	−0.216
	(0.076)	(0.066)	(0.061)	(0.056)	(0.060)	(0.068)	(0.078)
Small_hydro^2	−0.013	−0.011	−0.004	0.001	0.003	0.002	0.004
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Renewable_thermal	−0.440	−0.438	−0.412	−0.387	−0.344	−0.301	−0.218
	(0.085)	(0.074)	(0.067)	(0.068)	(0.071)	(0.081)	(0.087)
Renewable_thermal^2	0.005	0.005	0.005	0.005	0.005	0.005	0.003
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)
Non_renewable_thermal	0.044	0.020	−0.006	−0.044	−0.092	−0.091	−0.053
	(0.041)	(0.034)	(0.033)	(0.035)	(0.044)	(0.051)	(0.059)
Non_renewable_thermal^2	−0.007	−0.005	−0.003	−0.001	−0.001	−0.002	−0.004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)

TABLE E.2. Estimated coefficients for a quadratic specification of the impact of distributed generation on the ramp duration curve. Robust standard errors clustered at the substation level are reported.

	Dependent variable:						
	Q01	Q10	Q25	Q50	Q75	Q90	Q99
PV	−0.071	−0.026	−0.017	−0.003	0.017	0.037	0.054
	(0.017)	(0.004)	(0.001)	(0.001)	(0.001)	(0.005)	(0.015)
PV^2	−0.002	−0.001	0.00003	0.00001	0.00005	0.001	0.002
	(0.001)	(0.0001)	(0.00002)	(0.00001)	(0.00002)	(0.0002)	(0.001)
Wind	−0.136	−0.040	−0.016	0.0001	0.017	0.041	0.130
	(0.012)	(0.004)	(0.002)	(0.0001)	(0.002)	(0.004)	(0.012)
$Wind^2$	−0.0001	−0.0001	−0.00001	0.00000	0.00001	0.0001	0.0002
	(0.0002)	(0.0001)	(0.00002)	(0.00000)	(0.00002)	(0.0001)	(0.0002)
Small_hydro	0.004	−0.0004	0.005	0.0004	−0.002	−0.013	0.011
	(0.022)	(0.006)	(0.003)	(0.002)	(0.003)	(0.007)	(0.030)
$Small\_hydro^2$	−0.001	−0.0001	−0.0002	0.00002	0.0001	0.0005	0.001
	(0.001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.001)
Renewable_thermal	−0.017	−0.006	−0.001	−0.001	0.002	0.003	0.032
	(0.020)	(0.009)	(0.004)	(0.002)	(0.005)	(0.010)	(0.021)
$Renewable\_thermal^2$	0.00003	0.0003	0.0001	0.0002	−0.0001	−0.0003	−0.002
	(0.001)	(0.001)	(0.0003)	(0.0001)	(0.0004)	(0.001)	(0.002)
Non_renewable_thermal	−0.008	0.002	0.002	−0.0004	−0.004	−0.001	0.007
	(0.012)	(0.005)	(0.002)	(0.001)	(0.002)	(0.005)	(0.013)
$Non\_renewable\_thermal^2$	0.0003	−0.00004	−0.0001	−0.00004	0.0002	0.0001	−0.0003
	(0.0004)	(0.0002)	(0.0001)	(0.00004)	(0.0001)	(0.0002)	(0.0004)

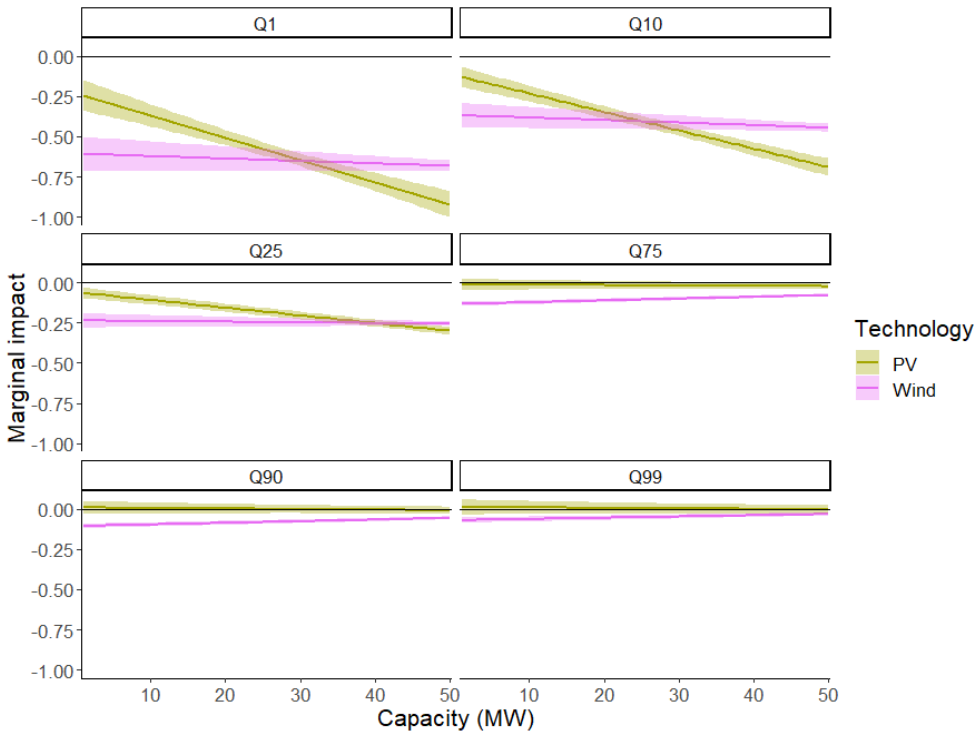


FIGURE E.1. Marginal impact of wind and solar on the load duration curve as a function of installed capacity (confidence intervals are built from the variance-covariance matrix with errors clustered at the sub-station level)

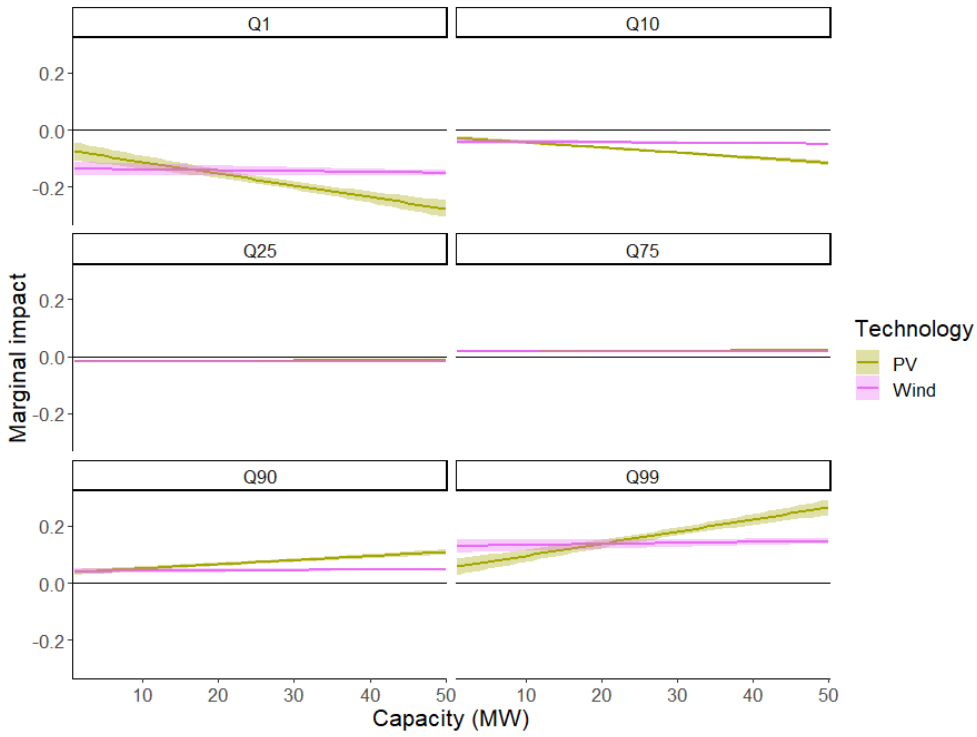


FIGURE E.2. Marginal impact of wind and solar on the ramp duration curve as a function of installed capacity (confidence intervals are built from the variance-covariance matrix with errors clustered at the sub-station level)

Appendix F: Discussion of External Validity

F.1. Spatial Heterogeneity within France

Figure F.1 shows how the installed capacity of distributed wind and solar at the department level has evolved over time. We observe very different adoption patterns. Wind has experienced a large growth in the North but is virtually absent in the South-East. Conversely, PV has grown significantly in the South but much less in the North East.

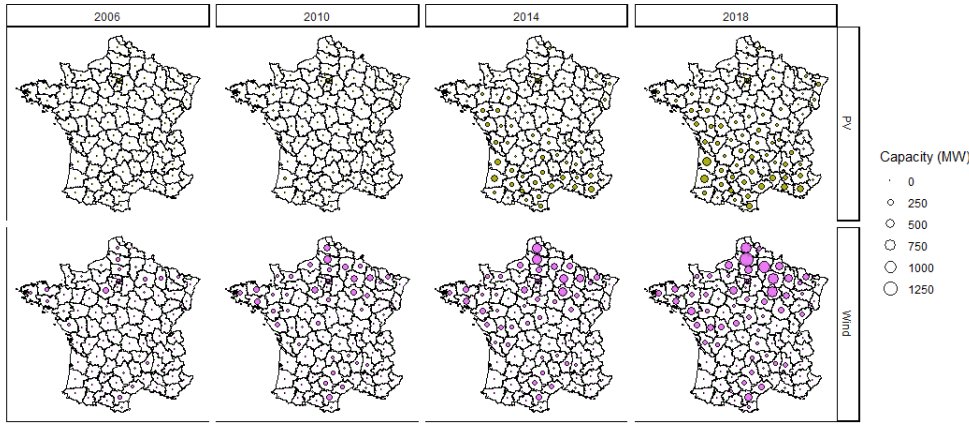


FIGURE F.1. Evolution of the installed capacities of distributed wind and PV at the department level.

Because most departments have low installed capacities of at least one distributed generation technology, we do not study each department in isolation but rather divide France into four "macro regions".[62] Figure F.2 shows how the installed capacity of distributed generation by technology has evolved over time in each of these regions.

Given the observed heterogeneity of adoption patterns across our four macro regions, we re-ran our main model for each macro region separately. Tables F.1

62. France has actually 12 administrative regions which we use to build year-by-region dummies in one of the robustness checks of Appendix B. We use larger "macro regions" here to avoid multiplying the number of different results reported and to benefit from higher statistical power.

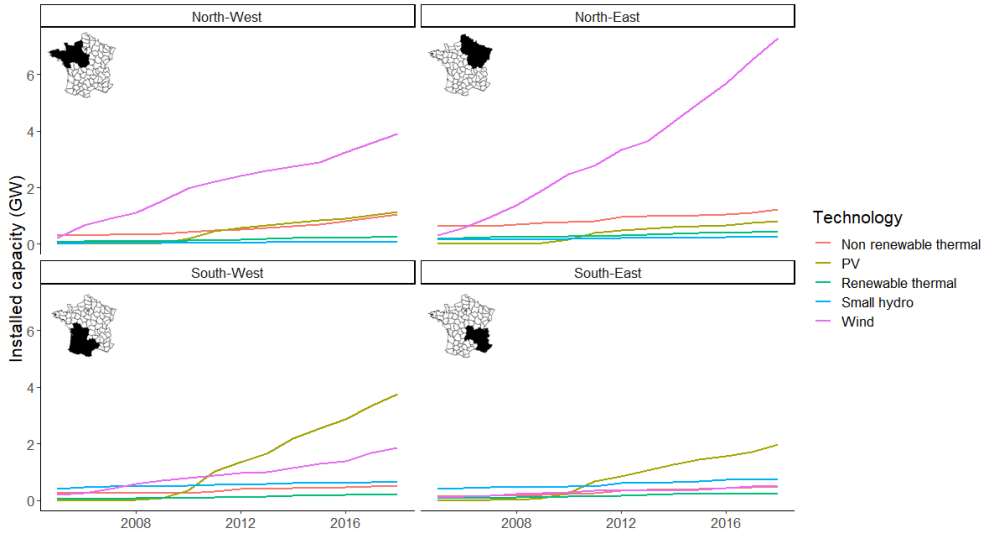


FIGURE F.2. Total installed capacities of distributed generation (in GW) by year and technology (as observed in our final dataset) when dividing France into four macro-region.

to F.4 report the results for the load duration curve. They are qualitatively similar to our main results with two caveats. First, the parameter estimates are noisier and cannot be precisely estimated in regions where a given technology is not sufficiently deployed (e.g. PV in the North). Second, consistently with the results of our quadratic specification, wind has a larger impact on the highest quantiles of the distribution of hourly net withdrawals in the South, where its installed capacity is significantly lower than in the North.

TABLE F.1. Estimated coefficients when regressing the main quantiles of the distribution of hourly net withdrawals (for a given substation in a given year) on the installed capacities of the different technologies, restricting the sample to the North-West macro region. Robust standard errors clustered at the substation level are reported

	Q01	Q10	Q25	Q50	Q75	Q90	Q99
Wind	−0.584	−0.361	−0.216	−0.124	−0.096	−0.077	−0.050
	(0.053)	(0.033)	(0.021)	(0.015)	(0.014)	(0.013)	(0.012)
PV	−0.044	0.029	0.020	−0.011	0.004	0.030	0.094
	(0.104)	(0.076)	(0.054)	(0.052)	(0.057)	(0.065)	(0.076)
Small hydro	−0.144	−0.147	−0.142	−0.131	−0.135	−0.136	−0.122
	(0.216)	(0.143)	(0.115)	(0.109)	(0.134)	(0.151)	(0.150)
Renewable thermal	−0.492	−0.505	−0.475	−0.424	−0.374	−0.329	−0.232
	(0.113)	(0.095)	(0.095)	(0.095)	(0.098)	(0.118)	(0.146)
Non renewable thermal	−0.210	−0.142	−0.115	−0.104	−0.156	−0.209	−0.255
	(0.038)	(0.031)	(0.028)	(0.037)	(0.047)	(0.045)	(0.050)
Observations	7,139	7,139	7,139	7,139	7,139	7,139	7,139
R^2	0.953	0.959	0.971	0.977	0.979	0.979	0.978
Adjusted R^2	0.949	0.956	0.969	0.975	0.977	0.977	0.976

TABLE F.2. Estimated coefficients when regressing the main quantiles of the distribution of hourly net withdrawals (for a given substation in a given year) on the installed capacities of the different technologies, restricting the sample to the North-East macro region. Robust standard errors clustered at the substation level are reported

	Q01	Q10	Q25	Q50	Q75	Q90	Q99
Wind	−0.722	−0.462	−0.259	−0.125	−0.075	−0.052	−0.027
	(0.025)	(0.019)	(0.011)	(0.005)	(0.006)	(0.007)	(0.007)
PV	−0.177	−0.122	−0.060	−0.033	−0.005	0.013	0.008
	(0.094)	(0.067)	(0.043)	(0.039)	(0.048)	(0.052)	(0.051)
Small hydro	−0.327	−0.256	−0.275	−0.309	−0.406	−0.484	−0.614
	(0.134)	(0.128)	(0.125)	(0.135)	(0.142)	(0.156)	(0.175)
Renewable thermal	−0.238	−0.273	−0.254	−0.218	−0.143	−0.107	−0.097
	(0.111)	(0.089)	(0.070)	(0.075)	(0.085)	(0.105)	(0.099)
Non renewable thermal	0.065	0.032	0.020	−0.005	−0.031	−0.028	0.024
	(0.043)	(0.041)	(0.039)	(0.039)	(0.049)	(0.054)	(0.062)
Observations	9,107	9,107	9,107	9,107	9,107	9,107	9,107
R^2	0.961	0.955	0.973	0.983	0.984	0.986	0.988
Adjusted R^2	0.957	0.951	0.971	0.981	0.982	0.985	0.987

TABLE F.3. Estimated coefficients when regressing the main quantiles of the distribution of hourly net withdrawals (for a given substation in a given year) on the installed capacities of the different technologies, restricting the sample to the South-West macro region. Robust standard errors clustered at the substation level are reported

	Q01	Q10	Q25	Q50	Q75	Q90	Q99
Wind	−0.494	−0.353	−0.258	−0.153	−0.114	−0.094	−0.082
	(0.062)	(0.047)	(0.034)	(0.018)	(0.012)	(0.014)	(0.017)
PV	−0.569	−0.378	−0.172	−0.069	−0.042	−0.025	−0.010
	(0.038)	(0.033)	(0.018)	(0.014)	(0.014)	(0.015)	(0.019)
Small hydro	−0.347	−0.339	−0.282	−0.165	−0.146	−0.169	−0.164
	(0.084)	(0.061)	(0.050)	(0.037)	(0.041)	(0.040)	(0.052)
Renewable thermal	−0.340	−0.344	−0.337	−0.368	−0.333	−0.296	−0.281
	(0.133)	(0.123)	(0.108)	(0.106)	(0.086)	(0.090)	(0.112)
Non renewable thermal	−0.039	−0.025	−0.033	−0.027	−0.040	−0.041	−0.014
	(0.103)	(0.078)	(0.061)	(0.058)	(0.070)	(0.077)	(0.093)
Observations	7,148	7,148	7,148	7,148	7,148	7,148	7,148
R^2	0.940	0.958	0.975	0.983	0.984	0.984	0.979
Adjusted R^2	0.936	0.954	0.973	0.982	0.983	0.982	0.978

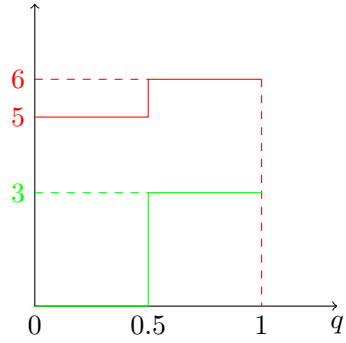
TABLE F.4. Estimated coefficients when regressing the main quantiles of the distribution of hourly net withdrawals (for a given substation in a given year) on the installed capacities of the different technologies, restricting the sample to the South-East macro region. Robust standard errors clustered at the substation level are reported

	Q01	Q10	Q25	Q50	Q75	Q90	Q99
Wind	−0.640	−0.426	−0.262	−0.156	−0.118	−0.092	−0.062
	(0.043)	(0.040)	(0.022)	(0.013)	(0.015)	(0.016)	(0.016)
PV	−0.586	−0.437	−0.200	−0.045	−0.014	0.013	0.029
	(0.065)	(0.058)	(0.029)	(0.014)	(0.015)	(0.017)	(0.019)
Small hydro	−0.452	−0.414	−0.258	−0.131	−0.116	−0.102	−0.073
	(0.084)	(0.079)	(0.044)	(0.033)	(0.043)	(0.045)	(0.051)
Renewable thermal	−0.345	−0.303	−0.323	−0.319	−0.284	−0.229	−0.147
	(0.114)	(0.110)	(0.107)	(0.110)	(0.115)	(0.119)	(0.145)
Non renewable thermal	0.128	0.046	0.028	−0.013	−0.062	−0.059	−0.041
	(0.072)	(0.058)	(0.060)	(0.072)	(0.095)	(0.116)	(0.129)
Observations	6,697	6,697	6,697	6,697	6,697	6,697	6,697
R^2	0.948	0.968	0.979	0.982	0.981	0.980	0.978
Adjusted R^2	0.944	0.966	0.977	0.981	0.979	0.978	0.977

F.2. External Validity to Other Power Systems

A simple observation of renewable generation patterns shows that the high level of contemporaneous correlation in the output of wind/solar units (e.g. night time for PV) mechanically creates hours with consistently low wind/solar generation.[63] In the absence of storage, the persistence of such hours with low generation even at high penetration levels significantly attenuate the ability of wind and solar to decrease peak net demand beyond a certain point.[64] As a result, it seems likely that many of our results for France would carry over at least partially to other power systems.

To formalize this intuition, we start by rephrasing it in the most simple setting. Assume there are only two time periods, with corresponding gross demand levels 5 and 6 MW. The corresponding load duration curve is depicted in red on the pictures below. Further assume that an intermittent distributed generation unit connects to this distribution network, with an output of 0 in the period where demand is 5, and an output of 3 when demand is 6.

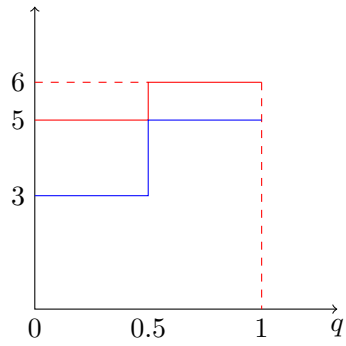


The "direct" impact of the distributed generation unit is thus 0 on the lower quantiles, and -3 on the upper quantiles. However, the period where

63. Wolak (2016) documents the significant degree of contemporaneous correlation in the hourly capacity factors across 40 wind generation units and across the 13 solar generation units in California.

64. Wolak (2022) demonstrates that the significant increase in solar capacity in California has shifted the net peak demand period into the late evening.

gross demand is 6 is no longer the peak period for net withdrawals, that is gross demand minus distributed generation. The new load duration curve is obtained by reshuffling periods in increasing order in terms of net withdrawal (in blue below).



After accounting for "peak shifting", the final impact of the distributed generation unit on the load duration curve is -2 on the lower quantiles, and -1 on the upper quantiles. In other words, the total effect of distributed generation, as estimated by our empirical strategy, may be decomposed into a "direct effect" ($(0, -3)$ in our example) and a "peak shifting effect" ($(-2, +2)$ in our example).

This decomposition can be generalized to a larger number T of time intervals (e.g. $T = 8760$ for our data). Indeed, let L_i denote gross consumption during time period i, where periods are sorted by increasing levels of gross consumption: $L_1 \leq L_2 \leq ... \leq L_T$. We further denote G_i the output from distributed generation in period i. The net withdrawal in period i is then $L_i - G_i$. Because G_i has different determinants than L_i, the metric $L_i - G_i$ needs no longer be increasing in i. We can however define a permutation $\sigma(.)$ such that $j = \sigma(i)$ is the ranking of period i when ordered by net withdrawals. For example, $10 = \sigma(2)$ means that the 2nd lowest value of $\{L_i\}_i$ is reached during the same physical time period (e.g. 10 January 3 am) as the 10th lowest value of $\{L_j - G_j\}_j$. Assuming a capacity K of distributed generation has been

installed and a linear effect, our empirical strategy for a given quantile $q \in \{0, 1\}$ estimates an impact β_q defined as:

$$\beta_q = \frac{(L_{\sigma^{-1}(qT)} - G_{\sigma^{-1}(qT)}) - L_{qT}}{K}$$

This total impact may be decomposed as:

$$\beta_q = \underbrace{-\frac{G_{qT}}{K}}_{\text{Direct effect}} + \underbrace{\frac{(L_{\sigma^{-1}(qT)} - G_{\sigma^{-1}(qT)}) - (L_{qT} - G_{qT})}{K}}_{\text{Peak-shifting effect}}$$

Because we do not observe gross consumption and distributed generation output separately in our data, performing the decomposition between a "direct" and a "peak shifting" effect would require to build a counterfactual gross electricity demand, which is very challenging when the spatial granularity is a distribution network and the time granularity is hourly observations. However, we can perform numerical simulations to get a sense of the likely relative importance of both effects in the case of PV. Indeed, we can (i) retrieve the hourly production profile from all PV capacities at the national level (publicly available from 2012 onward) and assume that local PV production follows the same profile (assuming a capacity factor of 12% below); and (ii) focus attention to a substation where very little distributed generation is installed so that the hourly net withdrawals we observe may be considered as gross electricity demand. Under these strong assumptions, we can disentangle the "direct" and "peak-shifting" effects under increasing installed capacities of distributed PV.

We ran this exercise in 2013 for a randomly chosen substation with low installed capacities of distributed generation for that year (results for other draws of substations were qualitatively similar). This substation is observed to withdraw on average 21 MW from the transmission grid, which we assume to be gross electricity demand. We then simulate the addition of 5 and 50 MW of distributed PV which, given the assumed capacity factor of 12%, generates respectively (over the course of the year) 3% and 29% of the total annual local electricity consumption. Figure F.3 shows the quantile impact functions (which

we can compute exactly given our simplifying assumptions), decomposing it between the "direct" and "peak shifting" effect. In this Figure, red dots represent the direct effect ($-\frac{G_{qT}}{K}$ for each quantile q), green dots the indirect effect ($\frac{(L_{\sigma^{-1}(qT)}-G_{\sigma^{-1}(qT)})-(L_{qT}-G_{qT})}{K}$ for each quantile q), and blue dots the resulting combined effect of the two, which is the total effect our empirical strategy seeks to recover.

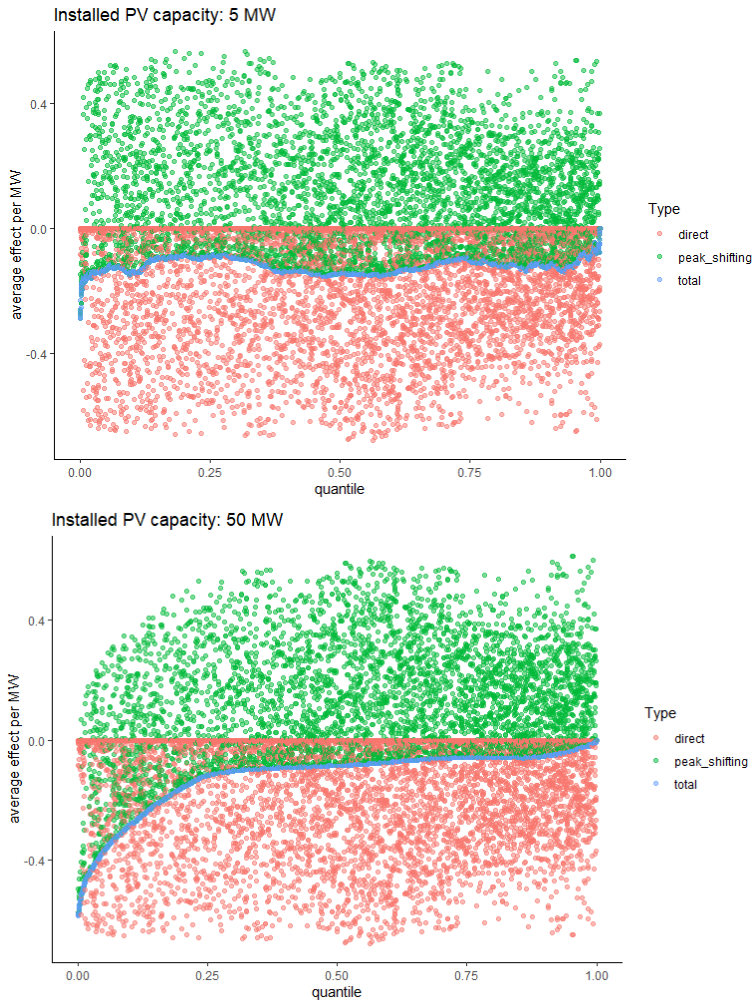


FIGURE F.3. Decomposition of the (average) total effect between a "direct" and a "peak shifting" effect for a randomly chosen substation with little distributed generation in 2013, assuming local hourly distributed PV generation follows the national production profile.

Consistently with our main results, the impact of distributed PV on the highest quantiles is very small when significant levels of installed capacities are assumed. In addition, the "peak shifting" effect greatly attenuates the "direct" impact of distributed generation, that is the impact that would be observed if the ranking of net withdrawals would remain the same as the ranking of gross electricity consumption. The intuition behind this result goes as follows. With low levels of distributed generation ($G << L$), the ranking of hours is almost identical for gross consumption and net withdrawals, so that the peak-shifting effect is close to zero and the direct effect $-\frac{G_{qT}}{K}$ dominates. In contrast, with high levels of distributed generation ($G >> L$), the total effect converge towards $-\frac{G_{\sigma^{-1}(qT)}}{K}$. The ranking σ^{-1} will in turn be significantly influenced by the hourly capacity factor of distributed generation units.